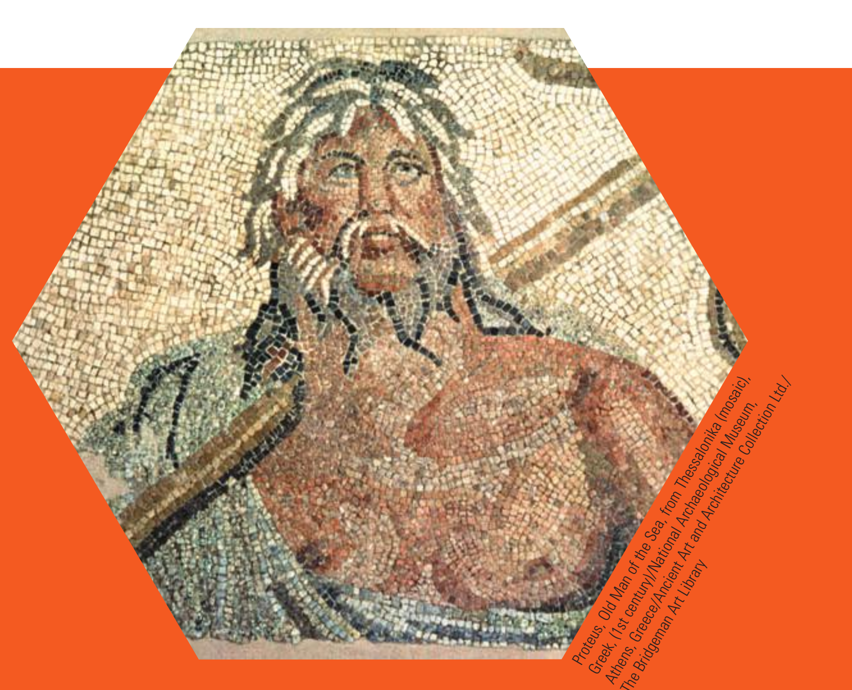
CHAPTER

15


Proteus, Old Man of the Sea, from Thessalonika (mosaic), Greek, (1st century)/National Archaeological Museum, Athens, Greece/Ancient Art and Architecture Collection Ltd./ The Bridgeman Art Library

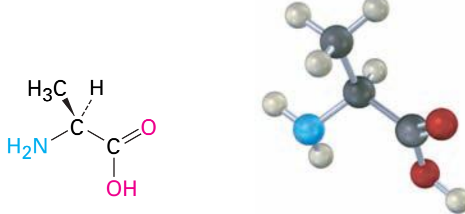
The building blocks of life that we call proteins are aptly named after Proteus, the early Greek sea-god whose name means "first" or "primordial."

Biomolecules: Amino Acids, Peptides, and Proteins

15.1 Structures of Amino Acids
15.2 Isoelectric Points
15.3 Peptides and Proteins
15.4 Covalent Bonding in Peptides
15.5 Peptide Structure Determination: Amino Acid Analysis
15.6 Peptide Sequencing: The Edman Degradation
15.7 Peptide Synthesis
15.8 Protein Structure
15.9 Enzymes and Coenzymes
15.10 How Do Enzymes Work? Citrate Synthase
Interlude—X-Ray Crystallography

***Proteins* occur in every living organism, are of many different types, and have many different biological functions.** The keratin of skin and fingernails, the fibroin of silk and spider webs, and the estimated 50,000 or so enzymes that catalyze the biological reactions in our bodies are all proteins. Regardless of their function, all proteins have a fundamentally similar structure and are made up of many *amino acids* linked together in a long chain.

Amino acids, as their name implies, are difunctional. They contain both a basic amino group and an acidic carboxyl group.



Alanine, an amino acid

Their value as building blocks to make proteins stems from the fact that amino acids can join together into long chains by forming amide bonds between the $-NH_2$ of one amino acid and the $-CO_2H$ of another. For classification

Online homework for this chapter can be assigned in OWL, an online homework assessment tool.

purposes, chains with fewer than 50 amino acids are often called **peptides**, while the term **protein** is generally used for larger chains.

Amide bonds

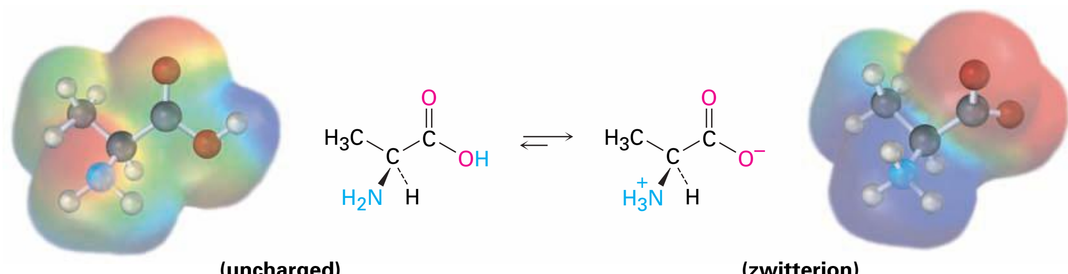
Many

WHY THIS CHAPTER?

Continuing our look at the main classes of biomolecules, we'll focus in this chapter on amino acids, the fundamental building blocks from which the 150,000 or so proteins in our bodies are made. We'll then see how amino acids are incorporated into proteins, a topic crucial to any understanding of biological chemistry.

15.1 Structures of Amino Acids

Because amino acids contain both basic amino and acidic carboxyl groups, they undergo an intramolecular acid–base reaction and exist in aqueous solution primarily in the form of dipolar ions, called **zwitterions** (from the German *zwitter,* meaning "hybrid").



(uncharged) (zwitterion)

Alanine

Amino acid zwitterions are internal salts and therefore have many of the physical properties associated with salts. They are relatively soluble in water but insoluble in hydrocarbons and are crystalline substances with relatively high melting points. In addition, amino acids are *amphiprotic,* meaning that they can react either as acids or as bases, depending on the circumstances. In aqueous acid solution, an amino acid zwitterion is a base that accepts a proton

onto its $-CO_2^-$ group to yield a cation; in aqueous base solution, the zwitterion is an acid that loses a proton from its $-NH_3^+$ group to form an anion.

In acid solution

$$RCH(NH_3^+)CO_2^- + H_3O^+ \rightleftharpoons RCH(NH_3^+)CO_2H + H_2O$$

In base solution

$$RCH(NH_3^+)CO_2^- + OH^- \rightleftharpoons RCH(NH_2)CO_2^- + H_2O$$

The structures, abbreviations (both one-letter and three-letter), and pK_a values of the 20 amino acids commonly found in proteins are shown in Table 15.1. All 20 are ***α*-amino acids**, meaning that the amino group in each is a substituent on the *α* carbon—the one next to the carbonyl group. Nineteen of the twenty are primary amines, RNH_2, and differ only in the identity of the **side chain**—the substituent attached to the *α* carbon. Proline is a secondary amine whose nitrogen and *α* carbon atoms are part of a five-membered pyrrolidine ring.

Side chain

A primary *α*-amino acid

Proline, a secondary *α*-amino acid

In addition to the 20 amino acids commonly found in proteins, 2 others—selenocysteine and pyrrolysine—are found in some organisms, and more than 700 nonprotein amino acids are also found in nature. *γ*-Aminobutyric acid (GABA), for instance, is found in the brain and acts as a neurotransmitter; homocysteine is found in blood and is linked to coronary heart disease; and thyroxine is found in the thyroid gland, where it acts as a hormone.

Selenocysteine

Pyrrolysine

***γ*-Aminobutyric acid**

Homocysteine

Thyroxine

Table 15.1 Structures of the 20 Common Amino Acids Found in Proteins

Name	Abbreviations		MW	Stucture	pK_a α-CO_2H	pK_a α-NH_3^+	pK_a side chain	p*I*
Neutral Amino Acids								
Alanine	Ala	A	89		2.34	9.69	—	6.01
Asparagine	Asn	N	132		2.02	8.80	—	5.41
Cysteine	Cys	C	121		1.96	10.28	8.18	5.07
Glutamine	Gln	Q	146		2.17	9.13	—	5.65
Glycine	Gly	G	75		2.34	9.60	—	5.97
Isoleucine	Ile	I	131		2.36	9.60	—	6.02
Leucine	Leu	L	131		2.36	9.60	—	5.98
Methionine	Met	M	149		2.28	9.21	—	5.74
Phenylalanine	Phe	F	165		1.83	9.13	—	5.48
Proline	Pro	P	115		1.99	10.60	—	6.30

continued

Table 15.1 Structures of the 20 Common Amino Acids Found in Proteins *(continued)*

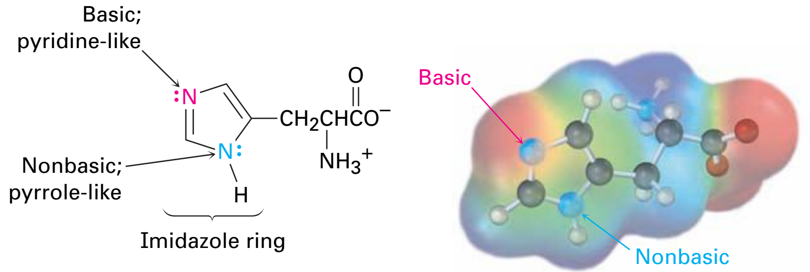
Name	Abbreviations		MW	Stucture	pK_a α-CO_2H	pK_a α-NH_3^+	pK_a side chain	p*I*
Neutral Amino Acids *continued*								
Serine	Ser	S	105		2.21	9.15	—	5.68
Threonine	Thr	T	119		2.09	9.10	—	5.60
Tryptophan	Trp	W	204		2.83	9.39	—	5.89
Tyrosine	Tyr	Y	181		2.20	9.11	10.07	5.66
Valine	Val	V	117		2.32	9.62	—	5.96
Acidic Amino Acids								
Aspartic acid	Asp	D	133		1.88	9.60	3.65	2.77
Glutamic acid	Glu	E	147		2.19	9.67	4.25	3.22
Basic Amino Acids								
Arginine	Arg	R	174		2.17	9.04	12.48	10.76
Histidine	His	H	155		1.82	9.17	6.00	7.59
Lysine	Lys	K	146		2.18	8.95	10.53	9.74

Except for glycine, $^{+}H_3NCH_2CO_2^{-}$, the α carbons of amino acids are chirality centers. Two enantiomers of each are therefore possible, but nature uses only one to build proteins. In Fischer projections, naturally occurring amino acids are represented by placing the carboxyl group at the top and the side chain at the bottom as if drawing a carbohydrate (Section 14.2) and then placing the amino group on the left. Because of their stereochemical similarity to L sugars (Section 14.3), the naturally occurring α-amino acids are often referred to as L amino acids.

