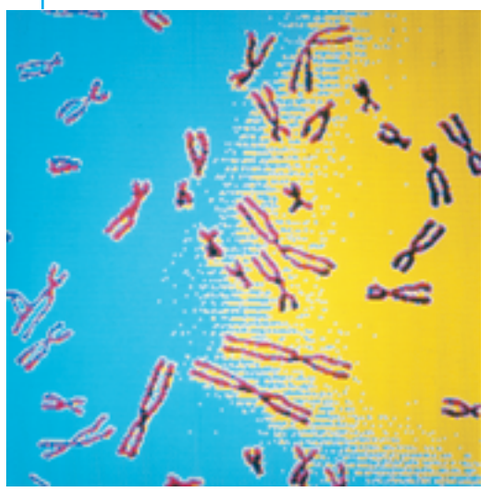
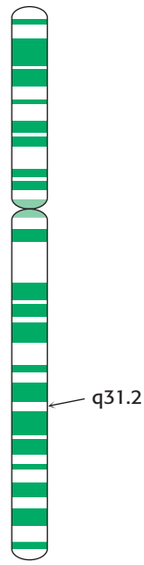
CHAPTER 1

Prelude: Biochemistry and the Genomic Revolution





Disease and the genome. Studies of the human genome are revealing disease origins and other biochemical mysteries. Human chromosomes, left, contain the DNA molecules that constitute the human genome. The staining pattern serves to identify specific regions of a chromosome. On the right is a diagram of human chromosome 7, with band q31.2 indicated by an arrow. A gene in this region encodes a protein that, when malfunctioning, causes cystic fibrosis. [(Left) Alfred Pasieka/Peter Arnold.]

GACTTCACTTCTAATGATGATTATGGGAGAACTGGAGCCT
TCAGAGGGTAAAAATTAAGCACAGTGGAAGAATTTCATTC
TGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAG
AAAATATCTTTGGTGTTTCCTATGATGAATATAGATACAG
AAGCGTCATCAAAGCATGCCAACTAGAAGAG…. This string of letters A, C, G, and T is a part of a DNA sequence. Since the biochemical techniques for DNA sequencing were first developed more than three decades ago, the genomes of dozens of organisms have been sequenced, and many more such sequences will be forthcoming. The information contained in these DNA sequences promises to shed light on many fascinating and important questions. What genes in *Vibrio cholera,* the bacterium that causes cholera, for example, distinguish it from its more benign relatives? How is the development of complex organisms controlled? What are the evolutionary relationships between organisms?

Sequencing studies have led us to a tremendous landmark in the history of biology and, indeed, humanity. *A nearly complete sequence of the entire human genome* has been determined. The string of As, Cs, Gs, and Ts with which we began this book is a tiny part of the human genome sequence, which is more than 3 billion letters long. If we included the entire sequence, our opening sentence would fill more than 500,000 pages.

The implications of this knowledge cannot be overestimated. By using this blueprint for much of what it means to be human, scientists can begin the identification and

OUTLINE

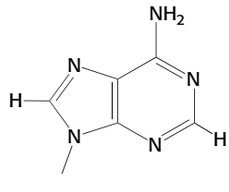
characterization of sequences that foretell the appearance of specific diseases and particular physical attributes. One consequence will be the development of better means of diagnosing and treating diseases. Ultimately, physicians will be able to devise plans for preventing or managing heart disease or cancer that take account of individual variations. Although the sequencing of the human genome is an enormous step toward a complete understanding of living systems, much work needs to be done. Where are the functional genes within the sequence, and how do they interact with one another? How is the information in genes converted into the functional characteristics of an organism? Some of our goals in the study of biochemistry are to learn the concepts, tools, and facts that will allow us to address these questions. It is indeed an exciting time, the beginning of a new era in biochemistry.

1.1 DNA ILLUSTRATES THE RELATION BETWEEN FORM AND FUNCTION

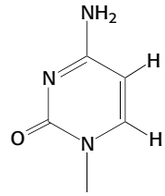
The structure of DNA, an abbreviation for *deoxyribonucleic acid*, illustrates a basic principle common to all biomolecules: the intimate relation between structure and function. The remarkable properties of this chemical substance allow it to function as a very efficient and robust vehicle for storing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

1.1.1 DNA Is Constructed from Four Building Blocks

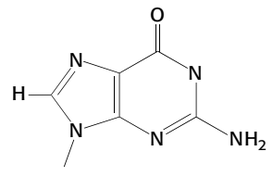
DNA is a *linear polymer* made up of four different monomers. It has a fixed backbone from which protrude variable substituents (Figure 1.1). The backbone is built of repeating sugar-phosphate units. The sugars are molecules of *deoxyribose* from which DNA receives its name. Joined to each deoxyribose is one of four possible bases: adenine (A), cytosine (C), guanine (G), and thymine (T).



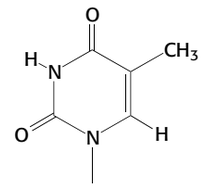
Adenine (A)



Cytosine (C)



Guanine (G)



Thymine (T)

All four bases are planar but differ significantly in other respects. Thus, the monomers of DNA consist of a sugar-phosphate unit, with one of four bases attached to the sugar. *These bases can be arranged in any order along a strand of DNA*. The order of these bases is what is displayed in the sequence that begins this chapter. For example, the first base in the sequence shown is G

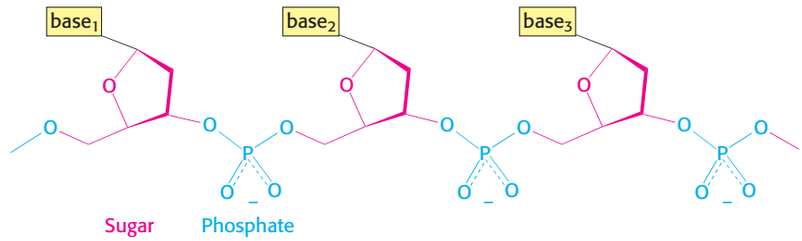


FIGURE 1.1 Covalent structure of DNA. Each unit of the polymeric structure is composed of a sugar (deoxyribose), a phosphate, and a variable base that protrudes from the sugar-phosphate backbone.

(guanine), the second is A (adenine), and so on. *The sequence of bases along a DNA strand constitutes the genetic information*—the instructions for assembling proteins, which themselves orchestrate the synthesis of a host of other biomolecules that form cells and ultimately organisms.

1.1.2 Two Single Strands of DNA Combine to Form a Double Helix

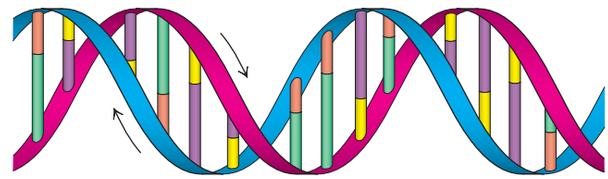


FIGURE 1.2 The double helix. The double-helical structure of DNA proposed by Watson and Crick. The sugar-phosphate backbones of the two chains are shown in red and blue and the bases are shown in green, purple, orange, and yellow.

