



5

Proteins & Amino Acids

Goal

To understand how common and dissimilar features of amino acids determine the chemical and physical properties of proteins.

Objectives

After this chapter, you should be able to

- explain why peptide bonds are polar and prefer the *trans* configuration.
- explain how side chains confer distinct chemical and physical properties on amino acids.
- draw a peptide of a given sequence at a specified pH.

Many of the most important macromolecules in living systems are polymers. These polymers are composed of small building blocks that are linked together in long, linear chains. Three of the most important biological polymers are **polysaccharides**, **polynucleotides**, and **polypeptides** (Figure 1). Polysaccharides, such as starch, are composed of sugar subunits whereas polynucleotides, such as DNA and RNA (the subject of Chapter 8), are built from nucleotides. Here and in the next chapter we focus on polypeptides. Polypeptides are chains of subunits called **amino acids** that are joined together by peptide bonds. Short polypeptides are called **peptides**, and long polypeptides are typically called **proteins**. Proteins are composed of 20 kinds of amino acids, which are at once alike and dissimilar. They share common features that allow them to form peptide bonds with each other while also exhibiting distinctive chemical features. This diversity of amino acids and the sheer number of possible combinations in their linear order allow for tremendous dissimilarity in the chemical and physical properties of proteins (Figure 2). Furthermore, proteins do not exist as unstructured chains. Rather, they fold in on themselves to form three-dimensional architectures with unique features.

Proteins have diverse functions

Owing to their enormous diversity, cells employ proteins to perform numerous tasks (Figure 3). Some proteins function as **enzymes**, which catalyze chemical reactions by reducing ΔG^\ddagger . Nearly all enzymes are proteins, and as we saw in the previous chapter, cells are able to carry

Figure 1 Polymers are macromolecules composed of small-molecule monomers linked together in chains

Carbohydrates, proteins, and nucleic acids are examples of biological polymers. Each class of molecule is made of monomer subunits covalently linked together in chains.

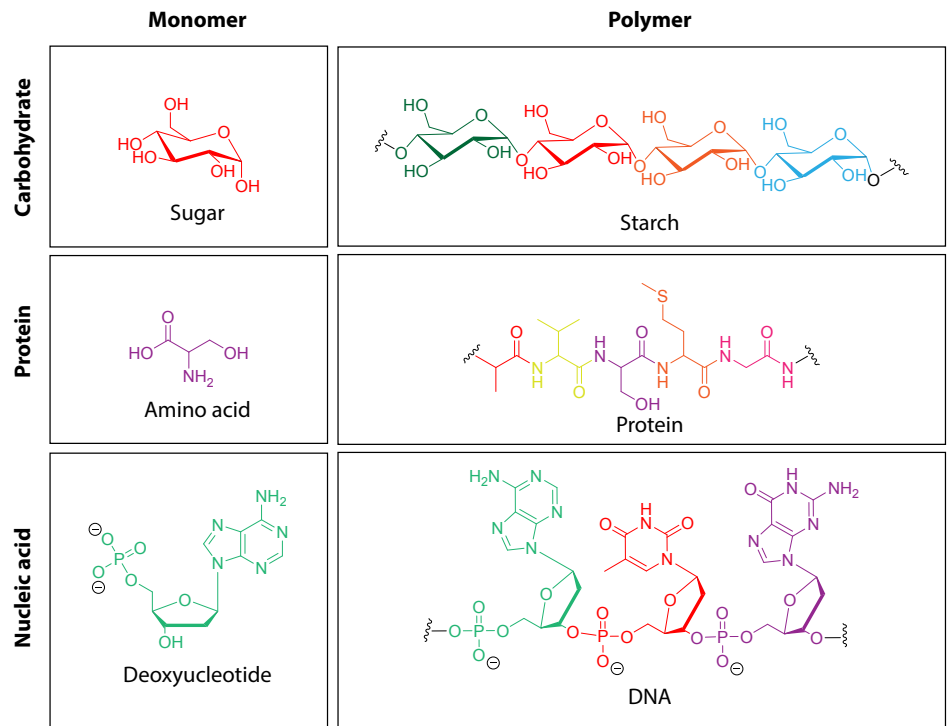
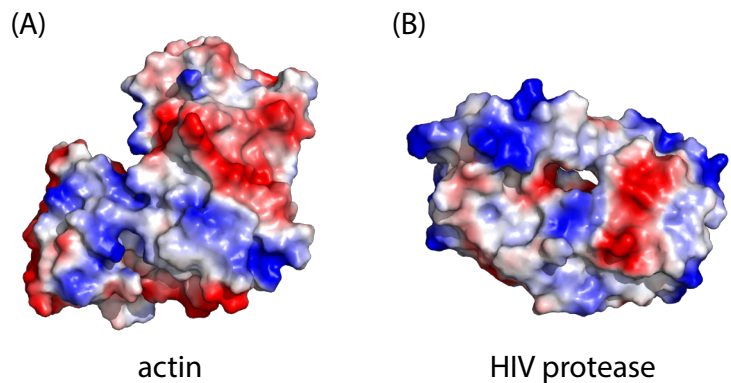


Figure 2 Proteins exhibit unique structures and chemical properties

Surface charge representation of the proteins actin (A) and HIV protease (B). Even though both proteins are chains of amino acids, they each feature distinct three-dimensional shapes with unique chemical properties, as evidenced by the unique distribution of surface charges on each molecule. Blue represents positive charge, red negative charge, and white neutral.



out controlled chemical reactions because they use enzymes to modulate reaction rates and couple favorable processes with unfavorable ones. In fact, nearly all transformations that occur in the cell are mediated by enzymes, and without them, living systems would carry out virtually no chemistry. Enzymes catalyze a wide variety of reactions and are often categorized according to the chemistry that they perform. Most enzymatic reactions involve either the transfer of electrons (oxidation and reduction reactions), the transfer of functional groups, the cleavage or formation of bonds, the rearrangement of bonds within individual molecules, or the use of ATP to covalently connect molecules. We will have more to say about how enzymes lower ΔG^\ddagger in subsequent chapters.

While enzymes come in many shapes and sizes and facilitate a vast number of specific chemical reactions, proteins as a whole are even more diverse. Not all proteins are enzymes; some proteins play structural roles. Hair is made of such proteins, as are fingernails and the outer layers of the skin. Many familiar materials, such as wool, silk, and leather are also made of protein. These structural proteins have evolved to withstand particular

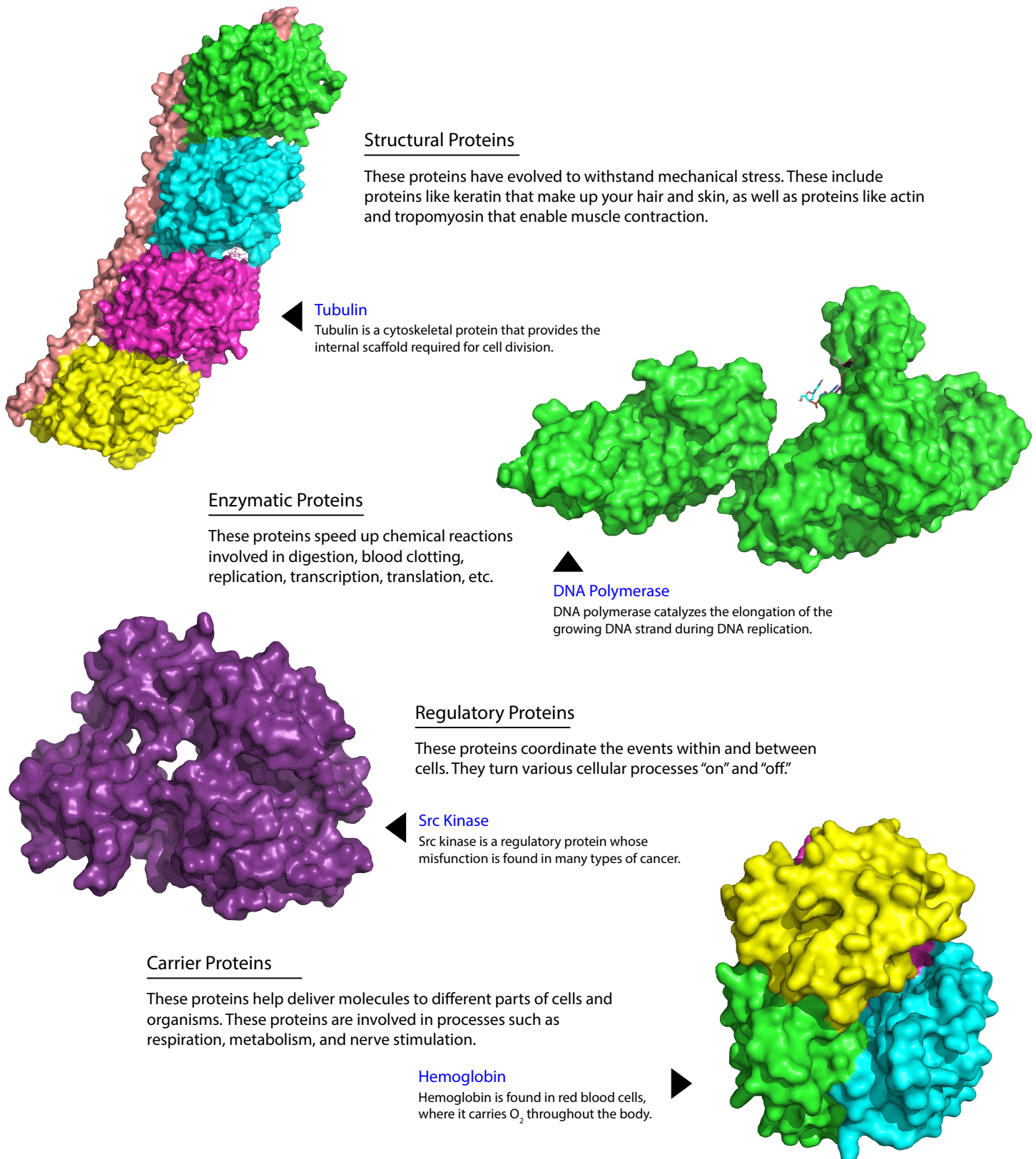


Figure 3 Proteins have a broad range of structures and functions

Example proteins are displayed using a surface representation. Proteins are colored by polypeptide chain.

mechanical stresses and afford protection to the organisms that produce them. On a cellular level, structural proteins contribute to the physical integrity of the cell and are responsible for much of the organization and compartmentalization found in living systems. For example, cytoskeletal

proteins like actin and tubulin are often used as scaffolds to position and control the movement and localization of other cellular molecules. Thus, even within the seemingly narrow category of “structural proteins,” we observe a wide range of structures, functions, and behaviors.