CO_2^- / $H_3\overset{+}{N}$—C—H / CH_3	CO_2^- / $H_3\overset{+}{N}$—C—H / CH_2OH	CO_2^- / $H_3\overset{+}{N}$—C—H / CH_2SH	[CHO / HO—C—H / CH_2OH]
L-Alanine (*S*)-Alanine	**L-Serine (*S*)-Serine**	**L-Cysteine (*R*)-Cysteine**	**L-Glyceraldehyde**

The 20 common amino acids can be further classified as neutral, acidic, or basic, depending on the structure of their side chains. Fifteen of the twenty have neutral side chains, two (aspartic acid and glutamic acid) have an extra carboxylic acid function in their side chains, and three (lysine, arginine, and histidine) have basic amino groups in their side chains. Note that both cysteine (a thiol) and tyrosine (a phenol), although usually classified as neutral, nevertheless have weakly acidic side chains that can be deprotonated in a sufficiently strong base solution.

At the physiological pH of 7.3 within cells, the side-chain carboxyl groups of aspartic acid and glutamic acid are deprotonated and the basic side-chain nitrogens of lysine and arginine are protonated. Histidine, however, which contains a heterocyclic imidazole ring in its side chain, is not quite basic enough to be protonated at pH 7.3. Note that only the pyridine-like, doubly bonded nitrogen in histidine is basic. The pyrrole-like singly bonded nitrogen is nonbasic because its lone pair of electrons is part of the six-π-electron aromatic imidazole ring (Section 12.6).



Histidine

Humans are able to synthesize only 11 of the 20 protein amino acids, called *nonessential amino acids.* The other 9, called *essential amino acids,* are biosynthesized only in plants and microorganisms and must be obtained in our diet. The division between essential and nonessential amino acids is not clear-cut, however. Tyrosine, for instance, is sometimes considered nonessential because humans can produce it from phenylalanine, yet phenylalanine itself is essential and must be obtained in the diet. Arginine can be synthesized by humans, but much of the arginine we need also comes from our diet.

Worked Example 15.1

Acid–Base Reactions of Amino Acids

Write an equation for the reaction of glycine hydrochloride, $Cl^{-}\,{}^{+}H_3NCH_2CO_2H$, with **(a)** 1 equivalent of NaOH and **(b)** 2 equivalents of NaOH.

Solution

(a) Reaction with the first equivalent of NaOH removes the acidic $-CO_2H$ proton to give the neutral zwitterion.

$$Cl^{-}\ {}^{+}H_3NCH_2\overset{O}{\overset{\|}{C}}OH + NaOH \longrightarrow {}^{+}H_3NCH_2\overset{O}{\overset{\|}{C}}O^{-} + H_2O + NaCl$$

(b) Once the zwitterion has formed, reaction with a second equivalent of NaOH removes the remaining acidic proton from the $-NH_3^{+}$ group to give an amino carboxylate anion.

$${}^{+}H_3NCH_2\overset{O}{\overset{\|}{C}}O^{-} + NaOH \longrightarrow H_2NCH_2\overset{O}{\overset{\|}{C}}O^{-}\ Na^{+} + H_2O$$

Problem 15.1 How many of the α-amino acids in Table 15.1 contain aromatic rings? How many contain sulfur? How many are alcohols? How many have hydrocarbon side chains?

Problem 15.2 Of the 19 L amino acids, 18 have the *S* configuration at the α carbon. Cysteine is the only L amino acid that has an *R* configuration. Explain.

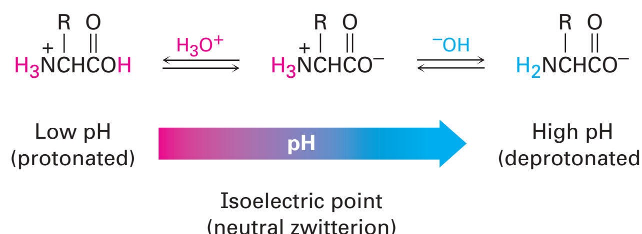
Problem 15.3 Draw L-alanine in the standard three-dimensional format using solid, wedged, and dashed lines.

Problem 15.4 Write the products of the following reactions:

(a) Phenylalanine + 1 equiv NaOH → **?**
(b) Product of **(a)** + 1 equiv HCl → **?**
(c) Product of **(a)** + 2 equiv HCl → **?**

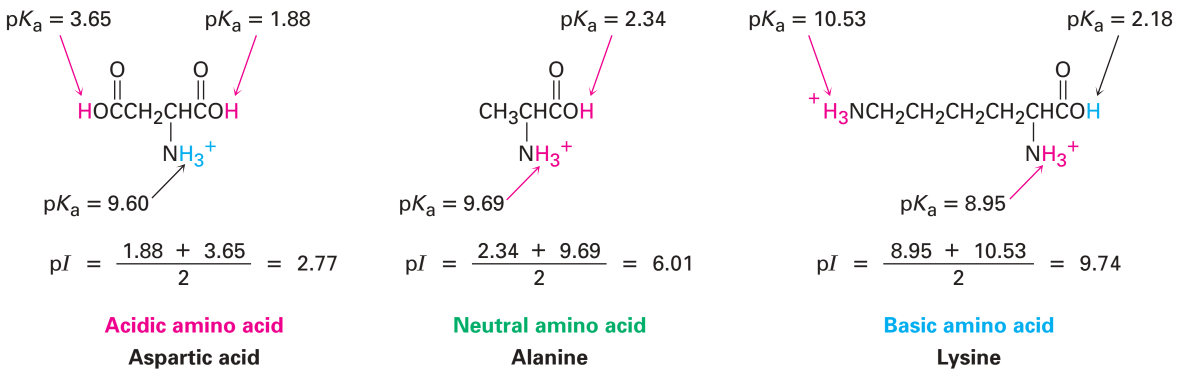
15.2 Isoelectric Points

In acid solution, an amino acid is protonated and exists primarily as a cation. In base solution, an amino acid is deprotonated and exists primarily as an anion. Thus, at some intermediate pH, the amino acid must be exactly balanced between anionic and cationic forms and exist primarily as the neutral zwitterion. This pH is called the amino acid's **isoelectric point, p*I*.**



The isoelectric point of an amino acid depends on its structure, with values for the 20 common amino acids given in Table 15.1. The 15 neutral amino acids have isoelectric points near neutrality, in the pH range 5.0 to 6.5. The two acidic amino acids have isoelectric points at lower pH so that deprotonation of the side-chain $-CO_2H$ is suppressed, and the three basic amino acids have isoelectric points at higher pH so that protonation of the side-chain amino group is suppressed.

More specifically, the p*I* of any amino acid is the average of the two acid-dissociation constants that involve the neutral zwitterion. For the 13 amino acids with a neutral side chain, p*I* is the average of pK_{a1} and pK_{a2}. For the four amino acids with either a strongly or weakly acidic side chain, p*I* is the average of the two *lowest* pK_a values. For the three amino acids with a basic side chain, p*I* is the average of the two *highest* pK_a values.



Just as individual amino acids have isoelectric points, entire proteins have an overall p*I* because of the cumulative effect of all the acidic or basic amino acids they may contain. The enzyme lysozyme, for instance, has a preponderance of basic amino acids and thus has a high isoelectric point (p*I* = 11.0). Pepsin, however, has a preponderance of acidic amino acids and a low isoelectric point (p*I* ~ 1.0). Not surprisingly, the solubilities and properties of proteins with different p*I*'s are strongly affected by the pH of the medium. Solubility in water is usually lowest at the isoelectric point, where the protein has no net charge, and is higher both above and below the p*I*, where the protein is charged.

We can take advantage of the differences in isoelectric points to separate a mixture of proteins into its pure constituents. Using a technique known as *electrophoresis,* a mixture of proteins is placed near the center of a strip of paper or gel. The paper or gel is then moistened with an aqueous buffer of a given pH, and electrodes are connected to the ends of the strip. When an electric potential is applied, those proteins with negative charges (those that are deprotonated because the pH of the buffer is above their isoelectric point) migrate slowly toward the positive electrode. At the same time, those amino acids with positive charges (those that are protonated because the pH of the buffer is below their isoelectric point) migrate toward the negative electrode.