Most DNA molecules consist of not one but two strands (Figure 1.2). How are these strands positioned with respect to one another? In 1953, James Watson and Francis Crick deduced the arrangement of these strands and proposed a three-dimensional structure for DNA molecules. This structure is a *double helix* composed of two intertwined strands arranged such that the sugar-phosphate backbone lies on the outside and the bases on the inside. The key to this structure is that the bases form *specific base pairs* (bp) held together by *hydrogen bonds* (Section 1.3.1): adenine pairs with thymine (A–T) and guanine pairs with cytosine (G–C), as shown in Figure 1.3. Hydrogen bonds are much weaker than covalent bonds such as the carbon–carbon or carbon–nitrogen bonds that define the structures of the bases themselves. Such weak bonds are crucial to biochemical systems; they are weak enough to be reversibly broken in biochemical processes, yet they are strong enough, when many form simultaneously, to help stabilize specific structures such as the double helix.

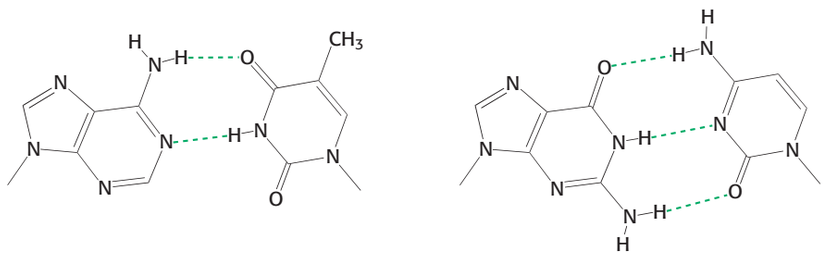


FIGURE 1.3 Watson-Crick base pairs. Adenine pairs with thymine (A–T), and guanine with cytosine (G–C). The dashed lines represent hydrogen bonds.

The structure proposed by Watson and Crick has two properties of central importance to the role of DNA as the hereditary material. First, the structure is compatible with *any sequence of bases*. The base pairs have essentially the same shape (Figure 1.4) and thus fit equally well into the center of the double-helical structure. Second, because of base-pairing, *the sequence of bases along one strand completely determines the sequence along the other strand*. As Watson and Crick so coyly wrote: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." Thus, if the DNA double helix is separated into two single strands, each strand can act as a template for the generation of its partner strand through specific base-pair formation (Figure 1.5). *The three-dimensional structure of DNA beautifully illustrates the close connection between molecular form and function.*

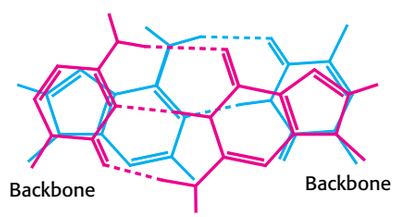


FIGURE 1.4 Base-pairing in DNA. The base-pairs A–T (blue) and C–G (red) are shown overlaid. The Watson-Crick base-pairs have the same overall size and shape, allowing them to fit neatly within the double helix.

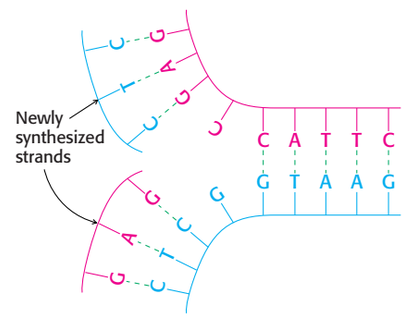
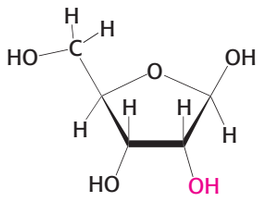
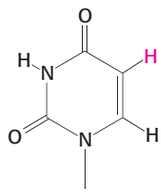


FIGURE 1.5 DNA replication. If a DNA molecule is separated into two strands, each strand can act as the template for the generation of its partner strand.



Ribose



Uracil (U)

1.1.3 RNA Is an Intermediate in the Flow of Genetic Information

An important nucleic acid in addition to DNA is *ribonucleic acid (RNA)*. Some viruses use RNA as the genetic material, and even those organisms that employ DNA must first convert the genetic information into RNA for the information to be accessible or functional. Structurally, RNA is quite similar to DNA. It is a linear polymer made up of a limited number of repeating monomers, each composed of a sugar, a phosphate, and a base. The sugar is ribose instead of deoxyribose (hence, *R*NA) and one of the bases is uracil (U) instead of thymine (T). Unlike DNA, an RNA molecule usually exists as a single strand, although significant segments within an RNA molecule may be double stranded, with G pairing primarily with C and A pairing with U. This intrastrand base-pairing generates RNA molecules with complex structures and activities, including catalysis.

RNA has three basic roles in the cell. First, it serves as the intermediate in the flow of information from DNA to protein, the primary functional molecules of the cell. The DNA is copied, or *transcribed,* into messenger RNA (mRNA), and the mRNA is *translated* into protein. Second, RNA molecules serve as adaptors that translate the information in the nucleic acid sequence of mRNA into information designating the sequence of constituents that make up a protein. Finally, RNA molecules are important functional components of the molecular machinery, called ribosomes, that carries out the translation process. As will be discussed in Chapter 2, the unique position of RNA between the storage of genetic information in DNA and the functional expression of this information as protein as well as its potential to combine genetic and catalytic capabilities are indications that RNA played an important role in the evolution of life.

1.1.4 Proteins, Encoded by Nucleic Acids, Perform Most Cell Functions

A major role for many sequences of DNA is to encode the sequences of *proteins,* the workhorses within cells, participating in essentially all processes. Some proteins are key structural components, whereas others are specific catalysts (termed *enzymes*) that promote chemical reactions. Like DNA and RNA, proteins are linear polymers. However, proteins are more complicated in that they are formed from a selection of 20 building blocks, called *amino acids,* rather than 4.

The functional properties of proteins, like those of other biomolecules, are determined by their three-dimensional structures. Proteins possess an extremely important property: a protein spontaneously folds into a well-defined and elaborate three-dimensional structure that is dictated entirely by the sequence of amino acids along its chain (Figure 1.6). *The self-folding nature of proteins constitutes the transition from the one-dimensional world of sequence information to the three-dimensional world of biological function.* This marvelous ability of proteins to self assemble into complex structures is responsible for their dominant role in biochemistry.