Proteins are also used to regulate a multitude of cellular processes. We have already pointed out that life depends on chemical reactions occurring on a rapid timescale; however, it is also necessary for these reactions to be precisely coordinated. Thus, the activities of various enzymes are regulated (turned on and off) by other proteins that coordinate cellular processes by responding to environmental conditions. In later chapters we will look at how proteins regulate when and where RNA is produced from DNA, and later we will examine a specific regulatory protein, Abl, whose malfunction results in cancer via deregulation of cell division.

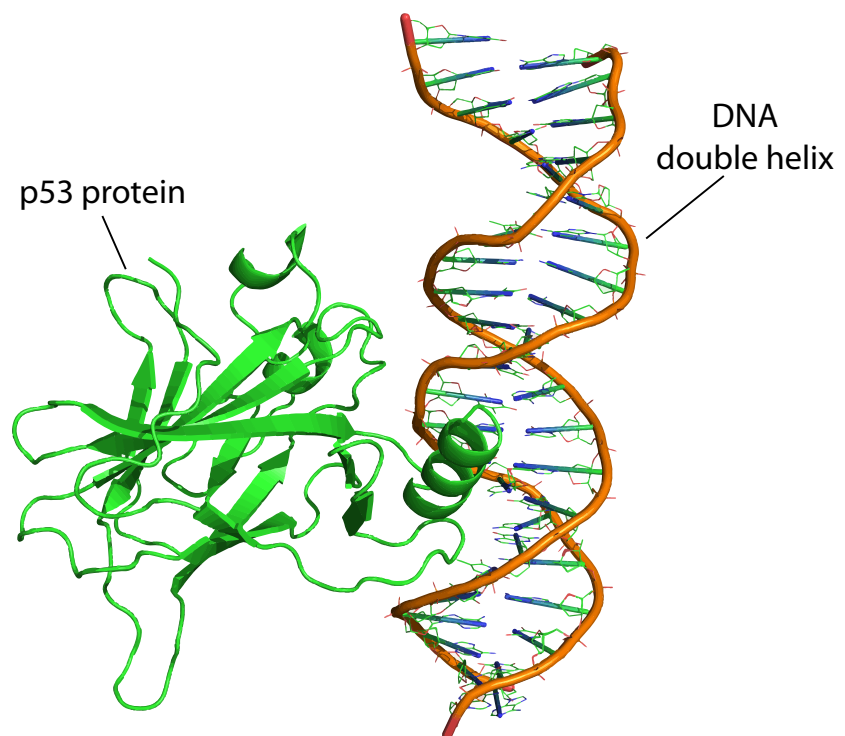
Some proteins act as carriers for other molecules. One example is hemoglobin, which carries oxygen gas to facilitate its transport through the circulatory system. Cholesterol and lipids are also carried throughout the bloodstream by protein carriers. Some proteins are not as easy to classify. For example, ion channels, which allow ions to pass through cell membranes and are critical to muscle contraction and neurotransmission, can be thought of as enzymes (because they catalyze the transport of ions across lipid bilayers) or as carriers (because they deliver ions from one side of the membrane to the other).

A protein's shape determines its function

Figure 4 shows the X-ray crystal structure of DNA with a protein bound to it, as we saw in Chapter 1. An X-ray crystal structure describes the near-exact position of each atom in the molecule in three-dimensional space. The technique for obtaining such a structure involves the use of X-rays (a form of electromagnetic radiation) and a crystal of the molecule in question. A

Figure 4 The structure of DNA is predictable, but proteins fold into a wide range of three-dimensional structures

Shown is an X-ray crystal structure of the p53 protein bound to a DNA double helix. DNA adopts a predictable double-helix structure, regardless of its sequence. The structures of proteins, however, vary widely depending on the specific sequence of amino acids of which they are composed.



beam of X-rays penetrates through the crystal lattice. Some of the X-rays are diffracted as they encounter the individual atoms that make up the molecules in the crystal lattice. The diffraction creates a characteristic pattern of reflections that is collected on an X-ray film and from which the arrangement of atoms can be reconstructed. This technique is a common method that scientists use to determine the structures of proteins and other molecules of life, and you will see many representations such as these throughout this book.

Figure 4 shows the iconic double-helical structure of DNA. DNA is almost always a double helix regardless of its nucleotide sequence; it essentially always has the same three-dimensional shape. In contrast, the protein in Figure 4 looks quite different from the proteins in Figures 2 and 3. This is because the three-dimensional structures of proteins are highly varied. The ability of different proteins to assume a wide range of distinct three-dimensional conformations is in large part why they are so versatile; a protein's shape determines its function. Conversely, DNA has just one function—genetic information storage—for which it assumes a single shape.

The sequence of amino acids in a protein determines its folded structure

The specific order of amino acids in a protein is known as its **primary structure**. It is this sequence that determines the three-dimensional architecture of a protein. A famous experiment that proves that all the information necessary for proper folding of a protein is contained in its primary structure is presented in the next chapter.

Amino acids share common structural features

All amino acids are composed of an **amino group** ($-\text{NH}_2$), a **carboxylic acid group** ($-\text{COOH}$), and an intervening carbon atom to which these two groups are connected. The intervening carbon atom is the **alpha** (α)

Box 1 Sickle-cell disease: one amino acid makes a difference

You might imagine that changing a single amino acid in a protein consisting of a hundred or more amino acids would have little effect on the protein's folded shape or function. Often this is the case, but sometimes a single change makes a profound difference. Figure 5 shows one example in which the substitution of a single amino acid significantly alters the protein. Hemoglobin, the oxygen carrier in red blood cells, contains the amino acid glutamate at position 6 in the primary sequence. Hemoglobin typically folds into a globular (i.e., roughly spherical) shape that associates to form a tetramer (a complex consisting of four molecules of hemoglobin). People who have sickle-cell anemia have a single amino acid substitution in their hemoglobin. The substitution is a switch from this glutamate to valine. This alteration causes hemoglobin molecules to clump, as reflected in a change in the shape of the red blood cells, which have a distorted, sickle shape. The sickle-shaped cells do not carry as much oxygen and therefore deliver less oxygen to the body's tissues. These cells are also fragile and can break, causing painful "crises" because they disrupt blood flow. The sickle-cell mutation is recessive, but a single copy of the mutant allele enables people to resist infection by the malaria-causing pathogen *Plasmodium*, which propagates in red blood cells. This protection against *Plasmodium* explains why the allele is common in areas where malaria is endemic.

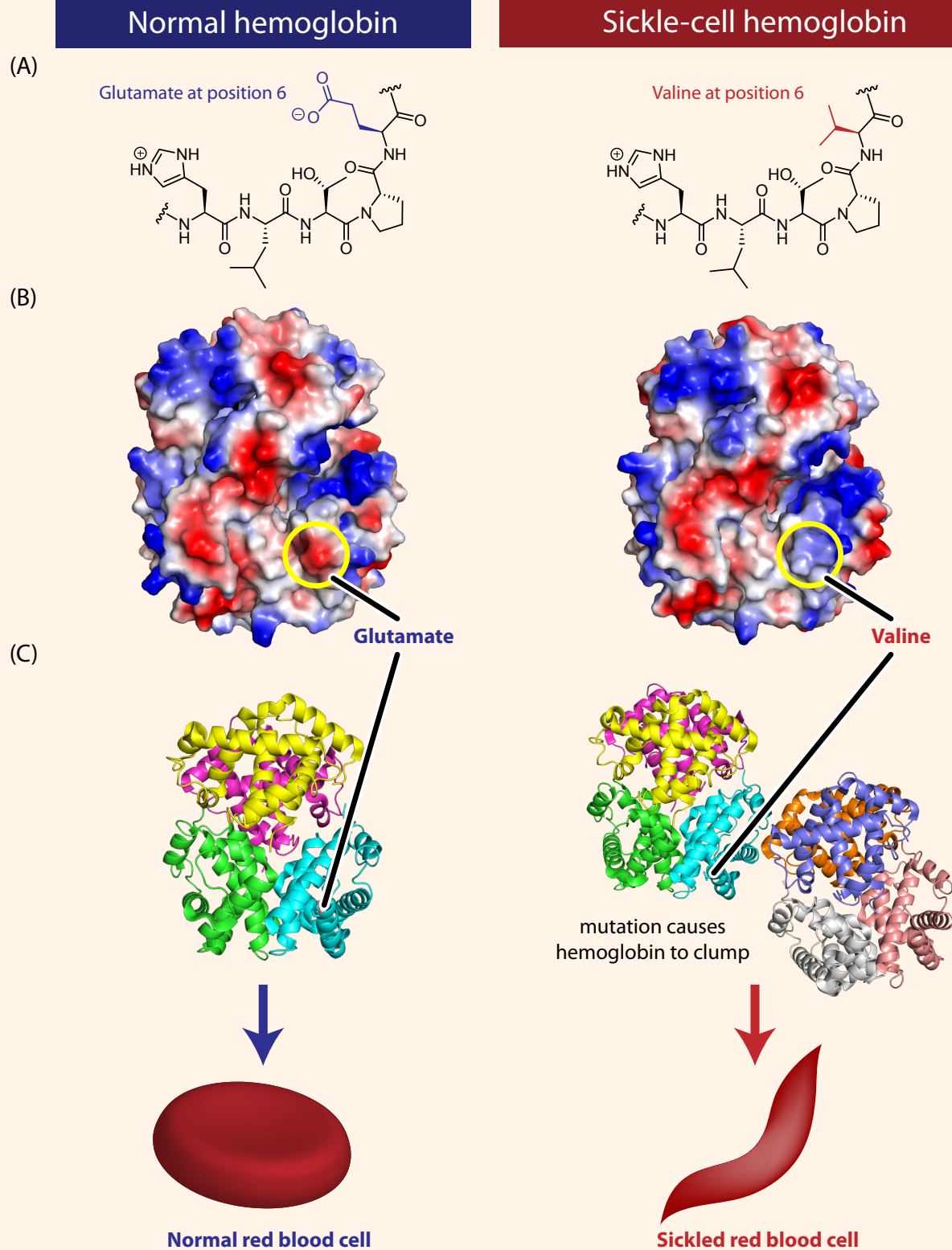
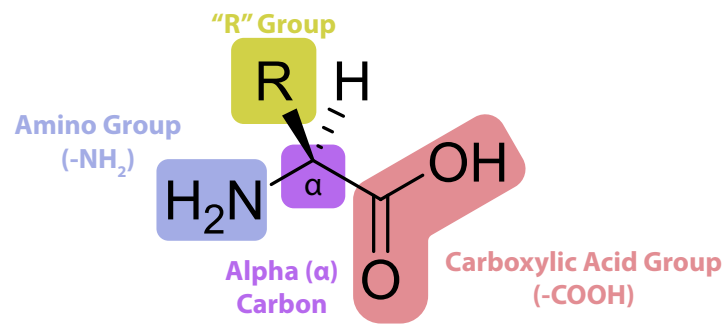