Different proteins migrate at different rates, depending on their isoelectric points and on the pH of the aqueous buffer, thereby effecting a separation of

the mixture into its components. Figure 15.1 illustrates the separation for a mixture containing basic, neutral, and acidic components.

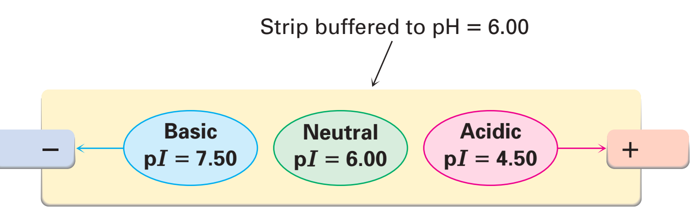


Figure 15.1 Separation of a protein mixture by electrophoresis. At pH = 6.00, a neutral protein does not migrate, a basic protein is protonated and migrates toward the negative electrode, and an acidic protein is deprotonated and migrates toward the positive electrode.

Problem 15.5 For the mixtures of amino acids indicated, predict the direction of migration of each component (toward the positive or negative electrode) and the relative rate of migration during electrophoresis.

(a) Valine, glutamic acid, and histidine at pH 7.6
(b) Glycine, phenylalanine, and serine at pH 5.7
(c) Glycine, phenylalanine, and serine at pH 6.0

Problem 15.6 Hemoglobin has pI = 6.8. Does hemoglobin have a net negative charge or net positive charge at pH = 5.3? At pH = 7.3?

15.3 Peptides and Proteins

Proteins and peptides are amino acid polymers in which the individual amino acids, called **residues**, are joined together by amide bonds, or *peptide bonds*. An amino group from one residue forms an amide bond with the carboxyl of a second residue, the amino group of the second forms an amide bond with the carboxyl of a third, and so on. For example, alanylserine is the dipeptide that results when an amide bond forms between the alanine carboxyl and the serine amino group.

Alanine (Ala)

+

Serine (Ser)

⟶

Alanylserine (Ala-Ser)

Note that two dipeptides can result from reaction between alanine and serine, depending on which carboxyl group reacts with which amino group.

If the alanine amino group reacts with the serine carboxyl, serylalanine results.

Serine (Ser)

+

Alanine (Ala)

Serylalanine (Ser-Ala)

The long, repetitive sequence of –N–CH–CO– atoms that makes up a continuous chain is called the protein's **backbone**. By convention, peptides are written with the **N-terminal amino acid** (the one with the free $-NH_3^+$ group) on the left and the **C-terminal amino acid** (the one with the free $-CO_2^-$ group) on the right. The name of the peptide is indicated by using the abbreviations listed in Table 15.1 for each amino acid. Thus, alanylserine is abbreviated Ala-Ser or A-S, and serylalanine is abbreviated Ser-Ala or S-A. The one-letter abbreviations are more convenient, though less immediately recognizable, than the three-letter abbreviations.

Worked Example 15.2

Drawing the Structure of a Dipeptide

Draw the structure of Ala-Val.

Strategy By convention, the N-terminal amino acid is written on the left and the C-terminal amino acid on the right. Thus, alanine is N-terminal, valine is C-terminal, and the amide bond is formed between the alanine $-CO_2H$ and the valine $-NH_2$.

Solution

Ala-Val

Worked Example 15.3

Identifying Tripeptides

There are six tripeptides that contain methionine, lysine, and isoleucine. Name them using both three- and one-letter abbreviations.

Solution

Met-Lys-Ile (M-K-I)	Lys-Met-Ile (K-M-I)	Ile-Met-Lys (I-M-K)
Met-Ile-Lys (M-I-K)	Lys-Ile-Met (K-I-M)	Ile-Lys-Met (I-K-M)

Problem 15.7 Draw structures of the two dipeptides made from leucine and cysteine.

Problem 15.8 Using both three- and one-letter notations for each amino acid, name the six possible isomeric tripeptides that contain valine, tyrosine, and glycine.

Problem 15.9 Draw the structure of Met-Pro-Val-Gly, and indicate the positions of the amide bonds.

15.4 Covalent Bonding in Peptides

The amide bond that links different amino acids together in peptides is no different from any other amide bond (Section 10.10). An amide nitrogen is nonbasic because its unshared electron pair is delocalized by interaction with the carbonyl group. This overlap of the nitrogen *p* orbital with the *p* orbitals of the carbonyl group imparts a certain amount of double-bond character to the C–N bond and restricts rotation around it. The amide bond is therefore planar, and the N–H is oriented 180° to the C=O.

Restricted rotation

Planar

A second kind of covalent bonding in peptides occurs when a disulfide linkage, RS—SR, is formed between two cysteine residues. As we saw in Section 8.8, a disulfide is formed by mild oxidation of a thiol, RSH, and is cleaved by mild reduction.

Cysteine Cysteine

Disulfide bond

A disulfide bond between cysteine residues in different peptide chains links the otherwise separate chains together, while a disulfide bond between cysteine residues in the same chain forms a loop. Insulin, for instance, is

composed of two chains that total 51 amino acids and are linked by two cysteine disulfide bridges.

A chain (21 units)

Gly-Ile-Val-Glu-Gln-Cys-Cys-Thr-Ser-Ile-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn

B chain (30 units)

Phe-Val-Asn-Glu-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Thr

Insulin

15.5 Peptide Structure Determination: Amino Acid Analysis

To determine the structure of a protein or peptide, we need to answer three questions: What amino acids are present? How much of each is present? In what sequence do the amino acids occur in the peptide chain? The answers to the first two questions are provided by an automated instrument called an *amino acid analyzer*.

In preparation for analysis, a peptide is broken into its constituent amino acids by reducing all disulfide bonds, capping the –SH groups of cysteine residues by S_N2 reaction with iodoacetic acid, and hydrolyzing the amide bonds by heating with aqueous 6 M HCl at 110 °C for 24 hours. The resultant amino acid mixture is then separated into its constituents by a technique called ion-exchange chromatography. The mixture is placed at the top of a glass column filled with a special absorbent material, and a series of aqueous buffers is pumped through the column. The various amino acids migrate down the column at different rates depending on their structures and are thus separated.

As each amino acid exits the column, it is mixed with a solution of a substance called *ninhydrin* and undergoes a rapid reaction that produces an intense purple color. The color is detected by a spectrometer, and a plot of exit time versus spectrometer absorbance is obtained.

Ninhydrin + α-Amino acid $\xrightarrow[H_2O]{NaOH}$ (purple color) + RCHO + CO_2

Because the time required for a given amino acid to pass down a standard column is reproducible, the identities of the amino acids in a peptide can be determined. The amount of each amino acid in the sample is determined by measuring the intensity of the purple color resulting from its reaction with ninhydrin. Figure 15.2 shows the results of amino acid analysis of a standard equimolar mixture of 17 α-amino acids. Typically, amino acid analysis requires about 100 picomoles (2–3 μg) of sample for a protein containing about 200 residues.

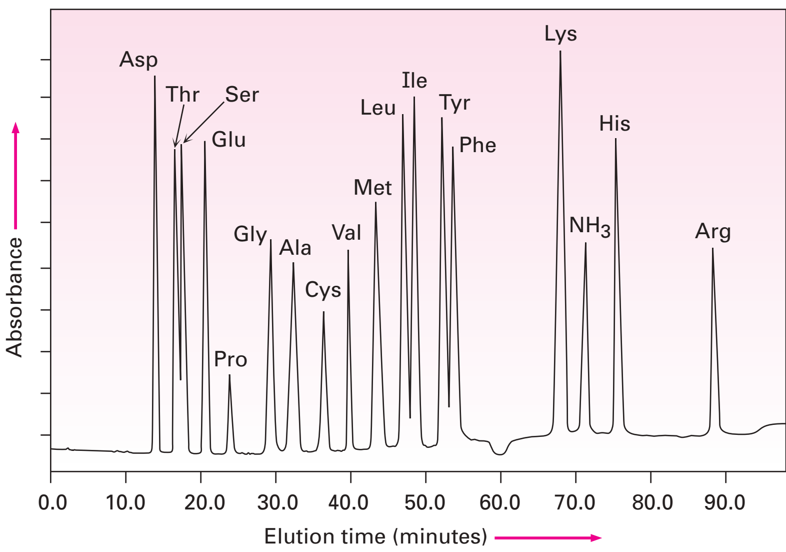


Figure 15.2 Amino acid analysis of an equimolar mixture of 17 amino acids.

Problem 15.10 Show the structure of the product you would expect to obtain by S_N2 reaction of a cysteine residue with iodoacetic acid.

Problem 15.11 Show the structures of the products obtained on reaction of valine with ninhydrin.

15.6 Peptide Sequencing: The Edman Degradation

With the identities and relative amounts of amino acids known, the peptide is then *sequenced* to find out in what order the amino acids are linked together. Much peptide sequencing is now done by mass spectrometry (Section 13.1), but a chemical method of peptide sequencing called the *Edman degradation* is also used.

The general idea of peptide sequencing by Edman degradation is to cleave one amino acid at a time from the N terminus of the peptide chain. That

terminal amino acid is then separated and identified, and the cleavage reactions are repeated on the chain-shortened peptide until the entire peptide sequence is known. Automated protein sequencers are available that allow as many as 50 repetitive sequencing cycles, with an efficiency that allows sample sizes of 1 to 5 picomoles—less than 0.1 μg of peptide.