How is the sequence of bases along DNA translated into a sequence of amino acids along a protein chain? We will consider the details of this process in later chapters, but the important finding is that *three bases along a DNA chain encode a single amino acid*. The specific correspondence between a set of three bases and 1 of the 20 amino acids is called the *genetic code*. Like the use of DNA as the genetic material, the genetic code is essentially universal; the same sequences of three bases encode the same amino acids in all life forms from simple microorganisms to complex, multicellular organisms such as human beings.

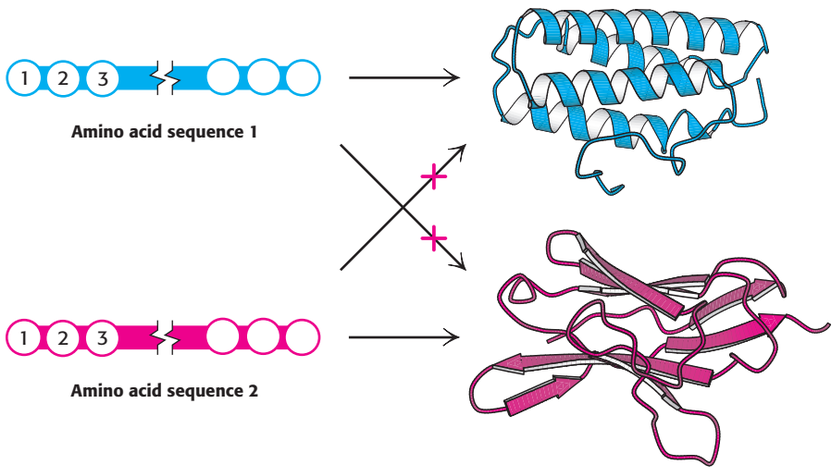


FIGURE 1.6 Folding of a protein. The three-dimensional structure of a protein, a linear polymer of amino acids, is dictated by its amino acid sequence.

Knowledge of the functional and structural properties of proteins is absolutely essential to understanding the significance of the human genome sequence. For example, the sequence at the beginning of this chapter corresponds to a region of the genome that differs in people who have the genetic disorder *cystic fibrosis*. The most common mutation causing cystic fibrosis, the loss of three consecutive Ts from the gene sequence, leads to the loss of a single amino acid within a protein chain of 1480 amino acids. This seemingly slight difference—a loss of 1 amino acid of nearly 1500—creates a life-threatening condition. What is the normal function of the protein encoded by this gene? What properties of the encoded protein are compromised by this subtle defect? Can this knowledge be used to develop new treatments? These questions fall in the realm of biochemistry. Knowledge of the human genome sequence will greatly accelerate the pace at which connections are made between DNA sequences and disease as well as other human characteristics. However, these connections will be nearly meaningless without the knowledge of biochemistry necessary to interpret and exploit them.

Cystic fibrosis—
A disease that results from a decrease in fluid and salt secretion by a transport protein referred to as the cystic fibrosis transmembrane conductance regulator (CFTR). As a result of this defect, secretion from the pancreas is blocked, and heavy, dehydrated mucus accumulates in the lungs, leading to chronic lung infections.

1.2 BIOCHEMICAL UNITY UNDERLIES BIOLOGICAL DIVERSITY

The stunning variety of living systems (Figure 1.7) belies a striking similarity. The common use of DNA and the genetic code by all organisms underlies one of the most powerful discoveries of the past century—namely, that *organisms are remarkably uniform at the molecular level*. All organisms are built from similar molecular components distinguishable by relatively minor variations. *This uniformity reveals that all organisms on Earth have*

FIGURE 1.7 The diversity of living systems. The distinct morphologies of the three organisms shown—a plant (the false hellebora, or Indian poke) and two animals (sea urchins and a common house cat)—might suggest that they have little in common. Yet biochemically they display a remarkable commonality that attests to a common ancestry. [(Left and right) John Dudak/Phototake. (Middle) Jeffrey L. Rotman/Peter Arnold.]

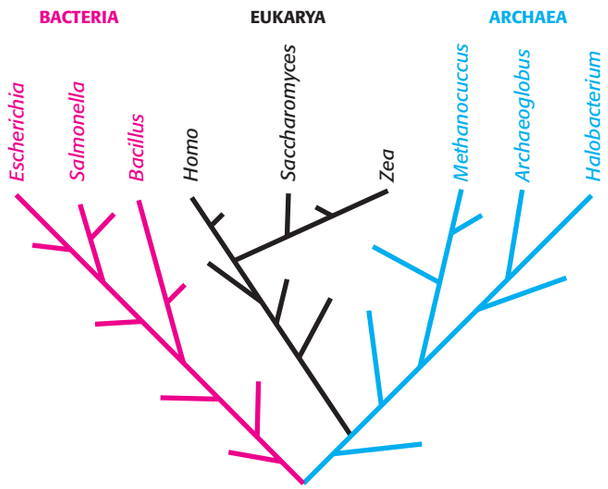


FIGURE 1.8 The tree of life. A possible evolutionary path from a common ancestral cell to the diverse species present in the modern world can be deduced from DNA sequence analysis.

arisen from a common ancestor. A core of essential biochemical processes, common to all organisms, appeared early in the evolution of life. The diversity of life in the modern world has been generated by evolutionary processes acting on these core processes through millions or even billions of years. As we will see repeatedly, the generation of diversity has very often resulted from the adaptation of existing biochemical components to new roles rather than the development of fundamentally new biochemical technology. The striking uniformity of life at the molecular level affords the student of biochemistry a particularly clear view into the essence of biological processes that applies to all organisms from human beings to the simplest microorganisms.

On the basis of their biochemical characteristics, the diverse organisms of the modern world can be divided into three fundamental groups called *domains: Eukarya* (eukaryotes), *Bacteria* (formerly Eubacteria), and *Archaea* (formerly Archaebacteria). Eukarya comprise all macroscopic organisms, including human beings as well as many microscopic, unicellular organisms such as yeast. The defining characteristic of *eukaryotes* is the presence of a well-defined nucleus within each cell. Unicellular organisms such as bacteria, which lack a nucleus, are referred to as *prokaryotes*. The prokaryotes were reclassified as two separate domains in response to Carl Woese's discovery in 1977 that certain bacteria-like organisms are biochemically quite distinct from better-characterized bacterial species. These organisms, now recognized as having diverged from bacteria early in evolution, are archaea. Evolutionary paths from a common ancestor to modern organisms can be developed and analyzed on the basis of biochemical information. One such path is shown in Figure 1.8.

By examining biochemistry in the context of the tree of life, we can often understand how particular molecules or processes helped organisms adapt to specific environments or life styles. We can ask not only *what* biochemical processes take place, but also *why* particular strategies appeared in the course of evolution. In addition to being sources of historical insights, *the answers to such questions are often highly instructive with regard to the biochemistry of contemporary organisms*.