Figure 5 Sickle-cell disease is caused by a single amino acid change in the hemoglobin protein

(A) Line drawings of a portion of the hemoglobin (*left*) and sickle-cell hemoglobin (*right*) proteins. Normal hemoglobin contains the amino acid glutamate at position 6 in the primary sequence. In individuals with sickle-cell disease, this glutamate is replaced with the amino acid valine. (B) Computer-generated structure showing the charges present on the surfaces of hemoglobin (*left*) and sickle-cell hemoglobin (*right*). As we will see shortly in this chapter, some amino acids can be negatively or positively charged, while many others are neutral. In the figure, blue represents positive charge, red represents negative charge, and white represents neutral atoms. The substitution of the negatively charged amino acid glutamate at position 6 with the neutral valine removes a negative charge that is normally present in hemoglobin. (C) Computer-generated models showing the structures of the hemoglobin (*left*) and sickle-cell hemoglobin (*right*) proteins. The substitution of glutamate with valine causes hemoglobin tetramers to clump together.

Figure 6 Amino acids have four component parts



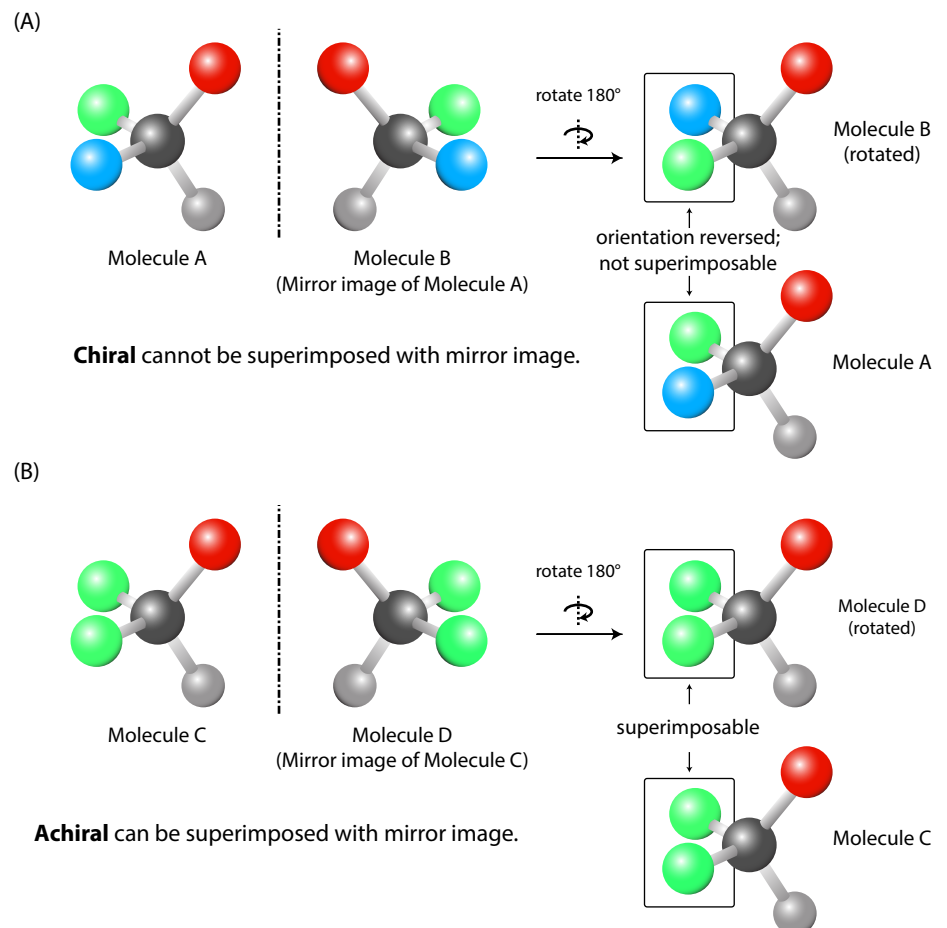
carbon. For 19 of the 20 amino acids, an additional chemical group, known as an **R group** or **side chain**, is attached to the α carbon. The twentieth amino acid, glycine, has two hydrogen atoms connected to the α carbon instead of an R group and a single hydrogen atom. The unique chemical and structural properties of each amino acid are determined by the identity of the R group.

Amino acids are chiral

A key feature of amino acids is that the α carbon is **chiral**. When a carbon atom is bound to four unique groups, it creates a **chiral center** (also known as a **stereocenter**). Simply put, a chiral molecule is one that cannot be superimposed with its mirror image. This is represented in Figure 7A by a cartoon in which four colored spheres represent the four different substituents of the α carbon of one of the 19 amino acids that have an R

Figure 7 Chiral molecules cannot be superimposed with their mirror images

(A) Atoms that are bound to four unique substituents cannot be superimposed with their mirror images and are therefore chiral. (B) Atoms that are bound to fewer than four unique substituents can be superimposed with their mirror images and are therefore achiral.



substituent (molecule A). You can see that the mirror image of molecule A (molecule B) is chiral because it cannot be rotated to match molecule A. Figure 7B is a cartoon of the substituents attached to the α carbon of glycine. Glycine is an achiral molecule for which two of the substituents are the same (hydrogen atoms), as represented by the green spheres in molecule C. You can see that rotating the mirrored molecule (molecule D) makes it indistinguishable from molecule C.

The configuration, or **stereochemistry**, of a chiral center is binary; a chiral center can either have one configuration or the mirror image of that configuration. To describe stereochemistry using standard line drawings, we use two additional types of lines: a solid, wedge-shaped line representing bonds that extend outward from the page and a dashed, wedge-shaped line representing bonds that extend into the page. By adding a third dimension of depth to standard line drawings, we can specify the stereochemistry of any molecule.

Although chirality may seem like an abstract concept, it is not. You are surrounded by chiral objects; many macroscopic structures and most molecules in your body, large and small, are chiral. For example, your feet are chiral, which is why you cannot wear your left shoe on your right foot. Your hands are chiral, which is why if you are left-handed it is hard to use most scissors, which are designed for right-handed people simply because the majority of people happen to be right-handed. Chirality in molecules can have as profound of an effect on function as chirality in hands or feet. Take for example the two small molecules shown in Figure 8. Both molecules have the same number and type of atoms and the same bond connections, and both are called carvone. Carvone has one chiral center, however, so the two molecules are actually different. The molecule on the left is the dominant odorant in spearmint, whereas the molecule on the right is the dominant odorant in caraway, the seed used in rye bread and Swedish cookies. No one would confuse the smells of caraway and spearmint. The reason these molecules smell so different is that the receptors to which they bind in the nose are themselves chiral. Just as your left shoe fits differently to your right foot than to your left foot, the two forms of carvone (called **stereoisomers**) bind differently to the chiral receptors in the nose. In other words, they may look similar on paper, but in three dimensions, their shapes are fundamentally different.

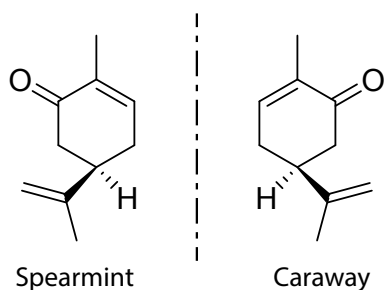


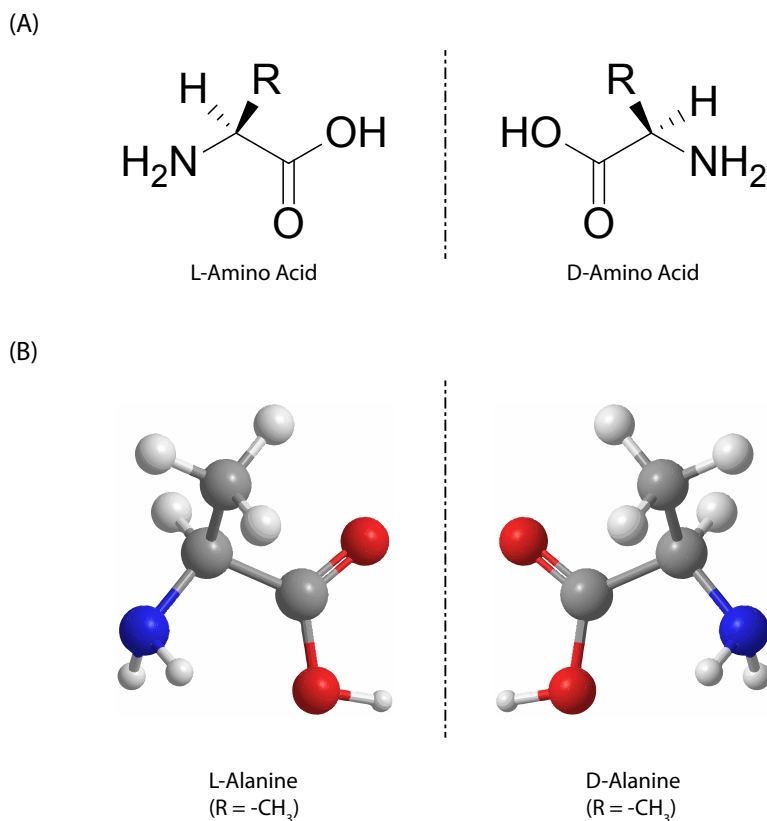
Figure 8 Chirality can strongly affect molecular properties

The two stereoisomers of carvone exhibit distinct properties. The stereoisomer of carvone on the left is responsible for the scent of spearmint, whereas the stereoisomer on the right is responsible for the scent of caraway. These molecules are mirror images of each other (reflected across the dashed line) that differ only in the configuration of one stereocenter, but our noses sense them quite differently.