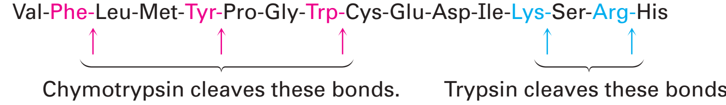
Edman degradation involves treatment of a peptide with phenyl isothiocyanate (PITC), $C_6H_5—N{=}C{=}S$, followed by mild acid hydrolysis. PITC first attaches to the $-NH_2$ group of the N-terminal amino acid, and the N-terminal residue then splits from the chain giving a *phenylthiohydantoin* derivative (PTH) along with chain-shortened peptide. The PTH is identified by comparison with known derivatives of the common amino acids, and the chain-shortened peptide is automatically resubmitted to another round of Edman degradation.

1. PITC
2. H_3O^+

A phenylthiohydantoin (PTH)

Complete sequencing of large proteins by Edman degradation is impractical because of the buildup of unwanted by-products. To get around the problem, a large peptide chain is first cleaved by partial hydrolysis into a number of smaller fragments, the sequence of each fragment is determined, and the individual fragments are fitted together by matching the overlapping ends. In this way, protein chains with more than 400 amino acids have been sequenced.

Partial hydrolysis of a protein can be carried out either chemically with aqueous acid or enzymatically. Acid hydrolysis is unselective and gives a more-or-less random mixture of small fragments, but enzymatic hydrolysis is quite specific. The enzyme trypsin, for instance, catalyzes hydrolysis of peptides only at the carboxyl side of the basic amino acids arginine and lysine; chymotrypsin cleaves only at the carboxyl side of the aryl-substituted amino acids phenylalanine, tyrosine, and tryptophan.



Worked Example 15.4

Sequencing a Simple Peptide

Angiotensin II is a hormonal octapeptide involved in controlling hypertension by regulating the sodium–potassium salt balance in the body. Amino acid analysis shows the presence of eight different amino acids in equimolar amounts: Arg,

Asp, His, Ile, Phe, Pro, Tyr, and Val. Partial hydrolysis of angiotensin II with dilute hydrochloric acid yields the following fragments:

Asp-Arg-Val-Tyr, Ile-His-Pro, Pro-Phe, Val-Tyr-Ile-His

What is the sequence of angiotensin II?

Strategy Line up the fragments to identify the overlapping regions, and then write the sequence.

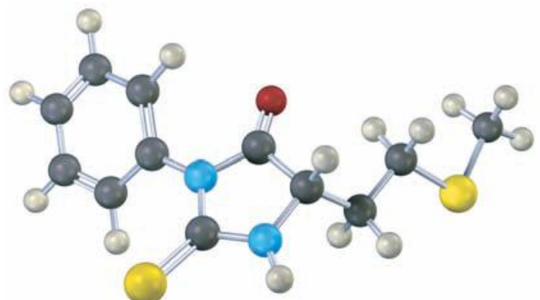
```
Asp-Arg-Val-Tyr
        :   :
        Val-Tyr-Ile-His
                :   :
                Ile-His-Pro
                        :
                        Pro-Phe
```

Solution The sequence is Asp-Arg-Val-Tyr-Ile-His-Pro-Phe.

Problem 15.12 What fragments would result if angiotensin II (Worked Example 15.4) were cleaved with trypsin? With chymotrypsin?

Problem 15.13 Give the amino acid sequence of a hexapeptide containing Arg, Gly, Ile, Leu, Pro, and Val that produces the following fragments on partial acid hydrolysis: Pro-Leu-Gly, Arg-Pro, Gly-Ile-Val.

Problem 15.14 What is the N-terminal residue on a peptide that gives the following PTH derivative on Edman degradation?



15.7 Peptide Synthesis

Once the structure of a peptide is known, synthesis is often the next goal—perhaps to obtain a larger amount for biological evaluation. A simple amide can be formed by treating an amine and a carboxylic acid with dicyclohexylcarbodiimide (DCC; Section 10.6), but peptide synthesis is more difficult

because of the need for specificity. Many different amide links must be formed in a precise order rather than at random.

The solution to the specificity problem is *protection* (Section 9.8). If we wanted to couple alanine with leucine to synthesize Ala-Leu, for instance, we could protect the $-NH_2$ group of alanine and the $-CO_2H$ group of leucine to shield them from reacting, then form the desired Ala-Leu amide bond by reaction with DCC, and then remove the protecting groups.

Alanine — Protect $-NH_2$ → RHN–CH(CH_3)–CO_2H

Leucine — Protect $-CO_2H$ → H_2N–CH($CH_2CH(CH_3)_2$)–C(=O)–O–R

1. DCC (form amide) 2. Deprotect → Ala-Leu

Carboxyl groups are often protected simply by converting them into methyl or benzyl esters. Esters are easily made from carboxylic acids (Section 10.6) and are easily hydrolyzed by mild treatment with aqueous NaOH.

Leucine — CH_3OH, HCl → Methyl leucinate — 1. NaOH 2. H_3O^+ → Leucine

Leucine — $PhCH_2OH$, HCl → Benzyl leucinate — H_2/Pd → Leucine

Amino groups are often protected as their *tert*-butyloxycarbonyl amide (Boc) or fluorenylmethyloxycarbonyl amide (Fmoc) derivatives. The Boc protecting group is introduced by reaction of the amino acid with di-*tert*-butyl dicarbonate in a nucleophilic acyl substitution reaction and is removed by brief treatment with a strong acid such as trifluoroacetic acid, CF_3CO_2H. The

Fmoc protecting group is introduced by reaction with an acid chloride and is removed by treatment with base.

Alanine → Di-*tert*-butyl dicarbonate, $(CH_3CH_2)_3N$ → Boc-Ala

Alanine → Fluorenylmethyloxycarbonyl chloride, $(CH_3CH_2)_3N$ → Fmoc-Ala

In summary, five steps are needed to synthesize a dipeptide such as Ala-Leu (Figure 15.3).

Figure 15.3 The five steps needed for peptide synthesis.

1. The amino group of alanine is protected as the Boc derivative, and
2. the carboxyl group of leucine is protected as the methyl ester.
3. The two protected amino acids are coupled using DCC.
4. The Boc protecting group is removed by acid treatment.
5. The methyl ester is removed by basic hydrolysis.

Ala + $(t\text{-BuOC})_2O$ —(1)→ Boc–Ala

Leu + CH_3OH —(2) H^+ catalyst→ Leu–OCH_3

—(3) DCC→ Boc–Ala-Leu–OCH_3

—(4) CF_3CO_2H→ Ala-Leu–OCH_3

—(5) NaOH, H_2O→ Ala-Leu

Although the steps just shown can be repeated to add one amino acid at a time to a growing chain, the synthesis of a large peptide by this sequential addition is arduous. Much simpler is the *Merrifield solid-phase* method, in which peptide synthesis is carried out with the growing amino acid chain covalently bonded to beads of a polymer resin rather than in solution. After bonding the first amino acid to the resin, a repeating series of four steps is carried out to build a peptide:

1. A Boc-protected amino acid is covalently linked to the polystyrene polymer by formation of an ester bond (S_N2 reaction).
2. The polymer-bonded amino acid is washed free of excess reagent and then treated with trifluoroacetic acid to remove the Boc group.
3. A second Boc-protected amino acid is coupled to the first by reaction with DCC. Excess reagents are removed by washing them from the insoluble polymer.
4. The cycle of deprotection, coupling, and washing is repeated as many times as desired to add amino acid units to the growing chain.
5. After the desired peptide has been made, treatment with anhydrous HF removes the final Boc group and cleaves the ester bond to the polymer, yielding the free peptide.

Boc—NHCH(R)COOH + $ClCH_2$—Polymer

(1) Base

Boc—NHCH(R)COOCH$_2$—Polymer

(2) 1. Wash 2. CF_3CO_2H

H_2NCH(R)COOCH$_2$—Polymer

(3) 1. DCC, Boc—NHCH(R′)COOH 2. Wash

Boc—NHCH(R′)C(O)—NHCH(R)COOCH$_2$—Polymer

(4) Repeat cycle many times

Boc—NHCH(R″)C(O)—(NHCH(R′)C(O))$_n$—NHCH(R)COOCH$_2$—Polymer

(5) HF

H_2NCH(R″)C(O)—(NHCH(R′)C(O))$_n$—NHCH(R)COOH + HOCH$_2$—Polymer

Robotic peptide synthesizers automatically repeat the coupling, washing, and deprotection steps with different amino acids. Each step occurs in high yield, and mechanical losses are minimized because the peptide intermediates

are never removed from the insoluble polymer until the final step. Using this procedure, up to 30 mg of a peptide with 20 amino acids can be routinely prepared in a few hours.

Worked Example 15.5

Reactions of Amino Acids

Write equations for the reaction of methionine with:
(a) CH_3OH, HCl **(b)** Di-*tert*-butyl dicarbonate

Solution

(a)

$$\underset{\displaystyle CH_2CH_2SCH_3}{H_2NCHCOH}\ (C{=}O) + CH_3OH \xrightarrow{HCl} \underset{\displaystyle CH_2CH_2SCH_3}{H_2NCHCOCH_3}\ (C{=}O) + H_2O$$

(b)

$$\underset{\displaystyle CH_2CH_2SCH_3}{H_2NCHCOH}\ (C{=}O) + (CH_3)_3COCOCOC(CH_3)_3\ (2\times C{=}O) \longrightarrow \underset{\displaystyle CH_2CH_2SCH_3}{(CH_3)_3COCNHCHCOH}\ (2\times C{=}O)$$

Problem 15.15 Write the structures of the intermediates in the five-step synthesis of Leu-Ala from alanine and leucine.