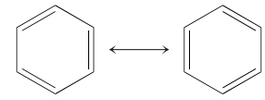
1.3 CHEMICAL BONDS IN BIOCHEMISTRY

The essence of biological processes—the basis of the uniformity of living systems—is in its most fundamental sense molecular interactions; in other words, the chemistry that takes place between molecules. Biochemistry is the *chemistry* that takes place within living systems. To truly understand biochemistry, we need to understand chemical bonding. We review here the types of chemical bonds that are important for biochemicals and their transformations.

The strongest bonds that are present in biochemicals are *covalent bonds*, such as the bonds that hold the atoms together within the individual bases shown in Figure 1.3. A covalent bond is formed by the sharing of a pair of electrons between adjacent atoms. A typical carbon–carbon (C–C) covalent bond has a bond length of 1.54 Å and bond energy of 85 kcal mol^{-1} (356 kJ mol^{-1}). Because this energy is relatively high, considerable energy must be expended to break covalent bonds. More than one electron pair can

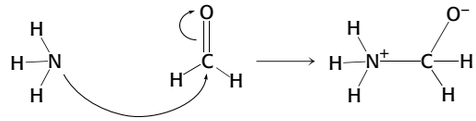
be shared between two atoms to form a multiple covalent bond. For example, three of the bases in Figure 1.4 include carbon–oxygen (C=O) double bonds. These bonds are even stronger than C–C single bonds, with energies near 175 kcal mol^{-1} (732 kJ mol^{-1}).

For some molecules, more than one pattern of covalent bonding can be written. For example, benzene can be written in two equivalent ways called *resonance structures*. Benzene's true structure is a composite of its two resonance structures. A molecule that can be written as several resonance structures of approximately equal energies has greater stability than does a molecule without multiple resonance structures. Thus, because of its resonance structures, benzene is unusually stable.



Benzene resonance structures

Chemical reactions entail the breaking and forming of covalent bonds. The flow of electrons in the course of a reaction can be depicted by curved arrows, a method of representation called "arrow pushing." Each arrow represents an electron pair.



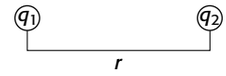
1.3.1 Reversible Interactions of Biomolecules Are Mediated by Three Kinds of Noncovalent Bonds

Readily reversible, noncovalent molecular interactions are key steps in the dance of life. Such weak, noncovalent forces play essential roles in the faithful replication of DNA, the folding of proteins into intricate three-dimensional forms, the specific recognition of substrates by enzymes, and the detection of molecular signals. Indeed, all biological structures and processes depend on the interplay of noncovalent interactions as well as covalent ones. The three fundamental noncovalent bonds are *electrostatic interactions, hydrogen bonds,* and *van der Waals interactions*. They differ in geometry, strength, and specificity. Furthermore, these bonds are greatly affected in different ways by the presence of water. Let us consider the characteristics of each:

1. *Electrostatic interactions*. An electrostatic interaction depends on the electric charges on atoms. The energy of an electrostatic interaction is given by *Coulomb's law:*

$$E = kq_1q_2/Dr$$

where E is the energy, q_1 and q_2 are the charges on the two atoms (in units of the electronic charge), r is the distance between the two atoms (in angstroms), D is the dielectric constant (which accounts for the effects of the intervening medium), and k is a proportionality constant ($k = 332$, to give energies in units of kilocalories per mole, or 1389, for energies in kilojoules per mole). Thus, the electrostatic interaction between two atoms bearing single opposite charges separated by 3 Å in water (which has a dielectric constant of 80) has an energy of 1.4 kcal mol^{-1} (5.9 kJ mol^{-1}).



2. *Hydrogen bonds*. Hydrogen bonds are relatively weak interactions, which nonetheless are crucial for biological macromolecules such as DNA and proteins. These interactions are also responsible for many of the properties of water that make it such a special solvent. The hydrogen atom in a hydrogen bond is partly shared between two relatively electronegative atoms such as nitrogen or oxygen. The *hydrogen-bond donor* is the group that includes both the atom to which the hydrogen is more tightly linked and the hydrogen atom itself, whereas the *hydrogen-bond acceptor* is the atom less tightly linked to the hydrogen atom (Figure 1.9). Hydrogen bonds are fundamentally

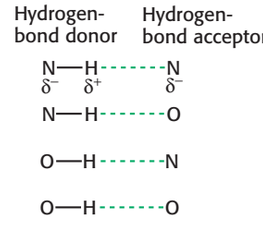
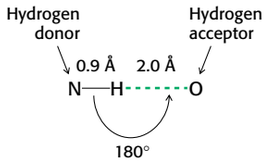


FIGURE 1.9 Hydrogen bonds that include nitrogen and oxygen atoms. The positions of the partial charges (δ^+ and δ^-) are shown.



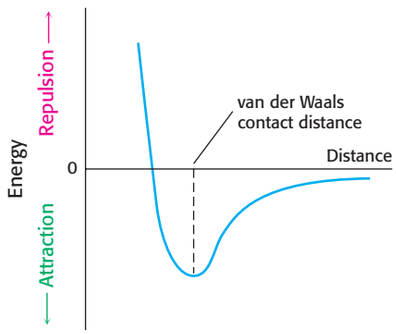


FIGURE 1.10 Energy of a van der Waals interaction as two atoms approach one another. The energy is most favorable at the van der Waals contact distance. The energy rises rapidly owing to electron–electron repulsion as the atoms move closer together than this distance.

electrostatic interactions. The relatively electronegative atom to which the hydrogen atom is covalently bonded pulls electron density away from the hydrogen atom so that it develops a partial positive charge (δ^+). Thus, it can interact with an atom having a partial negative charge (δ^-) through an electrostatic interaction.

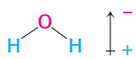
Hydrogen bonds are much weaker than covalent bonds. They have energies of 1–3 kcal mol^{-1} (4–13 kJ mol^{-1}) compared with approximately 100 kcal mol^{-1} (418 kJ mol^{-1}) for a carbon–hydrogen covalent bond. Hydrogen bonds are also somewhat longer than are covalent bonds; their bond distances (measured from the hydrogen atom) range from 1.5 to 2.6 Å; hence, distances ranging from 2.4 to 3.5 Å separate the two nonhydrogen atoms in a hydrogen bond. The strongest hydrogen bonds have a tendency to be approximately straight, such that the hydrogen-bond donor, the hydrogen atom, and the hydrogen-bond acceptor lie along a straight line.

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

Energies associated with van der Waals interactions are quite small; typical interactions contribute from 0.5 to 1.0 kcal mol^{-1} (from 2 to 4 kJ mol^{-1}) per atom pair. When the surfaces of two large molecules come together, however, a large number of atoms are in van der Waals contact, and the net effect, summed over many atom pairs, can be substantial.