The two stereoisomers for each of the 19 chiral amino acids are denoted as **D** and **L** (Figure 9). Only the L-stereoisomer is used in nature to construct proteins. While this may seem arbitrary, remember that stereoisomers are differently shaped, distinct structures. For example, imagine a hundred amino acid-long protein in which each amino acid could be either D or L. If so, then there would be 2^{100} possible versions of the same protein. All of these proteins would have wildly different structures that would fold in unpredictable ways. No such problem exists in living systems, which use only the L-stereoisomer, and as such, proteins always fold into the same predictable shapes. The use of a single stereoisomer is not restricted to proteins. All biological molecules with carbon atoms attached to four different groups, such as sugars and nucleotides, are restricted to a particular stereoisomer.

Figure 9 Amino acids are chiral

There are two stereoisomers of each amino acid, but only the L-stereoisomer is used in nature to construct proteins. (A) Line drawing of a generic L-amino acid with its D-stereoisomer. (B) Three-dimensional model of L-alanine with its mirror image, D-alanine, in which differently colored spheres represent different elements (white for hydrogen, blue for nitrogen, red for oxygen, and grey for carbon).



Amino acids in proteins are connected by peptide bonds

As we have learned, proteins are long chains of covalently connected amino acids. The covalent linkage between two amino acids is called a **peptide bond**, hence the name polypeptide. The peptide bond may look familiar, as it is an example of the amide functional group that we saw in Chapter 2. Peptide bonds connect the nitrogen atom of the amino group of one amino acid to the carbon atom of the carboxylic acid of another amino acid. This is equivalent to the removal of a water molecule from the two amino acids, that is, a hydroxyl group from the carboxylic acid and a hydrogen atom from the amino group (Figure 10A). However, and as we shall see in Chapter 11, peptide bonds in living systems are not formed by the direct removal of water molecules (because this would be energetically unfavorable) but rather through a series of intervening biochemical steps.

Polypeptide chains are directional, and the two ends of the chain feature different chemical groups. At one end is a free amino group, and at the other end is a free carboxylic acid group (Figure 10). The end of the polypeptide chain with a free amino group is the **amino terminus**, which is frequently written as “N-terminus” or “NH₃⁺-terminus.” The end with a free carboxylic acid group is the **carboxy terminus**, which is frequently written as “C-terminus” or “COO⁻-terminus.” When the amino acid sequence of a protein is specified, it *must* indicate the N-terminal-to-C-terminal directionality; reversing the directionality indicates a very different protein sequence! By convention, amino acid sequences are typically written left to right from amino to carboxy terminus.

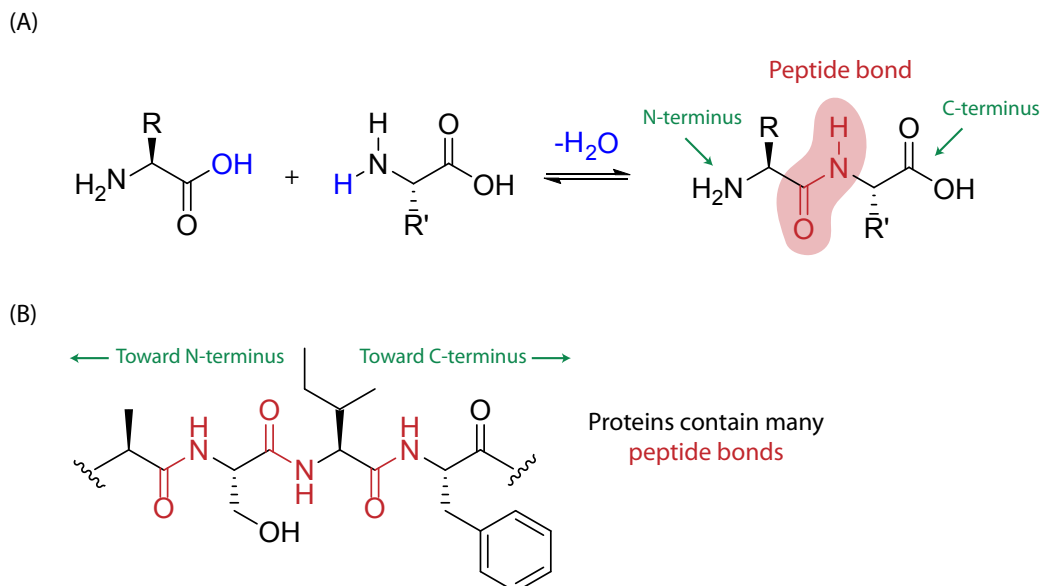


Figure 10 Amino acids are connected by peptide bonds to form polypeptide chains

(A) Amino acids are connected by peptide bonds in proteins. The amino and carboxylic acid groups of any two amino acids can be covalently connected by a peptide bond, with the equivalent of the removal of a water molecule (shown in blue). The resulting amino acid chain has two ends, one with a free amino group, the N-terminus, and one with a free carboxylic acid group, the C-terminus. (B) Proteins are often made from very long polypeptide chains that typically contain hundreds of peptide bonds. Only four amino acid subunits of a protein are shown, as depicted by the squiggly lines (“spinach”) at each end.

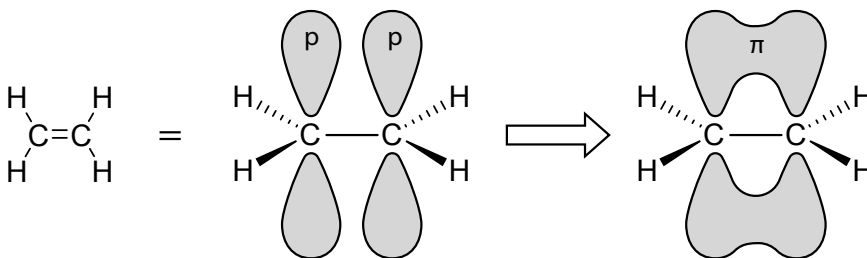
Peptide bonds are flat, polar, and not free to rotate

The chemical properties of the peptide bond are important for determining the shape of the polypeptide chain. Some of the bonds of the peptide chain are free to rotate, but the peptide bond itself can only adopt certain conformations. To understand why, we return to the topic of the orbitals that surround atoms. When a covalent bond forms, the orbitals of the two bonded atoms overlap and mix, creating a **bonding orbital** that holds the shared pair of electrons. Bonding orbitals are a combination of atomic orbitals from each atom, and much like atomic orbitals, they describe where the shared electrons are likely to be found. A single bond [which we sometimes call a **sigma (σ) bond**] is composed of a cylindrically symmetrical orbital that joins two nuclei. A double bond consists of one σ bond as well as a second bond called a **pi (π) bond**. The π bond forms when the p orbitals of two nuclei overlap.

Bonding orbitals can only hold two electrons. That means that single, double, and triple bonds must consist of one, two, and three bonding orbitals, respectively. In fact, single bonds always consist of one σ -orbital, double bonds always consist of one σ -orbital and one π -orbital, and triple bonds always consist of one σ -orbital and two π -orbitals. It turns out that this has an important consequence; bonds that contain π -orbitals cannot rotate freely. Sigma bonding orbitals have a cylindrical shape, and as such, single bonds can rotate freely without disrupting the orbital overlap that creates the bond. Conversely, the p orbitals that compose π -orbitals must be parallel with one another to overlap. If the π -bond were to rotate, the bond would break. Since breaking a covalent bond requires energy, the rotation of a π -bond has a high barrier. Because double bonds contain π -orbitals, they are not free to rotate. Shown in Figure 11 is the example of ethylene

Figure 11 Double bonds cannot rotate because they contain π -orbitals

Shown is the π -bonding orbital in ethylene (C_2H_4), which is formed by two overlapping p orbitals. The ethylene molecule is flat, and all of its atoms lie in the same plane because its π -bond can only exist when the p orbitals overlap.



(C_2H_4), which contains a carbon-carbon double bond. Because this double bond contains a π -bonding orbital, the bond cannot rotate. Notice that the hydrogen and carbon atoms in ethylene are arranged in the same plane as one another, allowing the p orbitals to overlap and form the π -bonding orbital (see Box 2).

We are still left with the question of why the peptide bond does not rotate. After all, we have drawn the peptide bond as a carbon-nitrogen single bond, so it should rotate freely. It turns out, however, that the peptide bond does not behave like a single bond because it has some double-bond character. The standard line drawing that we have shown for the peptide bond is somewhat misleading, as it fails to describe the bond's properties in their

Box 2 Hybrid orbitals

As we have seen, ethylene is unable to rotate around its double bond due to its π bond. Another feature of ethylene is that its hydrogen and carbon atoms are in a planar configuration; they have a trigonal planar geometry. This means that the atomic orbitals linking the hydrogen and carbon atoms are pointing to the vertices of a triangle. The significance of this trigonal planar geometry is that it keeps the electrons in the three bonds to carbon maximally separated. How can σ - and π -bonds achieve this geometry? As you will remember, the s orbital is spherical and does not point in any specific direction. The three p orbitals are dumbbell shaped and are directed orthogonally (at right angles) to each other. So it seems that no combination of these four orbitals could allow carbon to bond to three other atoms in a trigonal planar fashion. Therefore, scientists came up with a mathematical manipulation known as hybridization that allows these orbitals to “mix” to produce an equal number of hybrid orbitals. In the case of the ethylene carbons, the s and two p orbitals mix to form three sp^2 orbitals, which point at the three vertices of a triangle and allow bonding to two hydrogen atoms and one carbon atom, as shown below.