Problem 15.16 Show the mechanism for formation of a Boc derivative by reaction of an amino acid with di-*tert*-butyl dicarbonate.

15.8 Protein Structure

Proteins are usually classified as either *fibrous* or *globular,* according to their three-dimensional shapes. **Fibrous proteins**, such as the collagen in tendons and connective tissue and the myosin in muscle tissue, consist of polypeptide chains arranged side by side in long filaments. Because these proteins are tough and insoluble in water, they are used in nature for structural materials. **Globular proteins**, by contrast, are usually coiled into compact, roughly spherical shapes. These proteins are generally soluble in water and are mobile within cells. Most of the 3000 or so enzymes that have been characterized to date are globular proteins.

Proteins are so large that the word *structure* takes on a broader meaning than it does with simpler organic compounds. In fact, chemists speak of four different levels of structure when describing proteins:

- The **primary structure** of a protein is simply the amino acid sequence.
- The **secondary structure** of a protein describes how *segments* of the peptide backbone orient into a regular pattern.
- The **tertiary structure** describes how the *entire* protein molecule coils into an overall three-dimensional shape.
- The **quaternary structure** describes how different protein molecules come together to yield large aggregate structures.

Primary structure is determined, as we've seen, by sequencing the protein. Secondary, tertiary, and quaternary structures are determined either by NMR (Sections 13.7 through 13.13) or by a technique called X-ray crystallography, which is described in the *Interlude* at the end of this chapter.

The most common secondary structures are the α helix and the β-pleated sheet. An **α helix** is a right-handed coil of the protein backbone, much like the coil of a spiral staircase (Figure 15.4a). Each turn of the helix contains 3.6 amino acid residues, with a distance between coils of 540 pm, or 5.4 Å. The structure is stabilized by hydrogen bonds between amide N–H groups and C=O groups four residues away, with an N–H····O distance of 2.8 Å. The α helix is an extremely common secondary structure, and almost all globular proteins contain many helical segments. Myoglobin, a small globular protein containing 153 amino acid residues in a single chain, is an example (Figure 15.4b).

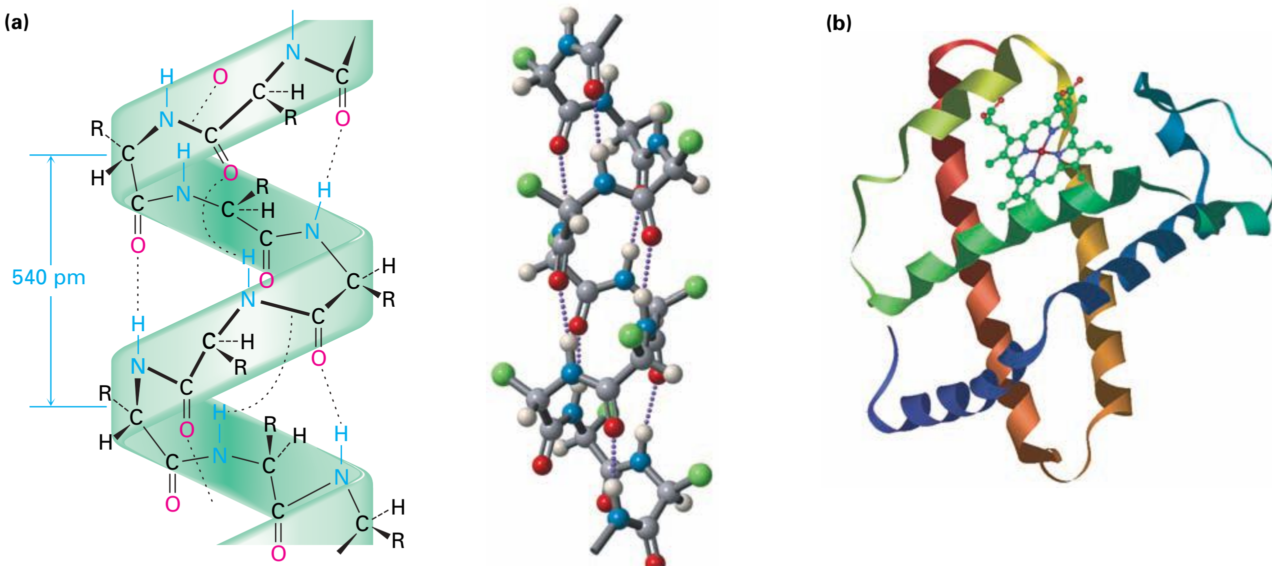
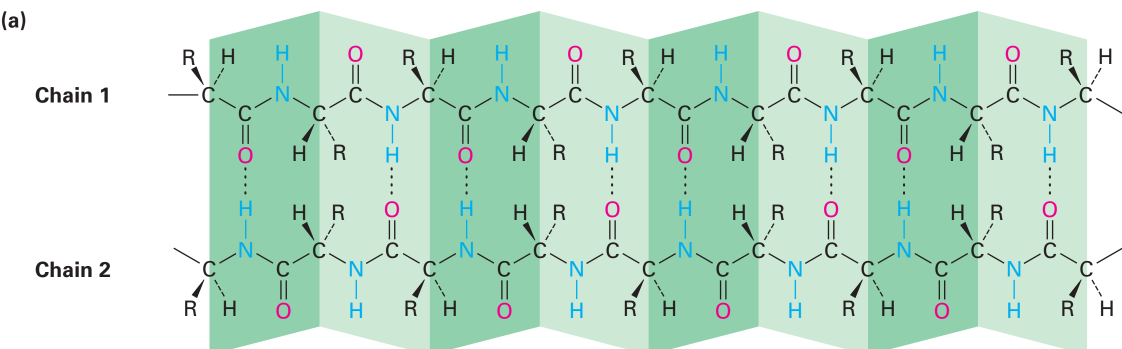


Figure 15.4 **(a)** The α-helical secondary structure of proteins is stabilized by hydrogen bonds between the N–H group of one residue and the C=O group four residues away. **(b)** The structure of myoglobin, a globular protein with extensive helical regions that are shown as ribbons in this representation.

A **β-pleated sheet** differs from an α helix in that the peptide chain is fully extended rather than coiled and the hydrogen bonds occur between residues in adjacent chains (Figure 15.5a). The neighboring chains can run either in the same direction (parallel) or in opposite directions (antiparallel), although the antiparallel arrangement is more common and energetically somewhat more favorable. Concanavalin A, for instance, consists of

two identical chains of 237 residues with extensive regions of antiparallel β sheets (Figure 15.5b).

(a)



(b)

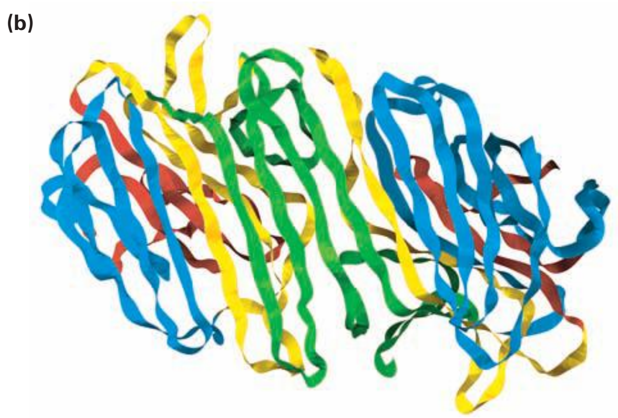


Figure 15.5 **(a)** The β-pleated sheet secondary structure of proteins is stabilized by hydrogen bonds between parallel or antiparallel chains. **(b)** The structure of concanavalin A, a protein with extensive regions of antiparallel β sheets, shown as ribbons.

What about tertiary structure? Why does a protein adopt the shape it does? The forces that determine the tertiary structure of a protein are the same forces that act on all molecules, regardless of size, to provide maximum stability. Particularly important are the hydrophilic (water-loving) interactions of the polar side chains on acidic or basic amino acids and the hydrophobic (water-fearing) interactions of nonpolar side chains. Those acidic or basic amino acids with charged side chains tend to congregate on the exterior of the protein, where they can be solvated by water. Those amino acids with neutral, nonpolar side chains tend to congregate on the hydrocarbon-like interior of a protein molecule, away from the aqueous medium.

Also important for stabilizing a protein's tertiary structure are the formation of disulfide bridges between cysteine residues, the formation of hydrogen bonds between nearby amino acid residues, and the presence of ionic attractions, called *salt bridges,* between positively and negatively charged sites on various amino acid side chains within the protein. The various kinds of stabilizing forces are summarized in Figure 15.6.

Figure 15.6 Kinds of interactions among amino acid side chains that stabilize a protein's tertiary structure.

15.9 Enzymes and Coenzymes

An **enzyme** is a substance—usually a protein—that acts as a catalyst for a biological reaction. Like all catalysts, enzymes don't affect the equilibrium constant of a reaction and can't bring about a chemical change that is otherwise unfavorable. An enzyme acts only to lower the activation energy for a reaction, thereby making the reaction take place more rapidly. Sometimes, in fact, the rate acceleration brought about by enzymes is extraordinary. Millionfold rate increases are common, and the glycosidase enzymes that hydrolyze polysaccharides increase the reaction rate by a factor of more than 10^{17}, changing the time required for the reaction from millions of years to milliseconds.