1.3.2 The Properties of Water Affect the Bonding Abilities of Biomolecules

Weak interactions are the key means by which molecules interact with one another—enzymes with their substrates, hormones with their receptors, antibodies with their antigens. The strength and specificity of weak interactions are highly dependent on the medium in which they take place, and the majority of biological interactions take place in water. Two properties of water are especially important biologically:



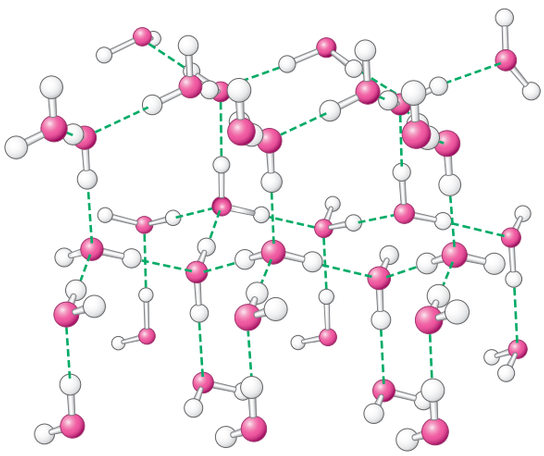
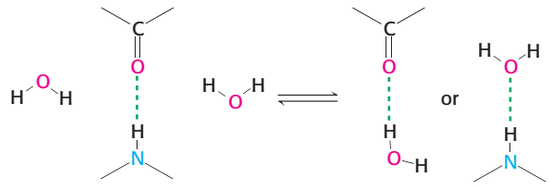


FIGURE 1.11 Structure of ice. Hydrogen bonds (shown as dashed lines) are formed between water molecules.

1. *Water is a polar molecule*. The water molecule is bent, not linear, and so the distribution of charge is asymmetric. The oxygen nucleus draws electrons away from the hydrogen nuclei, which leaves the region around the hydrogen nuclei with a net positive charge. The water molecule is thus an electrically polar structure.

2. *Water is highly cohesive*. Water molecules interact strongly with one another through hydrogen bonds. These interactions are apparent in the structure of ice (Figure 1.11). Networks of hydrogen bonds hold the structure together; similar interactions link molecules in liquid water and account for the cohesion of liquid water, although, in the liquid state, some of the hydrogen bonds are broken. The highly cohesive nature of water dramatically affects the interactions between molecules in aqueous solution.

What is the effect of the properties of water on the weak interactions discussed in Section 1.3.1? The polarity and hydrogen-bonding capability of water make it a highly interacting molecule. Water is an excellent solvent for polar molecules. The reason is that water greatly weakens electrostatic forces and hydrogen bonding between polar molecules by competing for their attractions. For example, consider the effect of water on hydrogen bonding between a carbonyl group and the NH group of an amide.



A hydrogen atom of water can replace the amide hydrogen atom as a hydrogen-bond donor, whereas the oxygen atom of water can replace the carbonyl oxygen atom as a hydrogen-bond acceptor. Hence, a strong hydrogen bond between a CO group and an NH group forms only if water is excluded.

The dielectric constant of water is 80, so water diminishes the strength of electrostatic attractions by a factor of 80 compared with the strength of those same interactions in a vacuum. The dielectric constant of water is unusually high because of its polarity and capacity to form oriented solvent shells around ions. These oriented solvent shells produce electric fields of their own, which oppose the fields produced by the ions. Consequently, the presence of water markedly weakens electrostatic interactions between ions.

The existence of life on Earth depends critically on the capacity of water to dissolve a remarkable array of polar molecules that serve as fuels, building blocks, catalysts, and information carriers. High concentrations of these polar molecules can coexist in water, where they are free to diffuse and interact with one another. However, the excellence of water as a solvent poses a problem, because it also weakens interactions between polar molecules. *The presence of water-free microenvironments within biological systems largely circumvents this problem*. We will see many examples of these specially constructed niches in protein molecules. Moreover, the presence of water with its polar nature permits another kind of weak interaction to take place, one that drives the folding of proteins (Section 1.3.4) and the formation of cell boundaries (Section 12.3).

The essence of these interactions, like that of all interactions in biochemistry, is energy. To understand much of biochemistry—bond formation, molecular structure, enzyme catalysis—we need to understand energy. Thermodynamics provides a valuable tool for approaching this topic. We will revisit this topic in more detail when we consider enzymes (Chapter 8) and the basic concepts of metabolism (Chapter 14).

1.3.3 Entropy and the Laws of Thermodynamics

The highly structured, organized nature of living organisms is apparent and astonishing. This organization extends from the organismal through the cellular to the molecular level. Indeed, biological processes can seem magical in that the well-ordered structures and patterns emerge from the chaotic and disordered world of inanimate objects. However, the organization visible in a cell or a molecule arises from biological events that are subject to the same physical laws that govern all processes—in particular, the *laws of thermodynamics*.

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

Another important thermodynamic concept is that of *entropy*. Entropy is a measure of the level of randomness or disorder in a system. *The Second Law of Thermodynamics states that the total entropy of a system and its surroundings always increases for a spontaneous process*. At first glance, this law appears to contradict much common experience, particularly about biological systems. Many biological processes, such as the generation of a well-defined structure such as a leaf from carbon dioxide gas and other nutrients, clearly increase the level of order and hence decrease entropy. Entropy may be decreased locally in the formation of such ordered structures only if the entropy of other parts of the universe is increased by an equal or greater amount.

An example may help clarify the application of the laws of thermodynamics to a chemical system. Consider a container with 2 moles of hydrogen gas on one side of a divider and 1 mole of oxygen gas on the other (Figure 1.12). If the divider is removed, the gases will intermingle spontaneously to form a uniform mixture. The process of mixing increases entropy as an ordered arrangement is replaced by a randomly distributed mixture.

Other processes within this system can decrease the entropy locally while increasing the entropy of the universe. A spark applied to the mixture initiates a chemical reaction in which hydrogen and oxygen combine to form water:

$$2\,H_2 + O_2 \longrightarrow 2\,H_2O$$

If the temperature of the system is held constant, the entropy of the system decreases because 3 moles of two differing reactants have been combined to form 2 moles of a single product. The gas now consists of a uniform set of indistinguishable molecules. However, the reaction releases a significant amount of heat into the surroundings, and this heat will increase the entropy of the surrounding molecules by increasing their random movement.

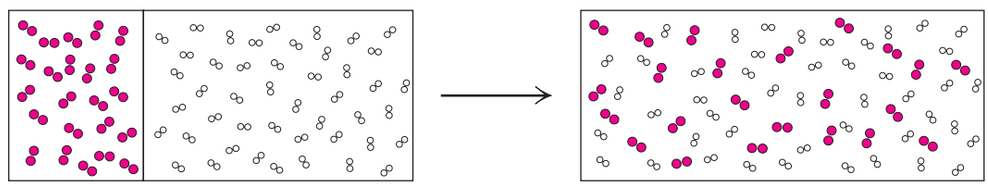


FIGURE 1.12 From order to disorder. The spontaneous mixing of gases is driven by an increase in entropy.