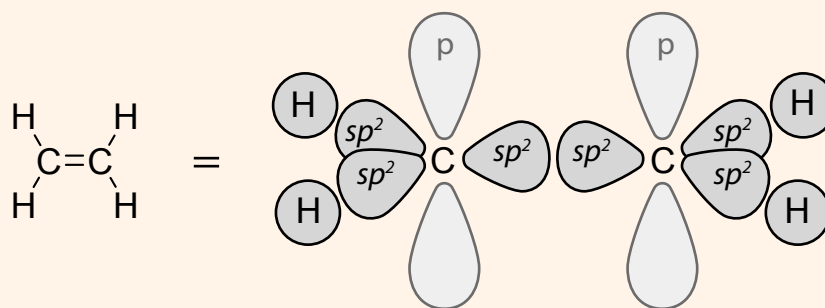
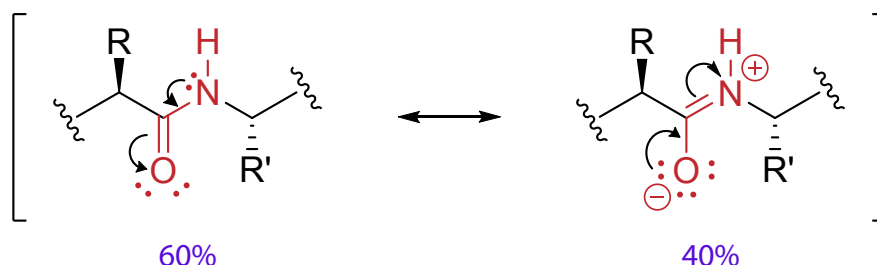


Figure 12 Resonance stabilization causes the peptide bond to have double-bond character

The double-headed arrow signifies that the peptide bond is a hybrid of these two states.



entirety. The peptide bond behaves like ethylene in that the atoms attached to nitrogen and carbon are in the same plane and rotation is restricted. To understand this behavior, consider the bonding between nitrogen and the carbonyl carbon of the peptide bond. The nitrogen has a tendency to share its lone pair of electrons with the carbonyl carbon, delocalizing electrons among the nitrogen, carbon and oxygen atoms and creating a lower-energy state (Figure 12). Bonding in which electrons are delocalized and distributed among multiple atoms is referred to as **resonance stabilization**. Resonance stabilization does not mean that the peptide bond oscillates between the structures shown in Figure 12; rather, it exists as a hybrid of these two extremes. The distribution of electrons in this hybrid state is not, however, uniform. Rather, it is biased such that the single bond-like character of the carbon-nitrogen bond is slightly favored (60%) over its double bond-like character. Because the C-N bond has some double-bond character, the peptide bond is flat, meaning that the nitrogen and carbonyl carbon atoms both have trigonal planar geometries.

Another consequence of resonance stabilization is that the peptide bond is polarized. As a result of resonance in this hybrid structure, the oxygen carries a partial-negative charge and the nitrogen a partial-positive charge. Because of the charge separation between oxygen and nitrogen, the peptide bond is polar. As we will see later, the partial-positive charge of the nitrogen and hydrogen and the partial-negative charge of the oxygen are important properties that influence how proteins fold.

The **backbone** of polypeptide chains is not simply a continuous series of peptide bonds. Rather, it also consists of bonds between the amide nitrogens and α carbons and the bonds between the α carbons and carbonyl carbons. As we now explain, the peptide bonds within this framework are not free to rotate, whereas the other two bonds are.

As in the example of ethylene from Figure 11, the double-bond character of the peptide bond causes it to be flat. All of the atoms immediately bound to the peptide bond are forced to be coplanar. The six coplanar atoms are the carbonyl carbon, the carbonyl oxygen, the peptide bond nitrogen, the hydrogen bound to that nitrogen, and the two adjacent α carbons. Each set of six coplanar atoms is highlighted as colored rectangles in Figure 13. Each rectangle is free to rotate relative to the adjacent coplanar sets of atoms. Notice that the α carbon is coplanar with two different (and differently colored) adjacent blocks of atoms. The requirement that sets of six atoms are coplanar rigidifies the polypeptide backbone and constrains it into a specific set of conformations, thereby influencing how the overall protein will fold.

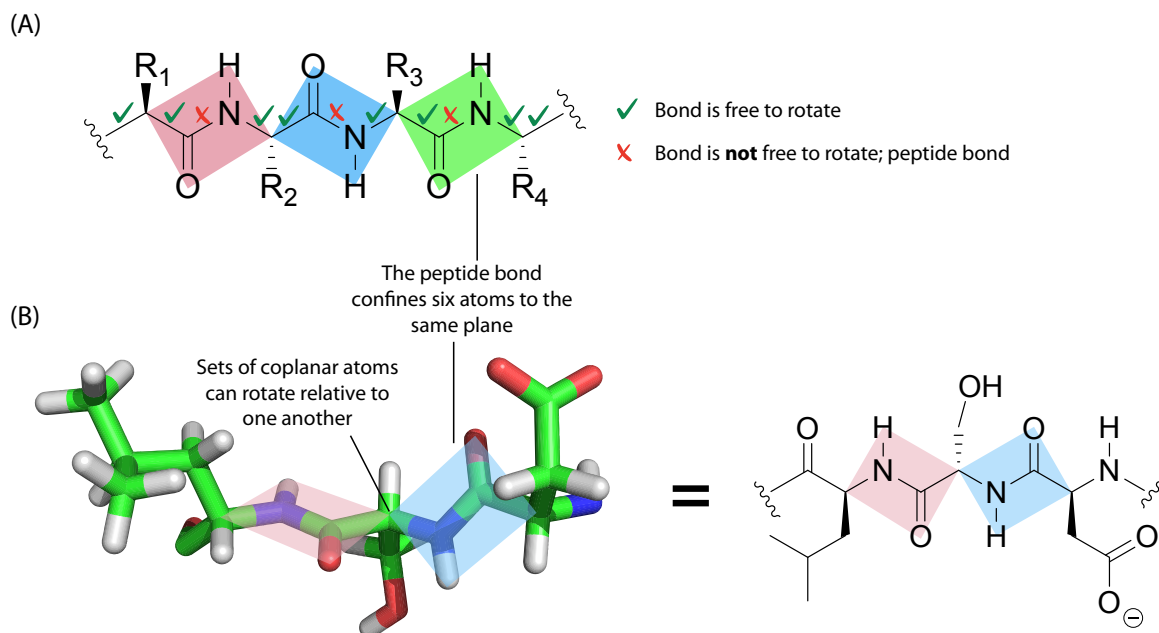


Figure 13 The peptide bond is flat and not free to rotate

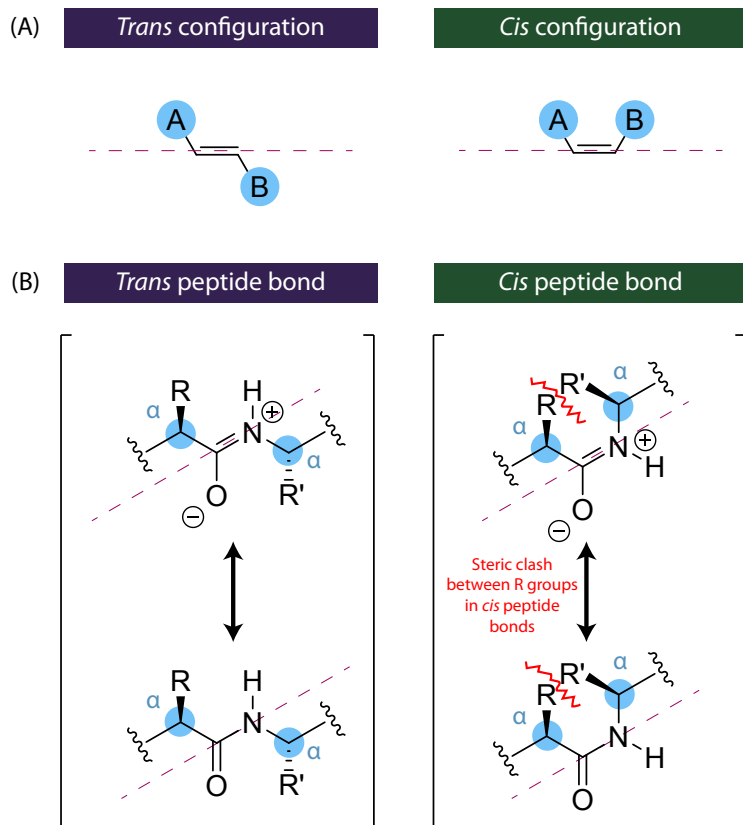
(A) Diagram of a generic polypeptide chain. Colored rectangles indicate sets of six atoms that are coplanar due to the double-bond character of the peptide bond. Green checks indicate bonds in the peptide backbone that are free to rotate, and red Xs indicate bonds that are not free to rotate. Note that only peptide backbone bonds are labeled. (B) A computer-generated model of a short peptide is shown on the left, and a standard line drawing of the same structure is shown on the right. Notice that the rotation of bonds other than the peptide bond in the peptide backbone enables sets of coplanar atoms to rotate relative to one another.

Peptide bonds favor the *trans* configuration

Because the peptide bond is not able to rotate freely, it can exist in two possible configurations known as **geometric isomers**. To understand the concept of geometric isomers, let us first consider the simple case of the hypothetical molecule shown in Figure 16A. It can exist as two geometric isomers referred to as *cis* and *trans*, which describe how two atoms or groups of atoms on opposite sides of a double bond are oriented relative to one another. Two atoms or groups of atoms are *cis* to each other when they are on the same side of an imaginary line drawn along the length of the double bond. If they are on opposite sides of that imaginary line, they are *trans* to each other. Likewise, peptide bonds can exist as *cis* or *trans* isomers depending on whether the α carbons of adjacent amino acids are on the same side of the peptide bond (Figure 14, right panel) or are on opposite sides (Figure 14, left panel) across the peptide bond connecting them. In most cases, the *trans* isomer is favored over the *cis* isomer. The reason for this is that the *cis* isomer places the R groups of the two connected amino acids in close proximity to each other. This results in a steric clash, a non-bonded interaction in which electrons involved in a bond get too close to electrons involved in an adjacent bond. This concept can be understood intuitively as the electron clouds of two chemical groups trying to occupy the same space at the same time and bumping into one another. As you might imagine, steric clashes are not favorable. In the *trans* isomer of the peptide bond, the steric clash is avoided, and the R groups are held far apart. The extent to which the *trans* configuration is favored over the *cis* configuration is dependent on the chemical identity of the R groups, as we will see shortly.

Figure 14 The *trans* configuration of the peptide bond is favored

Shown are geometric isomers of a hypothetical molecule with a double bond (A) and of the peptide bond (B). Shown in B are both extreme resonance states for the *cis* and *trans* configurations of the peptide bond.



The 20 amino acids have diverse properties

All 20 amino acids have distinct chemical and physical properties that are determined by their R groups (or lack thereof in the case of glycine) (Figure 15). Despite this diversity, amino acids can be grouped according to their size, charge, polarity, and, in certain cases, by unusual conformational features they impart on the polypeptide backbone. Because of their central importance in biology, you should learn the structures, the three-letter abbreviations, and the one-letter codes for all 20 amino acids. They are part of the language of life.