Unlike many of the catalysts that chemists use in the laboratory, enzymes are usually specific in their action. Often, in fact, an enzyme will catalyze only a single reaction of a single compound, called the enzyme's *substrate.* For example, the enzyme amylase found in the human digestive tract catalyzes only the hydrolysis of starch to yield glucose; cellulose and other polysaccharides are untouched by amylase.

Different enzymes have different specificities. Some, such as amylase, are specific for a single substrate, but others operate on a range of substrates. Papain, for instance, a globular protein of 212 amino acids isolated from papaya fruit, catalyzes the hydrolysis of many kinds of peptide bonds. In fact, it's this ability to hydrolyze peptide bonds that makes papain useful as a meat tenderizer and a cleaner for contact lenses.

$$\text{-(NHCH(R)C(=O)-NHCH(R')C(=O)-NHCH(R'')C(=O))-} \xrightarrow[H_2O]{\text{Papain}} \text{-(NHCH(R)C(=O)OH} + H_2N\text{CH(R')C(=O)-NHCH(R'')C(=O))-}$$

Enzymes are classified into six categories depending on the kind of reaction they catalyze, as shown in Table 15.2. *Oxidoreductases* catalyze oxidations and reductions; *transferases* catalyze the transfer of a group from one substrate to another; *hydrolases* catalyze hydrolysis reactions of esters, amides, and related substrates; *lyases* catalyze the elimination or addition of a small molecule such as H_2O from or to a substrate; *isomerases* catalyze isomerizations; and *ligases* catalyze the bonding together of two molecules, often coupled with the hydrolysis of ATP. The systematic name of an enzyme has two parts, ending with *-ase.* The first part identifies the enzyme's substrate, and the second part identifies its class. For example, hexose kinase is a transferase that catalyzes the transfer of a phosphate group from ATP to a hexose sugar.

Table 15.2 Classification of Enzymes

Class	Some subclasses	Function
Oxidoreductases	Dehydrogenases	Introduction of double bond
	Oxidases	Oxidation
	Reductases	Reduction
Transferases	Kinases	Transfer of phosphate group
	Transaminases	Transfer of amino group
Hydrolases	Lipases	Hydrolysis of ester
	Nucleases	Hydrolysis of phosphate
	Proteases	Hydrolysis of amide
Lyases	Decarboxylases	Loss of CO_2
	Dehydrases	Loss of H_2O
Isomerases	Epimerases	Isomerization of chirality center
Ligases	Carboxylases	Addition of CO_2
	Synthetases	Formation of new bond

In addition to their protein part, most enzymes also contain a small nonprotein part called a *cofactor.* A **cofactor** can be either an inorganic ion, such as Zn^{2+}, or a small organic molecule, called a **coenzyme**. A coenzyme is not a catalyst but is a reactant that undergoes chemical change during the reaction and requires an additional step or series of steps to return it to its initial state.

Many, although not all, coenzymes are derived from *vitamins*—substances that an organism requires for growth but is unable to synthesize and must receive in its diet. Coenzyme A from pantothenate (vitamin B_3), NAD^+ from niacin, FAD from riboflavin (vitamin B_2), tetrahydrofolate from folic acid, pyridoxal phosphate from pyridoxine (vitamin B_6), and thiamin diphosphate from thiamin (vitamin B_1) are examples. Table 15.3 shows the structures of some common coenzymes.

Table 15.3 Structures of Some Common Coenzymes

Adenosine triphosphate—ATP (phosphorylation)

Coenzyme A (acyl transfer)

Nicotinamide adenine dinucleotide—NAD^+ (oxidation/reduction)
($NADP^+$)

Flavin adenine dinucleotide—FAD (oxidation/reduction)

continued

Table 15.3 Structures of Some Common Coenzymes *(continued)*

Tetrahydrofolate (transfer of C_1 units)

S-Adenosylmethionine (methyl transfer)

Lipoic acid (acyl transfer)

Pyridoxal phosphate (amino acid metabolism)

Thiamin diphosphate (decarboxylation)

Biotin (carboxylation)

Problem 15.17 To what classes do the following enzymes belong?

(a) Pyruvate decarboxylase **(b)** Chymotrypsin
(c) Alcohol dehydrogenase

15.10 How Do Enzymes Work? Citrate Synthase

Enzymes work by bringing reactant molecules together, holding them in the orientation necessary for reaction, and providing any acidic or basic sites needed for catalysis. Let's look, for example, at citrate synthase, an enzyme that catalyzes the aldol-like addition of acetyl CoA to oxaloacetate to give citrate (Section 11.11). The reaction is the first step in the citric acid cycle, in which acetyl groups produced by degradation of food molecules are metabolized to yield CO_2 and H_2O. We'll look at the details of the citric acid cycle in Section 17.4.

$$\underset{\text{Oxaloacetate}}{^{-}O_2C{-}CH_2{-}C({=}O){-}CO_2^{-}} + \underset{\text{Acetyl CoA}}{H_3C{-}C({=}O){-}SCoA} \xrightarrow[\text{synthase}]{\text{Citrate}} \underset{\text{Citrate}}{^{-}O_2C{-}CH_2{-}C(OH)(CO_2^{-}){-}CH_2{-}CO_2^{-}} + HSCoA$$

Citrate synthase is a globular protein of 433 amino acids with a deep cleft lined by an array of functional groups that can bind to the substrate oxaloacetate. On binding oxaloacetate, the original cleft closes and another opens up nearby to bind acetyl CoA. This second cleft is also lined by appropriate functional groups, including a histidine at position 274 and an aspartic acid at position 375. The two reactants are now held by the enzyme in close proximity and with a suitable orientation for reaction. Figure 15.7 shows the structure of citrate synthase as determined by X-ray crystallography, along with a close-up of the active site.

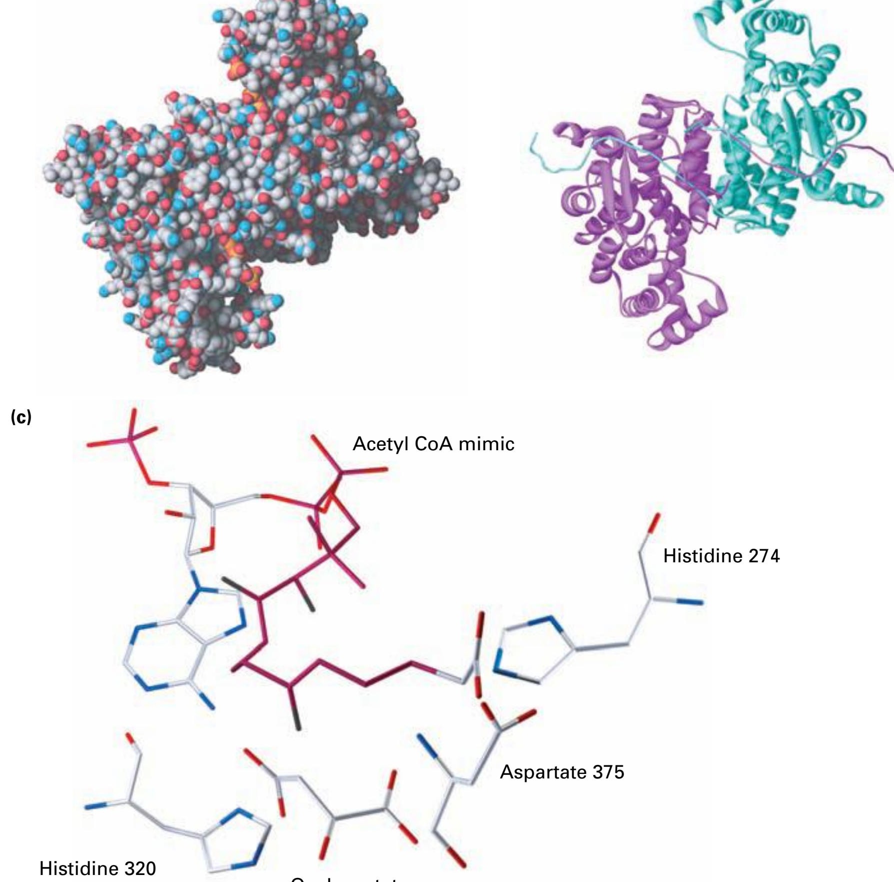


Figure 15.7 X-ray crystal structure of citrate synthase. Part **(a)** is a space-filling model and part **(b)** is a ribbon model, which emphasizes the α-helical segments of the protein chain and indicates that the enzyme is dimeric; that is, it consists of two identical chains held together by hydrogen bonds and other intermolecular attractions. Part **(c)** is a close-up of the active site, in which oxaloacetate and an unreactive acetyl CoA mimic are bound.

As shown in Figure 15.8, the first step in the aldol reaction is generation of the enol of acetyl CoA. The side-chain carboxyl of an aspartate residue acts as base to abstract an acidic α proton, while at the same time the side-chain imidazole ring of a histidine donates H^+ to the carbonyl oxygen. The enol thus produced then does a nucleophilic addition to the ketone carbonyl group of oxaloacetate. The first histidine acts as a base to remove the –OH hydrogen from the enol, while a second histidine residue simultaneously donates a proton to the oxaloacetate carbonyl group, giving citryl CoA. Water then hydrolyzes the thiol ester group in citryl CoA in a nucleophilic acyl substitution reaction, releasing citrate and coenzyme A as the final products.