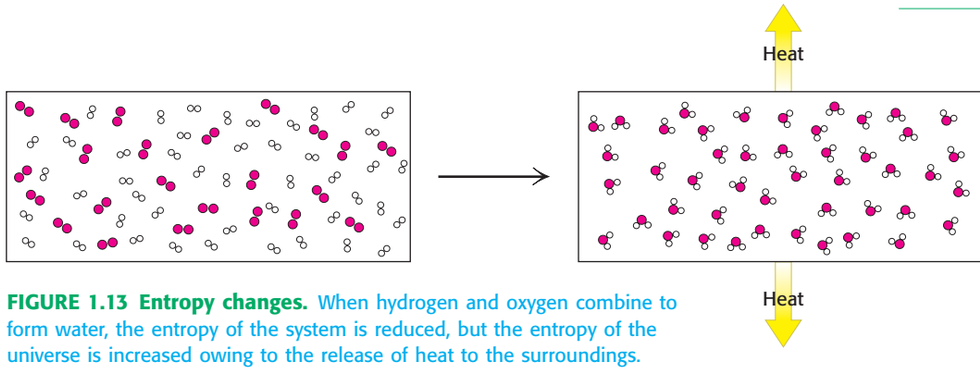


FIGURE 1.13 Entropy changes. When hydrogen and oxygen combine to form water, the entropy of the system is reduced, but the entropy of the universe is increased owing to the release of heat to the surroundings.

The entropy increase in the surroundings is enough to allow water to form spontaneously from hydrogen and oxygen (Figure 1.13).

The change in the entropy of the surroundings will be proportional to the amount of heat transferred from the system and inversely proportional to the temperature of the surroundings, because an input of heat leads to a greater increase in entropy at lower temperatures than at higher temperatures. In biological systems, T [in kelvin (K), absolute temperature] is assumed to be constant. If we define the heat content of a system as *enthalpy (H),* then we can express the relation linking the entropy *(S)* of the surroundings to the transferred heat and temperature as a simple equation:

$$\Delta S_{\text{surroundings}} = -\Delta H_{\text{system}}/T \tag{1}$$

The total entropy change is given by the expression

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{surroundings}} \tag{2}$$

Substituting equation 1 into equation 2 yields

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} - \Delta H_{\text{system}}/T \tag{3}$$

Multiplying by $-T$ gives

$$-T\Delta S_{\text{total}} = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \tag{4}$$

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

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \tag{5}$$

The free-energy change, ΔG, will be used throughout this book to describe the energetics of biochemical reactions.

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

$$\Delta S_{\text{system}} > \Delta H_{\text{system}}/T \tag{6}$$

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

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

In other words, *the free-energy change must be negative for a reaction to be spontaneous*. A negative free-energy change occurs with an increase in the overall entropy of the universe. Thus, we need to consider only one term, the free energy of the system, to decide whether a reaction can occur spontaneously; any effects of the changes within the system on the rest of the universe are automatically taken into account.

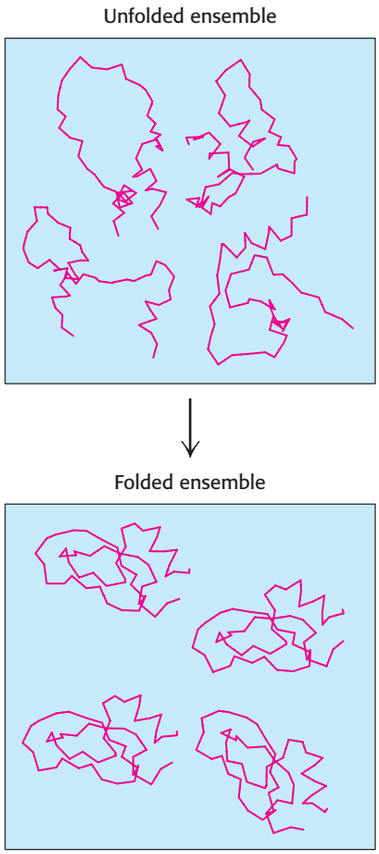


FIGURE 1.14 Protein folding. Protein folding entails the transition from a disordered mixture of unfolded molecules to a relatively uniform solution of folded protein molecules.

1.3.4 Protein Folding Can Be Understood in Terms of Free-Energy Changes

The problem of protein folding illustrates the utility of the concept of free energy. Consider a system consisting of a solution of unfolded protein molecules in aqueous solution (Figure 1.14). Each unfolded protein molecule can adopt a unique conformation, so the system is quite disordered and the entropy of the collection of molecules is relatively high. Yet, protein folding proceeds spontaneously under appropriate conditions. Thus, entropy must be increasing elsewhere in the system or in the surroundings. How can we reconcile the apparent contradiction that proteins spontaneously assume an ordered structure, and yet entropy increases? The entropy decrease in the system on folding is not as large as it appears to be, because of the properties of water. Molecules in aqueous solution interact with water molecules through the formation of hydrogen and ionic interactions. However, some molecules (termed *nonpolar molecules*) cannot participate in hydrogen or ionic interactions. The interactions of nonpolar molecules with water are not as favorable as are interactions between the water molecules themselves. The water molecules in contact with these nonpolar surfaces form "cages" around the nonpolar molecule, becoming more well ordered (and, hence, lower in entropy) than water molecules free in solution. As two such nonpolar molecules come together, some of the water molecules are released, and so they can interact freely with bulk water (Figure 1.15). Hence, nonpolar molecules have a tendency to aggregate in water because the entropy of the water is increased through the release of water molecules. This phenomenon, termed the *hydrophobic effect*, helps promote many biochemical processes.

How does the hydrophobic effect favor protein folding? Some of the amino acids that make up proteins have nonpolar groups. These nonpolar amino acids have a strong tendency to associate with one another inside the interior of the folded protein. The increased entropy of water resulting from the interaction of these hydrophobic amino acids helps to compensate for the entropy losses inherent in the folding process.

Hydrophobic interactions are not the only means of stabilizing protein structure. Many weak bonds, including hydrogen bonds and van der Waals interactions, are formed in the protein-folding process, and heat is released into the surroundings as a consequence. Although these interactions replace interactions with water that take place in the unfolded protein, the net result is the release of heat to the surroundings and thus a negative (favorable) change in enthalpy for the system.

The folding process can occur when the combination of the entropy associated with the hydrophobic effect and the enthalpy change associated with hydrogen bonds and van der Waals interactions makes the overall free energy negative.