Some amino acids can be categorized as being polar or nonpolar. Nonpolar amino acids typically have side chains that lack polar bonds. Nonpolar amino acids generally have hydrocarbon side chains (e.g., isoleucine) that are built from carbon and hydrogen, which have a small electronegative difference. Therefore, the bonds in these hydrocarbon side chains are not polarized. One nonpolar amino acid, methionine, also contains sulfur, which has the same electronegativity as carbon. The nonpolar amino acids are **hydrophobic**, as they tend to segregate away from water for reasons that we will explain later. In contrast, the polar amino acids are those with polar bonds in their side chains. Polar amino acids are **hydrophilic**, meaning that their side chains interact strongly with water. An interesting case is cysteine, which you might think should be categorized as nonpolar. However, the S-H bond in cysteine has a tendency to ionize at neutral pH as we will discuss below and hence to some extent behaves like a polar side chain. Similarly, tryptophan and tyrosine cannot be easily categorized as hydrophobic or hydrophilic; each has a large side chain with polar and nonpolar features.

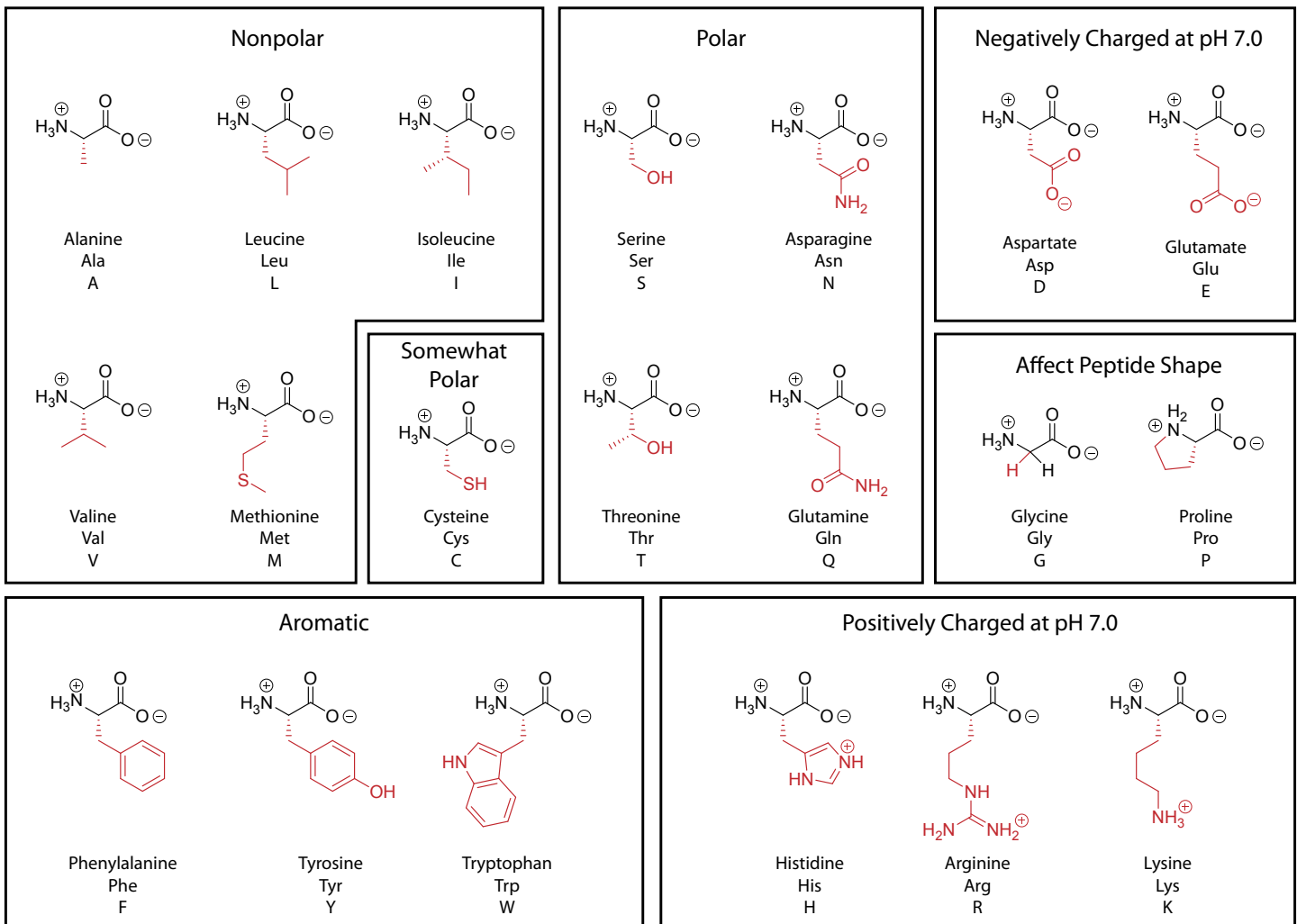


Figure 15 The 20 amino acids can be grouped by their properties





















Shown are the twenty amino acids as they predominantly exist at pH 7. The side chain of each amino acid is shown in red, whereas the amino group, carboxyl group, and α carbon are shown in black

Some amino acids have ionizable side chains

Some amino acids **ionize** (i.e., carry a full charge) at certain pH ranges (Figure 16). Aspartic acid and glutamic acid are acidic. They have side chains with pK_a values well below physiological pH, and as a result, both amino acids exist predominantly in their deprotonated, negatively charged states under physiological conditions. Aspartic acid and glutamic acid go by a different set of names when they are deprotonated; when deprotonated, we call them aspartate and glutamate. You will often hear these names used interchangeably. Cysteine can also be deprotonated, as its pK_a is about 8.0–9.0. Although the neutrally charged, protonated form of cysteine is favored at physiological pH, its low pK_a means that a significant amount of the deprotonated species is also present. Other amino acids, such as arginine and lysine, are basic, and at physiological pH they exist almost exclusively in their protonated, positively charged states. Histidine is also basic, but the pK_a of its conjugate acid is close to physiological pH; as a result, it exists at physiological pH as a mixture of its positively charged, protonated form and its neutrally charged, deprotonated form. Since histidine's pK_a is about 6.0–7.0, the neutrally charged state is slightly more abundant at physiological

Figure 16 Amino acids contain ionizable functional groups

The top portion of the table lists amino acids with ionizable side chains. The ionization state that predominantly exists at physiological pH is highlighted. The highlighted structures are highlighted in red if the indicated ionization form is at least 100 times more abundant than the other form and in purple if not. The bottom portion of the table describes the ionization of the free amino and carboxy groups at the termini of the polypeptide chain.

Amino Acid	Acid		Conjugate Base	pK _a
Aspartic Acid/Aspartate Asp D		⇌		3.9-4.0
Glutamic Acid/Glutamate Glu E		⇌		4.0-4.5
Histidine His H		⇌		6.0-7.0
Cysteine Cys C		⇌		8.0-9.0
Tyrosine Tyr Y		⇌		10.0-10.3
Lysine Lys K		⇌		10.4-11.1
Arginine Arg R		⇌		12.5
Serine Ser S		⇌		13.0
Terminus	Acid		Conjugate Base	pK _a
N-terminus		⇌		9.6
C-terminus		⇌		2.5

pH. Because its pK_a is so close to physiological pH, histidine is often used by enzymes to transfer protons during chemical reactions. Other amino acids, such as tyrosine and serine, can be deprotonated at high pH, but at physiological pH they largely exist in their protonated, neutrally charged states.

In addition to side chains, the N- and C-termini of the polypeptide chain are ionized at physiological pH. The amino group at the N-terminus is basic, and its conjugate acid has a pK_a of about 9.6. As a result, the amino terminus is predominantly protonated at physiological pH. The carboxyl group at the C-terminus is acidic and has a pK_a of about 2.5. As a result, the carboxy terminus is predominantly deprotonated at physiological pH. The amide nitrogen in the peptide bond *cannot* be protonated in water due to the delocalization of its lone pair.

Phenylalanine, tyrosine, and tryptophan contain aromatic rings

The aromatic amino acids—phenylalanine, tyrosine, and tryptophan—are unique in that they contain rings with alternating double bonds. Actually, the double bonds in these rings are not fixed. Rather, the electrons are delocalized among all the carbon atoms due to resonance. Aromaticity imparts properties that are exploited both in nature and in the laboratory. For example, they are important in the function of certain enzymes in which they serve as conduits that allow the transfer of electrons from one aromatic side chain to another. And in the laboratory, they are useful because they readily absorb ultraviolet light, facilitating measurements of the concentrations of proteins in solution.

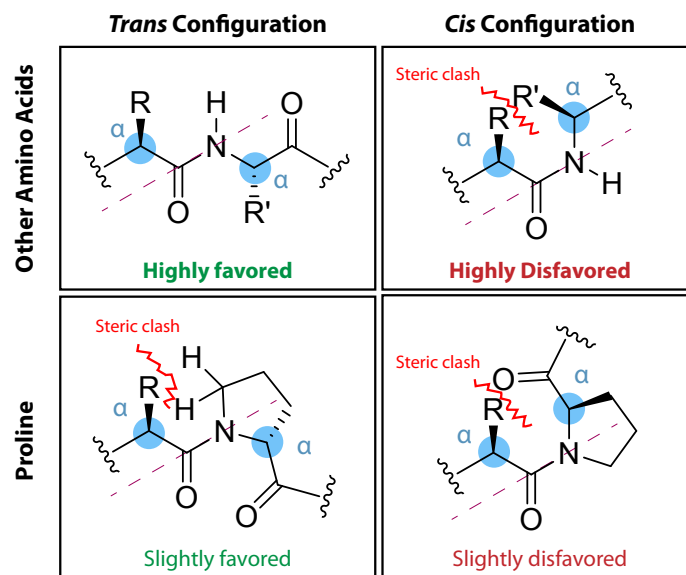
Amino acid side chains affect the conformation of the peptide backbone

As we mentioned earlier, the side chains in a polypeptide chain determine how it folds into a three-dimensional structure. This occurs in part because the amino acid side chains differ in their size and flexibility, which in turn biases the peptide backbone toward certain conformations. Some amino acids have large, bulky side chains (e.g., phenylalanine and tryptophan), whereas others take up much less space (e.g., alanine and glycine). The side chains of some amino acids are also more flexible than others. Methionine, for example, can adopt many more conformations than isoleucine or valine, both of which have a methyl ($-\text{CH}_3$) group on the side chain close to the peptide backbone (on the beta carbon, the second carbon from the carbonyl carbon). This methyl group restricts the conformations available to the peptide backbone because it creates steric clashes when certain conformations are adopted.