MECHANISM

Figure 15.8 Mechanism of action of the enzyme citrate synthase.

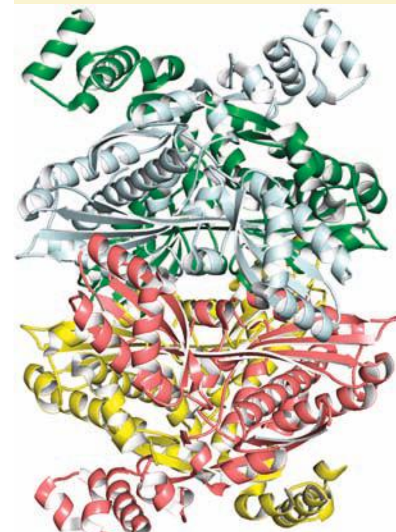
1 The side-chain carboxylate group of an aspartic acid acts as a base and removes an acidic α proton from acetyl CoA, while the N–H group on the side chain of a histidine acts as an acid and donates a proton to the carbonyl oxygen, giving an enol.

2 A histidine deprotonates the acetyl-CoA enol, which adds to the ketone carbonyl group of oxaloacetate in an aldol-like reaction. Simultaneously, an acid N–H proton of another histidine protonates the carbonyl oxygen, producing (*S*)-citryl CoA.

3 The thioester group of citryl CoA is hydrolyzed by a typical nucleophilic acyl substitution reaction to produce citrate plus coenzyme A.

© John McMurry

X-Ray Crystallography



A molecular model of HMG-CoA reductase, an enzyme crucial to the body's synthesis of cholesterol, as determined by X-ray crystallography and downloaded from the Protein Data Bank.

Determining the three-dimensional shape of an object around you is easy—you look at it, your eyes focus the light rays reflected from the object, and your brain assembles the data into a recognizable image. If the object is small, you use a microscope and let the microscope lens focus the visible light. Unfortunately, there is a limit to what you can see, even with the best optical microscope. Called the *diffraction limit,* you can't see anything smaller than the wavelength of light you are using for the observation. Visible light has wavelengths of several hundred nanometers, but atoms in molecules have dimensions on the order of 0.1 nm. Thus, to "see" a molecule—whether a small one in the laboratory or a large, complex enzyme with a molecular weight in the hundreds of thousands—you need wavelengths in the 0.1 nm range, which corresponds to X rays.

Let's say that we want to determine the structure of the enzyme HMG-CoA reductase, which catalyzes a crucial step in the process by which our bodies synthesize cholesterol. The technique used is called *X-ray crystallography.* First, the molecule is crystallized (which often turns out to be the most difficult and time-consuming part of the entire process), and a small crystal with a dimension of 0.4 to 0.5 mm on its longest axis is glued to the end of a glass fiber. The fiber and attached crystal are then mounted in an instrument called an X-ray diffractometer, which consists of a radiation source, a sample positioning and orienting device that can rotate the crystal in any direction, a detector, and a controlling computer.

Once mounted, the crystal is irradiated with X rays, usually so-called Cu$K\alpha$ radiation with a wavelength of 0.154 nm. When the X rays strike the enzyme crystal, they interact with electrons in the molecule and are scattered into a diffraction pattern that, when detected and visualized, appears as a series of bright spots against a null background.

Manipulation of the diffraction pattern to extract three-dimensional molecular data is a complex process, but the final result is that an electron-density map of the molecule is produced. Because electrons are largely localized around atoms, any two centers of electron density located within bonding distance of each other are assumed to represent bonded atoms, leading to a recognizable chemical structure.

So important is structural information for biochemistry that an online database of more than 60,000 biological structures has been created. Operated by Rutgers University and funded by the U.S. National Science Foundation, the Protein Data Bank (PDB) is a worldwide repository for processing and distributing three-dimensional structural data for biological macromolecules.

Summary and Key Words

α-amino acid 505
α helix 522
backbone 512
β-pleated sheet 522
C-terminal amino acid 512
coenzyme 525
cofactor 525
enzyme 524
fibrous protein 521
globular protein 521
isoelectric point, pI 509
N-terminal amino acid 512
peptide 504
primary structure 521
protein 504
quaternary structure 521
residue 511
secondary structure 521
side chain 505
tertiary structure 521
zwitterion 504

Proteins and **peptides** are large biomolecules made of **α-amino acid residues** linked together by amide bonds. Twenty α-amino acids are commonly found in proteins, and all except glycine have stereochemistry similar to that of L sugars.

Determining the structure of a peptide or protein begins with amino acid analysis. The peptide is first hydrolyzed to its constituent α-amino acids, which are then separated and identified. Next, the peptide is sequenced. Edman degradation by treatment with phenyl isothiocyanate (PITC) cleaves one residue from the **N terminus** of the peptide and forms an easily identifiable phenylthiohydantoin (PTH) derivative of that residue. An automated series of Edman degradations can sequence peptide chains up to 50 residues in length.

Peptide synthesis involves the use of protecting groups. An N-protected amino acid with a free $-CO_2H$ group is coupled using DCC to an O-protected amino acid with a free $-NH_2$ group. Amide formation occurs, the protecting groups are removed, and the sequence is repeated. Amines are usually protected as their *tert*-butyloxycarbonyl (Boc) or fluorenylmethyloxycarbonyl (Fmoc) derivatives; acids are usually protected as esters. The synthesis is often carried out by the Merrifield solid-phase method, in which the peptide is bonded to insoluble polymer beads.

Proteins have four levels of structure. **Primary structure** describes a protein's amino acid sequence; **secondary structure** describes how segments of the protein chain orient into regular patterns—either **α helix** or **β-pleated sheet**; **tertiary structure** describes how the entire protein molecule coils into an overall three-dimensional shape; and **quaternary structure** describes how individual protein molecules aggregate into larger structures.

Proteins are classified as either **globular** or **fibrous**. Fibrous proteins such as α-keratin are tough and water-insoluble; globular proteins such as myoglobin are water-soluble and mobile within cells. Most of the 3000 or so known enzymes are globular proteins.

Enzymes are biological catalysts that act by bringing reactant molecules together, holding them in the orientation necessary for reaction and providing any acidic or basic sites needed for catalysis. They are classified into six groups according to the kind of reaction they catalyze: *oxidoreductases* catalyze oxidations and reductions; *transferases* catalyze transfers of groups; *hydrolases* catalyze hydrolysis; *isomerases* catalyze isomerizations; *lyases* catalyze bond breakages; and *ligases* catalyze bond formations.

In addition to their protein part, many enzymes contain **cofactors**, which can be either metal ions or small organic molecules called **coenzymes**. Often, the coenzyme is a vitamin, a small molecule that must be obtained in the diet and is required in trace amounts for proper growth and functioning.

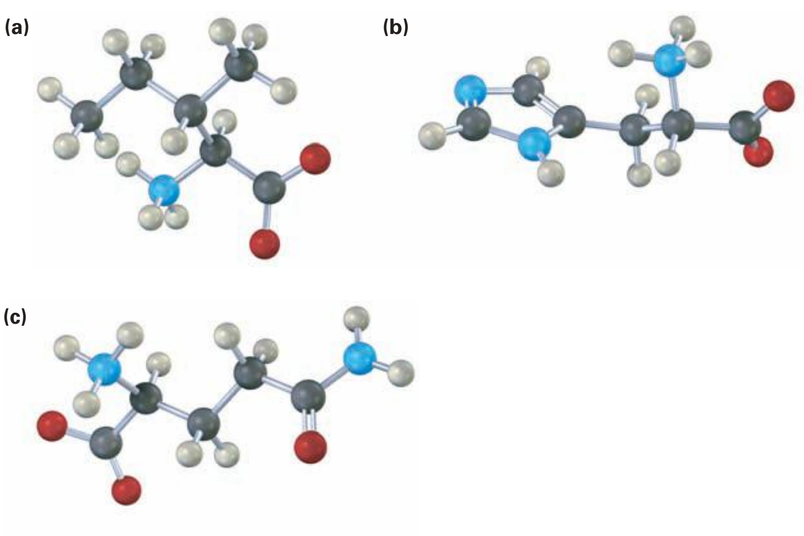
Exercises

Visualizing Chemistry

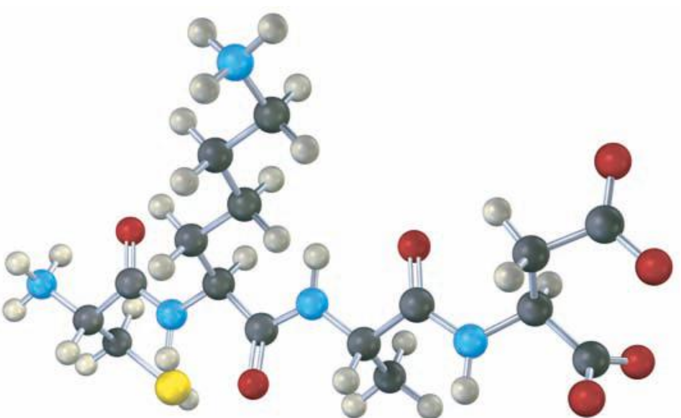
(Problems 15.1–15.17 appear within the chapter.)

Interactive versions of these problems are assignable in OWL.