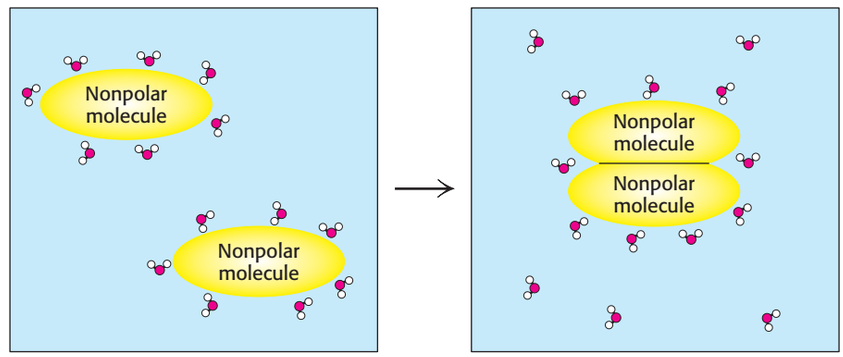


FIGURE 1.15 The hydrophobic effect. The aggregation of nonpolar groups in water leads to an increase in entropy owing to the release of water molecules into bulk water.

1.4 BIOCHEMISTRY AND HUMAN BIOLOGY

Our understanding of biochemistry has had and will continue to have extensive effects on many aspects of human endeavor. *First, biochemistry is an intrinsically beautiful and fascinating body of knowledge*. We now know the essence and many of the details of the most fundamental processes in biochemistry, such as how a single molecule of DNA replicates to generate two identical copies of itself and how the sequence of bases in a DNA molecule determines the sequence of amino acids in an encoded protein. Our ability to describe these processes in detailed, mechanistic terms places a firm chemical foundation under other biological sciences. Moreover, the realization that we can understand essential life processes, such as the transmission of hereditary information, as chemical structures and their reactions has significant philosophical implications. What does it mean, biochemically, to be human? What are the biochemical differences between a human being, a chimpanzee, a mouse, and a fruit fly? Are we more similar than we are different?

Second, biochemistry is greatly influencing medicine and other fields. The molecular lesions causing sickle-cell anemia, cystic fibrosis, hemophilia, and many other genetic diseases have been elucidated at the biochemical level. Some of the molecular events that contribute to cancer development have been identified. An understanding of the underlying defects opens the door to the discovery of effective therapies. Biochemistry makes possible the rational design of new drugs, including specific inhibitors of enzymes required for the replication of viruses such as human immunodeficiency virus (HIV). Genetically engineered bacteria or other organisms can be used as "factories" to produce valuable proteins such as insulin and stimulators of blood-cell development. Biochemistry is also contributing richly to clinical diagnostics. For example, elevated levels of telltale enzymes in the blood reveal whether a patient has recently had a myocardial infarction (heart attack). DNA probes are coming into play in the precise diagnosis of inherited disorders, infectious diseases, and cancers. Agriculture, too, is benefiting from advances in biochemistry with the development of more effective, environmentally safer herbicides and pesticides and the creation of genetically engineered plants that are, for example, more resistant to insects. All of these endeavors are being accelerated by the advances in genomic sequencing.

Third, advances in biochemistry are enabling researchers to tackle some of the most exciting questions in biology and medicine. How does a fertilized egg give rise to cells as different as those in muscle, brain, and liver? How do the senses work? What are the molecular bases for mental disorders such as Alzheimer disease and schizophrenia? How does the immune system distinguish between self and nonself? What are the molecular mechanisms of short-term and long-term memory? The answers to such questions, which once seemed remote, have been partly uncovered and are likely to be more thoroughly revealed in the near future.

Because all living organisms on Earth are linked by a common origin, evolution provides a powerful organizing theme for biochemistry. This book is organized to emphasize the unifying principles revealed by evolutionary considerations. We begin in the next chapter with a brief tour along a plausible evolutionary path from the formation of some of the chemicals that we now associate with living organisms through the evolution of the processes essential for the development of complex, multicellular organisms. The remainder of Part I of the book more fully introduces the most important classes of biochemicals as well as catalysis and regulation. Part II, Transducing and Storing Energy, describes how energy from chemicals or from sunlight is converted into usable forms and how this conversion is regulated. As we will see, a small set of molecules such as adenosine triphosphate

(ATP) act as energy currencies that allow energy, however captured, to be utilized in a variety of biochemical processes. This part of the text examines the important pathways for the conversion of environmental energy into molecules such as ATP and uncovers many unifying principles. Part III, Synthesizing the Molecules of Life, illustrates the use of the molecules discussed in Part II to synthesize key molecular building blocks, such as the bases of DNA and amino acids, and then shows how these precursors are assembled into DNA, RNA, and proteins. In Parts II and III, we will highlight the relation between the reactions within each pathway and between those in different pathways so as to suggest how these individual reactions may have combined early in evolutionary history to produce the necessary molecules. From the student's perspective, the existence of features common to several pathways enables material mastered in one context to be readily applied to new contexts. Part IV, Responding to Environmental Changes, explores some of the mechanisms that cells and multicellular organisms have evolved to detect and respond to changes in the environment. The topics range from general mechanisms, common to all organisms, for regulating the expression of genes to the sensory systems used by human beings and other complex organisms. In many cases, we can now see how these elaborate systems evolved from pathways that existed earlier in evolutionary history. Many of the sections in Part IV link biochemistry with other fields such as cell biology, immunology, and neuroscience. We are now ready to begin our journey into biochemistry with events that took place more than 3 billion years ago.

APPENDIX: DEPICTING MOLECULAR STRUCTURES

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

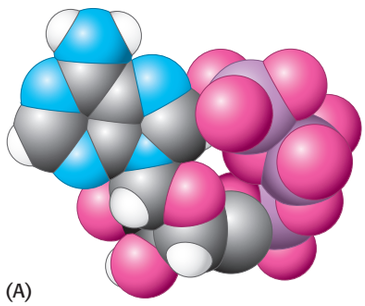
Stereochemical Renderings

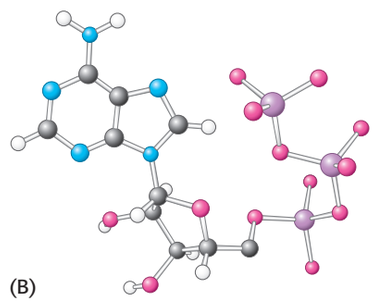
Most of the chemical formulas in this text are drawn to depict the geometric arrangement of atoms, crucial to chemical bonding and reactivity, as accurately as possible. For example, the carbon atom of methane is sp^3 hybridized and tetrahedral, with H–C–H angles of 109.5 degrees while the carbon atom in formaldehyde is sp^2 hybridized with bond angles of 120 degrees.