Glycine and proline have even more-pronounced effects on the peptide backbone. Glycine effectively lacks a side chain, and as such, it enables the peptide backbone to adopt a broader set of conformations around the two non-peptidic bonds in the repeating unit of the peptide backbone. Proline's side chain, which is connected to its amino group, prevents rotation around the bond connecting the amine and α carbon

Figure 17 Proline forms peptide bonds that can adopt either *cis* or *trans* configurations

Most amino acids form peptide bonds that strongly favor the *trans* configuration because of steric clashes in the *cis* configuration. Proline is an exception because its side chain causes steric clashes in both the *trans* and *cis* configurations. As a consequence, peptide bonds in which the nitrogen of the peptide bond comes from proline sometimes exist in the *cis* configuration.



atom (Figure 17). This is unique, as the amine- α carbon bond is free to rotate in every other amino acid. Furthermore, proline is the only amino acid for which the *trans* peptide bond configuration is similar in energy to the *cis* configuration. In the peptide backbones of all other amino acids, the nitrogen of the amino group contains a hydrogen substituent, and the α carbon is bound to a side chain. As a result, the *trans* configuration is preferred because the hydrogen atom is much less bulky than the α carbon (Figure 17). For proline, on the other hand, because its side chain is directly bonded to its amino group, there is always a steric clash between a carbon atom directly attached to nitrogen and the α carbon atom of the adjacent amino acid.

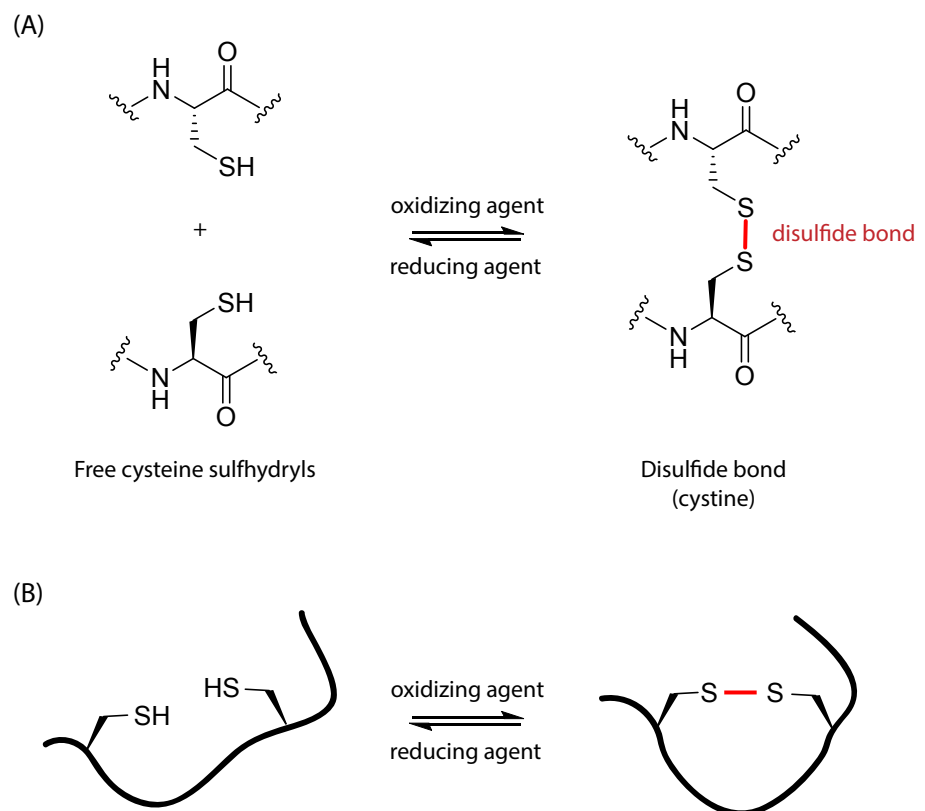
Cysteine side chains can form non-peptidic bonds in proteins

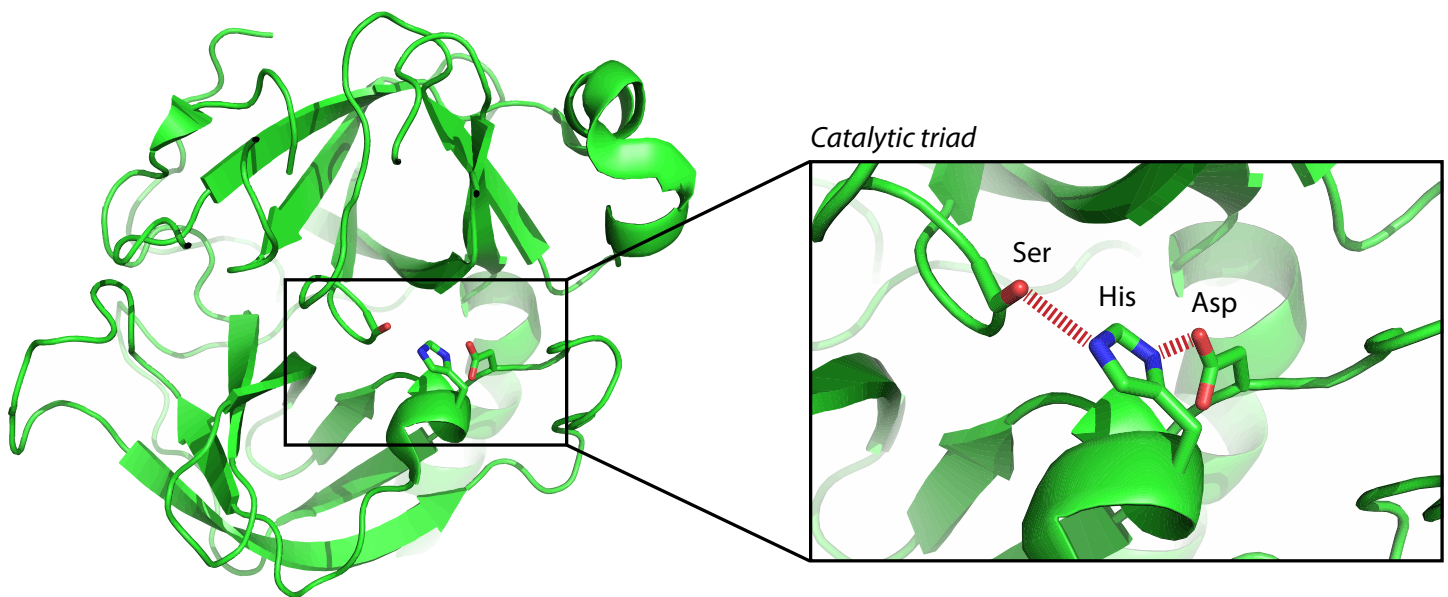
The amino acid cysteine contains a sulfhydryl (-SH) group on its side chain. Sulfhydryls can be oxidized to form **disulfide bonds** in which two cysteine side chains, often from distant locations in the primary sequence, come together to form a sulfur-sulfur covalent bond (Figure 18). The formation of such disulfide bonds in proteins rigidifies the protein and can stabilize conformations that are otherwise not particularly favored. A pair of disulfide-bonded cysteine residues is collectively known as **cystine**.

Your hair provides a familiar example of disulfide bonds. Hair features lots of these bonds, as they are important for its strength. Beauty salons take advantage of the disulfide bonds in your hair. If you want curly hair but were born with straight hair, you can go to the salon and get a “perm.” Your stylist adds a reducing agent, such as thioacetic acid, to your hair, which breaks the disulfide bonds and chemically transforms them back into free

Figure 18 Disulfide bonds are covalent linkages that form between cysteine side chains

(A) Shown is the reaction equation for the formation of a disulfide bond between two cysteine side chains. The disulfide bond is shown in red. (B) Shown is a diagram of a disulfide bond (*red*) forming between distant regions of a polypeptide chain (*black*). The disulfide bonds constrain the polypeptide chain into rigid conformations that promote protein folding.





Chymotrypsin (a protease enzyme)

Figure 19 Amino acid side chains are the functional components of proteins

Amino acids often function cooperatively, especially in enzymes. As an example, chymotrypsin is an enzyme that cleaves peptide bonds. This chemical reaction is facilitated using a set of three amino acids—aspertate, histidine, and serine—that function cooperatively to form a “catalytic triad.”

cysteines. Then the hair is curled around rollers, which places different sulfhydryl groups in close proximity, and finally, your hair is treated with an oxidizing agent, usually hydrogen peroxide, to form new disulfide bonds. Now instead of straight hair, you have curled hair.

Amino acids often function cooperatively in proteins

The amino acid side chains we have been describing are key functional components of proteins. Many proteins have enzymatic functions, and in these cases the side chains are directly involved in the reactions the enzymes catalyze. One class of enzymes, proteases, has evolved to catalyze the breakdown of peptide bonds in proteins. Many proteases contains a “catalytic triad” in which three amino acids in a precise arrangement work together to hydrolyze peptide bonds (Figure 19). The catalytic triad involves serine, histidine, and aspartate. The enzyme uses this catalytic triad to facilitate enzyme-catalyzed peptide bond hydrolysis. Of course, this can only occur because the folded protein precisely positions each of these amino acids in three dimensions, enabling the amino acids to interact with each other and with the substrate in just the right way. We will learn more about enzyme catalysis in later chapters, but before we do that, we need to examine the folded structures of proteins and the thermodynamics that underlie their formation.

Summary

Proteins are the most diverse and versatile macromolecules found in living systems. Proteins are polymeric chains of amino acid monomers connected by covalent peptide bonds. Unlike most other biological polymers, proteins fold into unique structures with distinct physical and chemical properties. As a result, cells use proteins for a broad range of structural, catalytic, regulatory, and transport functions. The folded structure of a protein is solely determined by its amino acid sequence, or primary structure. In fact, even small changes in the amino acid sequence can potentially alter a protein's folded structure, as we saw in the example of sickle-cell disease.