15.18 Identify the following amino acids:

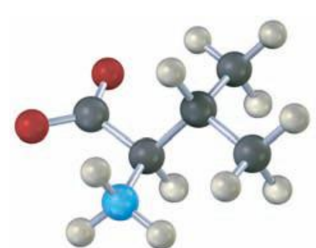


15.19 Give the sequence of the following tetrapeptide (yellow = S):



15.20 Isoleucine and threonine are the only two amino acids with two chirality centers. Assign *R* or *S* configuration to the methyl-bearing carbon atom of isoleucine.

15.21 Is the following molecule a D amino acid or an L amino acid? Identify it.



Additional Problems

Structure and Chirality

15.22 What does the prefix "α" mean when referring to an α-amino acid?

15.23 What amino acids do the following abbreviations stand for?
(a) Ser **(b)** Thr **(c)** Pro **(d)** F **(e)** Q **(f)** D

15.24 Why is cysteine such an important amino acid for determining the tertiary structure of a protein?

15.25 The *endorphins* are a group of naturally occurring compounds in the brain that act to control pain. The active part of an endorphin is a pentapeptide called an *enkephalin,* which has the structure Tyr-Gly-Gly-Phe-Met. Draw the structure.

15.26 Although only *S* amino acids occur in proteins, several *R* amino acids are found elsewhere in nature. For example, (*R*)-serine is found in earthworms and (*R*)-alanine is found in insect larvae. Draw Fischer projections of (*R*)-serine and (*R*)-alanine.

15.27 Draw a Fischer projection of (*S*)-proline, the only secondary amino acid.

15.28 Using both one- and three-letter code names for each amino acid, write the structures of all the peptides containing the following amino acids:
(a) Val, Leu, Ser **(b)** Ser, Leu_2, Pro

15.29 Write full structures for the following peptides, and indicate the positions of the amide bonds:
(a) Val-Phe-Cys **(b)** Glu-Pro-Ile-Leu

15.30 The amino acid threonine, (2*S*,3*R*)-2-amino-3-hydroxybutanoic acid, has two chirality centers and a stereochemistry similar to that of the four-carbon sugar D-threose. Draw a Fischer projection of threonine.

15.31 Draw the Fischer projection of a diastereomer of threonine (see Problem 15.30).

15.32 Draw the following amino acids in their zwitterionic forms:
(a) Serine **(b)** Tyrosine **(c)** Threonine

15.33 Draw structures of the predominant forms of lysine and aspartic acid at pH 3.0 and pH 9.7.

15.34 At what pH would you carry out an electrophoresis experiment if you wanted to separate a mixture of histidine, serine, and glutamic acid? Explain.

15.35 The amino acid analysis data in Figure 15.2 indicate that proline is not easily detected by reaction with ninhydrin. Suggest a reason.

Reactions

15.36 Draw the structure of the phenylthiohydantoin product you would expect to obtain from Edman degradation of the following peptides:
(a) Val-Leu-Gly **(b)** Ala-Pro-Phe

15.37 Predict the product of the reaction of valine with the following reagents:
(a) CH_3CH_2OH, H^+
(b) NaOH, H_2O
(c) Di-*tert*-butyl dicarbonate

15.38 Show the steps involved in a Merrifield synthesis of Phe-Ala-Val.

15.39 When an unprotected α-amino acid is treated with dicyclohexylcarbodiimide (DCC), a 2,5-diketopiperazine results. Explain.

DCC

An α-amino acid A 2,5-diketopiperazine

Peptide Sequencing

15.40 Which amide bonds in the following polypeptide are cleaved by trypsin? By chymotrypsin?

Phe-Leu-Met-Lys-Tyr-Asp-Gly-Gly-Arg-Val-Ile-Pro-Tyr

15.41 Propose a structure for an octapeptide that shows the composition Asp, Gly_2, Leu, Phe, Pro_2, Val on amino acid analysis. Edman analysis shows a glycine N-terminal group, and leucine is the C-terminal group. Acidic hydrolysis gives the following fragments:

Val-Pro-Leu, Gly, Gly-Asp-Phe-Pro, Phe-Pro-Val

15.42 Give the amino acid sequence of hexapeptides that produce the following fragments on partial acid hydrolysis:
(a) Arg, Gly, Ile, Leu, Pro, Val gives Pro-Leu-Gly, Arg-Pro, Gly-Ile-Val
(b) Asp, Leu, Met, Trp, Val_2 gives Val-Leu, Val-Met-Trp, Trp-Asp-Val

15.43 What is the structure of a nonapeptide that gives the following fragments when cleaved by chymotrypsin and by trypsin?

Trypsin cleavage: Val-Val-Pro-Tyr-Leu-Arg, Ser-Ile-Arg

Chymotrypsin cleavage: Leu-Arg, Ser-Ile-Arg-Val-Val-Pro-Tyr

General Problems

15.44 What kinds of reactions do the following classes of enzymes catalyze?
(a) Hydrolases **(b)** Lyases **(c)** Transferases

15.45 What kind of reaction does each of the following enzymes catalyze?
(a) A protease **(b)** A kinase **(c)** A carboxylase

15.46 How can you account for the fact that proline is never encountered in a protein α helix? The α-helical segments of myoglobin and other proteins stop when a proline residue is encountered in the chain.

15.47 Which of the following amino acids are more likely to be found on the outside of a globular protein, and which on the inside? Explain.
(a) Valine **(b)** Aspartic acid **(c)** Isoleucine **(d)** Lysine

15.48 Cysteine is the only amino acid that has L stereochemistry but an *R* configuration. Design another L amino acid of your own making that also has an *R* configuration.

15.49 Arginine, which contains a *guanidino* group in its side chain, is the most basic of the 20 common amino acids. How can you account for this basicity, using resonance structures to see how the protonated guanidino group is stabilized?

NH
H_2N C N H
CO_2^-
H $\overset{+}{N}H_3$
Arginine
Guanidino group

15.50 Look up the structure of human insulin in Section 15.4, and indicate where in each chain the molecule is cleaved by trypsin and by chymotrypsin.

15.51 Propose two structures for a tripeptide that gives Leu, Ala, and Phe on hydrolysis but does not react with phenyl isothiocyanate.

15.52 The reaction of ninhydrin with an α-amino acid occurs in several steps.
(a) The first step is loss of water to give a triketone. Show the mechanism of the reaction and the structure of the triketone.
(b) The second step is formation of an imine by reaction of the amino acid with the triketone. Show its structure.
(c) The third step is a decarboxylation. Show the structure of the product and the mechanism of the decarboxylation reaction.
(d) The fourth step is hydrolysis of an imine to yield an amine and an aldehyde. Show the structures of both products.
(e) The final step is formation of the purple anion. Show the mechanism of the reaction.

2 [Ninhydrin] + $H_2NCH(R)CO_2H$ ⟶ [purple anion] + RCHO + CO_2

Ninhydrin

15.53 Draw as many resonance forms as you can for the purple anion obtained by reaction of ninhydrin with an amino acid (Problem 15.52).

15.54 *Cytochrome c,* an enzyme found in the cells of all aerobic organisms, plays a role in respiration. Elemental analysis of cytochrome *c* reveals it to contain 0.43% iron. What is the minimum molecular weight of this enzyme?

15.55 A hexapeptide with the composition Arg, Gly, Leu, Pro_3 has proline at both C-terminal and N-terminal positions. What is the structure of the hexapeptide if partial hydrolysis gives Gly-Pro-Arg, Arg-Pro, and Pro-Leu-Gly?

15.56 *Aspartame,* a nonnutritive sweetener marketed under the trade name NutraSweet, is the methyl ester of a simple dipeptide, Asp-Phe-OCH_3.

(a) Draw the full structure of aspartame.

(b) The isoelectric point of aspartame is 5.9. Draw the principal structure present in aqueous solution at this pH.

(c) Draw the principal form of aspartame present at physiological pH 7.6.

(d) Show the products of hydrolysis on treatment of aspartame with H_3O^+.

IN THE MEDICINE CABINET

15.57 Leuprolide is a synthetic nonapeptide used to treat both endometriosis in women and prostate cancer in men.

Leuprolide

(a) Both C-terminal and N-terminal amino acids in leuprolide have been structurally modified. Identify the modifications.

(b) One of the nine amino acids in leuprolide has D stereochemistry rather than the usual L. Which one?

(c) Write the structure of leuprolide using both one- and three-letter abbreviations.

(d) What charge would you expect leuprolide to have at neutral pH?

15.58 *Oxytocin,* a nonapeptide hormone secreted by the pituitary gland, stimulates uterine contraction and lactation during childbirth. Its sequence was determined from the following evidence:

1. Oxytocin is a cyclic peptide containing a disulfide bridge between two cysteine residues.
2. When the disulfide bridge is reduced, oxytocin has the constitution Asn, Cys_2, Gln, Gly, Ile, Leu, Pro, Tyr.
3. Partial hydrolysis of reduced oxytocin yields seven fragments:

Asp-Cys, Ile-Glu, Cys-Tyr, Leu-Gly, Tyr-Ile-Glu,
Glu-Asp-Cys, Cys-Pro-Leu

4. Gly is the C-terminal group.
5. Both Glu and Asp are present as their side-chain amides (Gln and Asn) rather than as free side-chain acids.

What is the amino acid sequence of reduced oxytocin? What is the structure of oxytocin?

IN THE FIELD

15.59 The herbicide atrazine binds to an enzyme in weeds by anchoring to two critical residues, a phenylalanine and a serine. Describe three interactions that occur between these residues and the herbicide (see Figure 15.6).