Methane

Formaldehyde

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


(A)


(B)

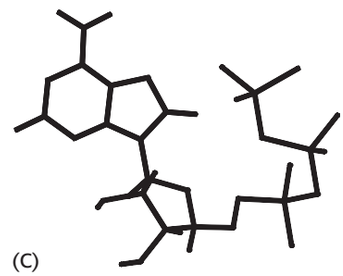
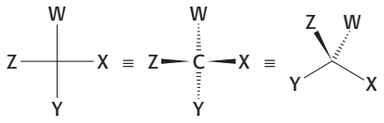

(C)

FIGURE 1.16 Molecular representations. Comparison of (A) space-filling, (B) ball-and-stick, and (C) skeletal models of ATP.

Fischer Projections

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



Fischer projection **Stereochemical rendering**

In a Fischer projection, the bonds to the central carbon are represented by horizontal and vertical lines from the substituent atoms to the carbon atom, which is assumed to be at the center of the cross. By convention, the horizontal bonds are assumed to project out of the page toward the viewer, whereas the vertical bonds are assumed to project into the page away from the viewer. Appendix xx at the back of the book is a structural glossary of the key molecules in biochemistry, presented both as stereochemically accurate structures and as Fisher projections.

For depicting molecular architecture in more detail, five types of models will be used: space filling, ball and stick, skeletal, ribbon, and surface representations (Figure 1.16). The first three types show structures at the atomic level.

1. *Space-filling models*. The space-filling models are the most realistic. The size and position of an atom in a space-filling model are determined by its bonding properties and van der Waals radius, or contact distance (Section 1.3.1). A van der Waals radius describes how closely two atoms can approach each other when they are not linked by a covalent bond. The colors of the model are set by convention.

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

Space-filling models of several simple molecules are shown in Figure 1.17.

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

3. *Skeletal models*. An even simpler image is achieved with a skeletal model, which shows only the molecular framework. In skeletal models, atoms are not shown explicitly. Rather, their positions are implied by the junctions and ends of bonds. Skeletal models are frequently used to depict larger, more complex structures.

As biochemistry has advanced, more attention has been focused on the structures of biological macromolecules and their complexes. These structures comprise thousands or even tens of thousands of atoms. Although these structures can be depicted at the atomic level, it is difficult to discern the relevant structural features because of the large number of atoms. Thus, more schematic representations—ribbon diagrams and surface representations—have been developed for the depiction of macromolecular structures in which atoms are not shown explicitly (Figure 1.18).

4. *Ribbon diagrams*. These diagrams are highly schematic and most commonly used to accent a few dramatic aspects of protein structure, such as the α helix (a coiled ribbon), the β strand (a broad arrow), and loops (simple lines), so as to provide simple and clear views of the folding patterns of proteins.

5. *Surface representations*. Often, the interactions between macromolecules take place exclusively at their surfaces. Surface representations have been developed to better visualize macromolecular surfaces. These representations display the overall shapes of macromolecules and can be shaded or colored to indicate particular features such as surface topography or the distribution of electric charges.

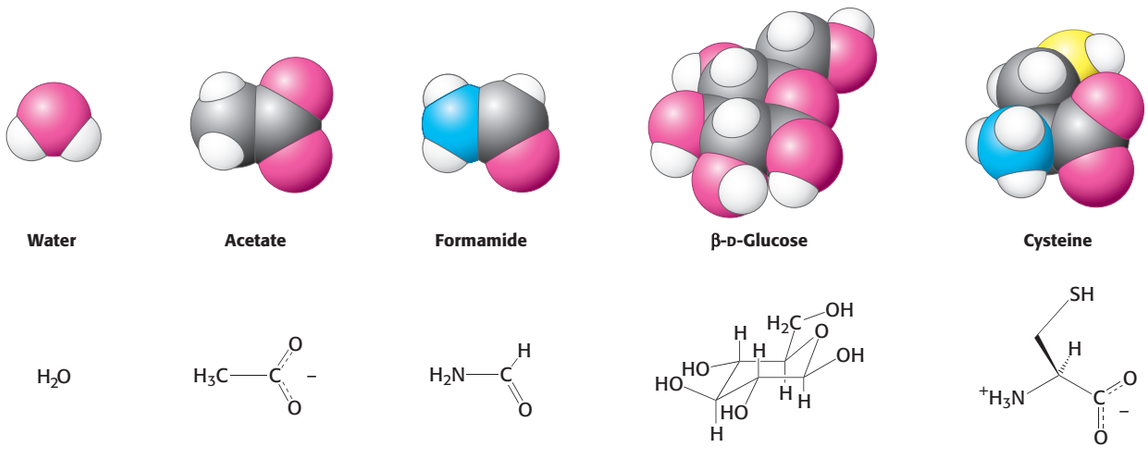


FIGURE 1.17 Space-filling models. Structural formulas and space-filling representations of selected molecules are shown.

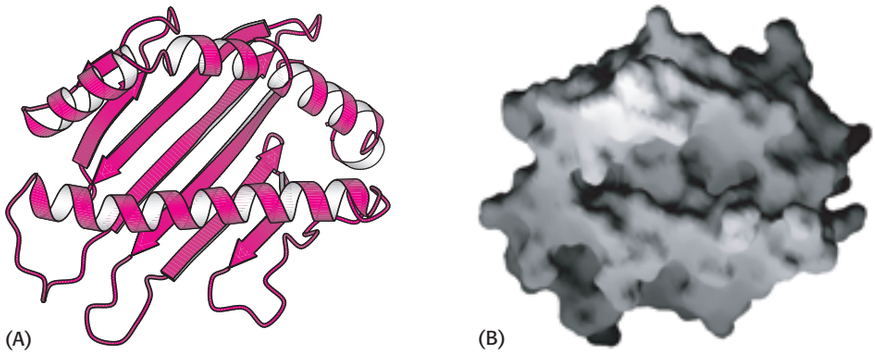


FIGURE 1.18 Alternative representations of protein structure. A ribbon diagram (A) and a surface representation (B) of a key protein from the immune system emphasize different aspects of structure.

KEY TERMS

deoxyribonucleic acid (DNA) (p. 4)
double helix (p. 5)
ribonucleic acid (RNA) (p. 6)
protein (p. 6)
amino acid (p. 6)
genetic code (p. 6)
Eukarya (p. 8)
Bacteria (p. 8)
Archaea (p. 8)
eukaryote (p. 8)
prokaryote (p. 8)
covalent bond (p. 8)
resonance structure (p. 9)
electrostatic interaction (p. 9)
hydrogen bond (p. 9)
van der Waals interaction (p. 9)
entropy (p. 12)
enthalpy (p. 13)
free energy (p. 13)
hydrophobic effect (p. 14)
sterochemistry (p. 16)
Fischer projection (p. 17)
space-filling model (p. 17)
ball-and stick-model (p. 17)
skeletal model (p. 17)
ribbon diagram (p. 17)
surface presentation (p. 17)