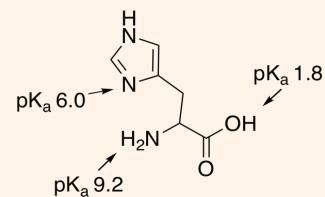
The amino acids that make up proteins all contain a central α carbon that is connected to an amino ($-\text{NH}_2$) group, a carboxylic acid ($-\text{COOH}$) group, an R group (i.e., side chain), and a hydrogen atom. The chemical composition of the side chain determines the identity of the amino acid as well as its chemical properties. For all amino acids except glycine (for which the side chain is a hydrogen atom), the α carbon is chiral, meaning that it is not superimposable with its mirror image. Stereochemistry, which is the specific configuration at a chiral center, can have enormous effects on molecular properties. We use the term enantiomers to describe molecules that are mirror images of each other, and we refer to one enantiomer as "D" and the other as "L." Living systems have evolved to use only the L-enantiomers of amino acids to construct proteins.

Amino acids are connected by peptide bonds to form polypeptide chains. Polypeptides have two ends, one with a free amino group, the N-terminus, and one with a free carboxylic acid group, the C-terminus. At physiological pH, both of these termini are ionized. The amide in the peptide backbone, however, cannot be protonated in water. Peptide bonds are flat and polar, and they do not rotate freely. These properties influence the conformations that the peptide backbone can adopt, ultimately influencing how proteins fold. Furthermore, the conformations of the peptide backbone are influenced by amino acid side chains, which create steric clashes that bias the peptide backbone into specific conformations. Glycine and proline have a particularly strong influence on backbone conformation. Glycine can adopt a wide range of conformations since it lacks a side chain, and as a result, polypeptides rich in glycine tend to be more flexible. Proline, however, constrains the backbone into a narrower range of conformations. Due to the unique nature of its side chain, proline can sometimes take part in cis peptide bonds, which are generally disfavored for other amino acids.

The amino acids found in proteins have a diverse range of properties that resist easy categorization. Some side chains are polar, some are nonpolar, and some are slightly polar. Several are positively or negatively charged at physiological pH. And yet others have ring structures that simultaneously have polar and nonpolar features. In the next chapter we will see how this diversity of amino acid properties influences the structure, folding, and myriad behaviors of proteins.

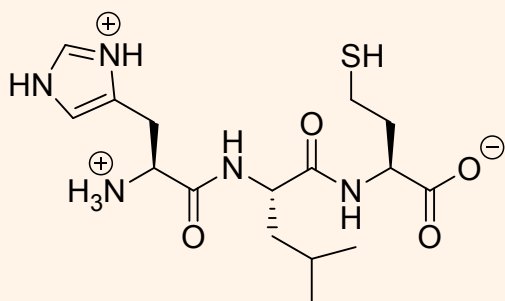
Practice Problems

1. At physiological pH of 7.4, most amino acids are predominantly (>99%) present in one protonation state. Only histidine is substantially (>1%) present in two different protonation states. What are the charges of those two states?

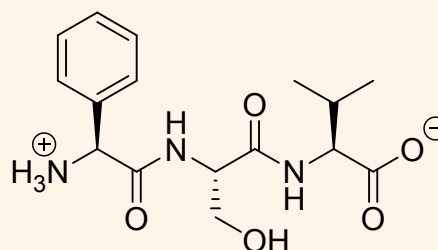


2. A mistake was made when drawing each of the following peptides. For each peptide, identify the mistake and correct the drawing.

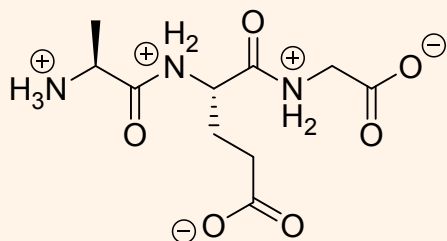
a. NH_2 -His-Leu-Cys-COOH at pH 7.4



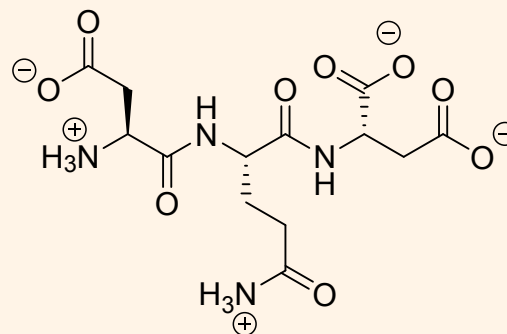
b. NH_2 -Phe-Ser-Val-COOH at pH 7.4



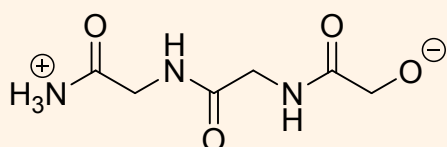
c. NH_2 -Ala-Glu-Gly-COOH at pH 7.4



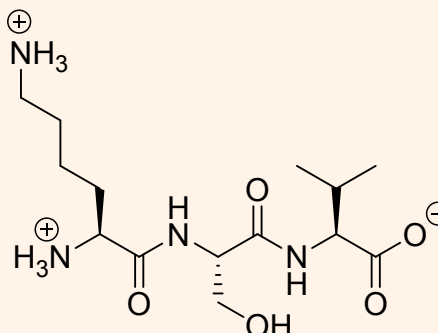
d. NH_2 -Asp-Gln-Glu-COOH at pH 7.4



e. NH_2 -Gly-Gly-Gly-COOH at pH 7.4



f. NH_2 -Val-Ser-Lys-COOH at pH 7.4

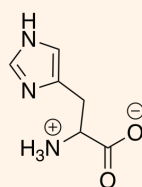
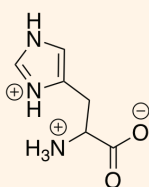


(Solutions are located on the next page.)

Solutions to practice problems

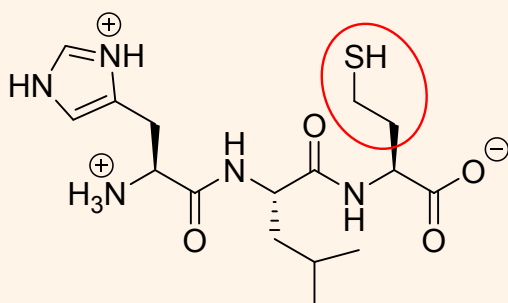
Question 1:

Positive and neutral. The important thing to recognize here is that one of the three pK_a values listed for histidine is close to the pH of 7.4. The Henderson-Hasselbalch equation gives us the ratio of the conjugate acid to the conjugate base form of an acid as a function of the difference between the pH of the solution and the pK_a of the acid. As we've seen, when that difference is 0, the ratio is 1. If we consider other numbers, we will see that a ratio of 10:1 results when the pH differs from the pK_a by 1 pH unit, or a ratio of 100:1 when the pH differs from the pK_a by two pH units. Thus, the only way we can have more than 1% of two different protonation states is if the pH differs from the pK_a by less than two pH units. The histidine pK_a of 6.0 of the 5-membered imidazole ring is sufficiently close to the pH of 7.4 that this is possible. We need to identify that group as a base as drawn, so it becomes cationic upon protonation to its conjugate acid below pH 6. Consequently, the two forms that are present at more than 1% each at a physiological pH of 7.4 are:



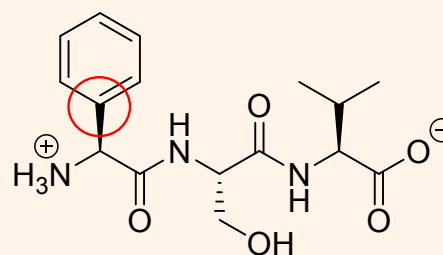
Question 2:

a. NH_2 -His-Leu-Cys-COOH at pH 7.4



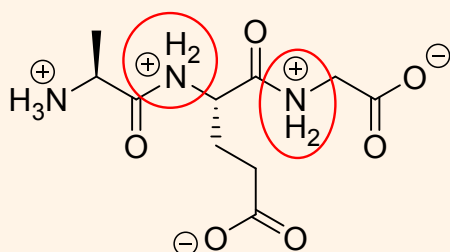
The side chain of cysteine contains one fewer carbon than is drawn here.

b. NH_2 -Phe-Ser-Val-COOH at pH 7.4



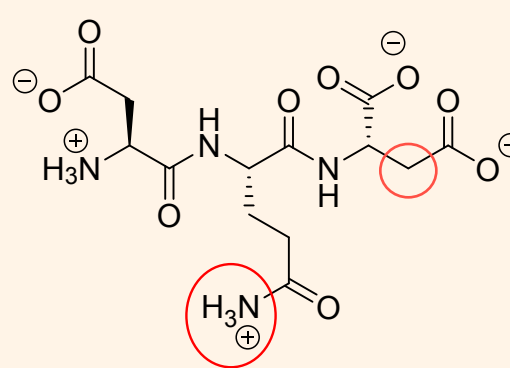
The side chain of phenylalanine contains one more carbon than is drawn here.

c. NH_2 -Ala-Glu-Gly-COOH at pH 7.4



The amide functional group found in the peptide backbone cannot be ionized. This structure would be correct if the circled nitrogens were drawn as "NH" instead of "NH₂⁺".

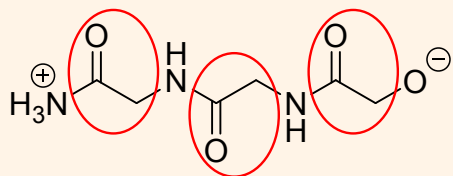
d. NH_2 -Asp-Gln-Glu-COOH at pH 7.4



The glutamine side chain cannot be ionized. This structure would be correct if the circled nitrogens were drawn as "NH₂" instead of "NH₃⁺".

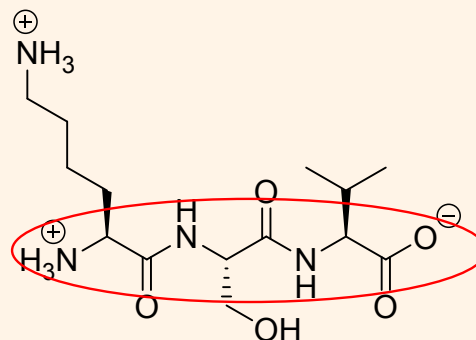
Solutions to practice problems

e. $\text{NH}_2\text{-Gly-Gly-Gly-COOH}$ at pH 7.4



The peptide backbone is not drawn consistently. This structure would be correct if all of the carbonyl oxygens were shifted to the right by one carbon atom.

f. $\text{NH}_2\text{-Val-Ser-Lys-COOH}$ at pH 7.4



This peptide is drawn backwards, and is the structure of " $\text{NH}_2\text{-Lys-Ser-Val-COOH}$ " instead of " $\text{NH}_2\text{-Val-Ser-Lys-COOH}$."