BIOCHEMISTRY: LS2101

Dr. Partha P. Datta

IISER Kolkata

AMINO ACIDS, PEPTIDES, AND PROTEINS

$$\text{H}_3\overset{+}{\text{N}}-\overset{\overset{\displaystyle \text{COO}^-}{|}}{\underset{\underset{\displaystyle \text{R}}{|}}{\text{C}}}-\text{H}$$
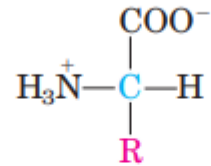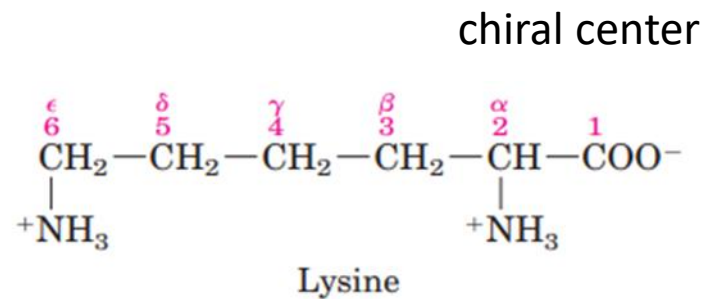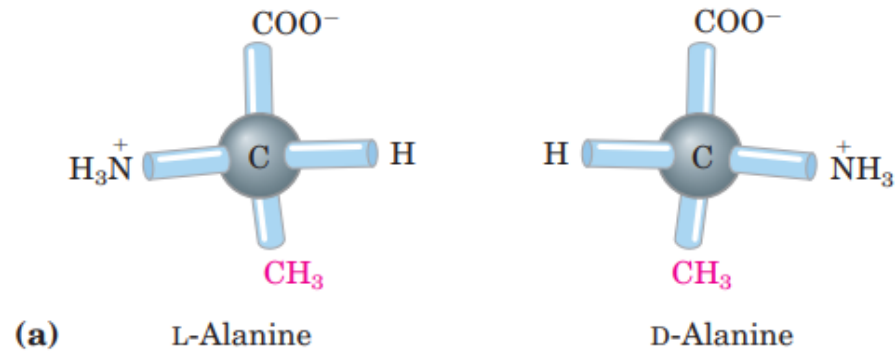
FIGURE 3–2 General structure of an amino acid. This structure is common to all but one of the α-amino acids. (Proline, a cyclic amino acid, is the exception.) The R group or side chain (red) attached to the α carbon (blue) is different in each amino acid.

chiral center

$$\overset{\overset{\displaystyle \epsilon}{6}}{\text{CH}_2}-\overset{\overset{\displaystyle \delta}{5}}{\text{CH}_2}-\overset{\overset{\displaystyle \gamma}{4}}{\text{CH}_2}-\overset{\overset{\displaystyle \beta}{3}}{\text{CH}_2}-\overset{\overset{\displaystyle \alpha}{2}}{\text{CH}}-\overset{1}{\text{COO}}^-$$

$^+\text{NH}_3$ $^+\text{NH}_3$

Lysine

enantiomers

(a)   L-Alanine   D-Alanine

absolute configuration

(b)   L-Alanine   D-Alanine
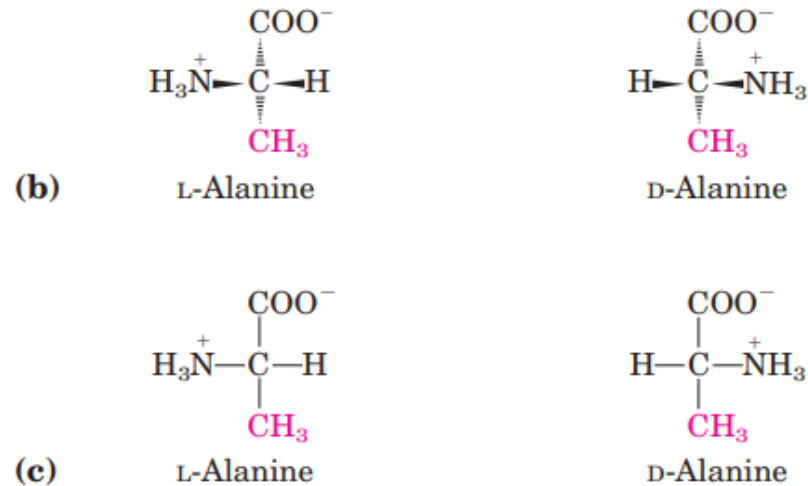
(c)   L-Alanine   D-Alanine

FIGURE 3–3 Stereoisomerism in α-amino acids. (a) The two stereoiso-mers of alanine, L- and D-alanine, are nonsuperimposable mirror im-ages of each other (enantiomers). **(b, c)** Two different conventions for showing the configurations in space of stereoisomers. In perspective formulas **(b)** the solid wedge-shaped bonds project out of the plane of the paper, the dashed bonds behind it. In projection formulas **(c)** the horizontal bonds are assumed to project out of the plane of the paper, the vertical bonds behind. However, projection formulas are often used casually and are not always intended to portray a specific stereochemical configuration.
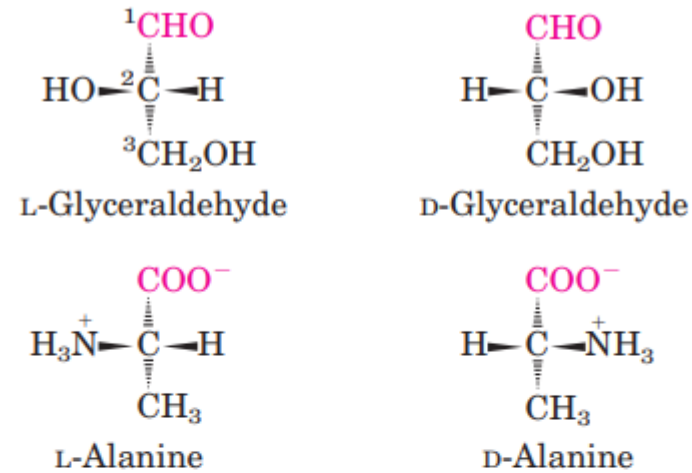
FIGURE 3–4 Steric relationship of the stereoisomers of alanine to the absolute configuration of L- and D-glyceraldehyde. In these perspective formulas, the carbons are lined up vertically, with the chiral atom in the center. The carbons in these molecules are numbered beginning with the terminal aldehyde or carboxyl carbon (red), 1 to 3 from top to bottom as shown. When presented in this way, the R group of the amino acid (in this case the methyl group of alanine) is always below the α carbon. L-Amino acids are those with the α-amino group on the left, and D-amino acids have the α-amino group on the right.

The Amino Acid Residues in Proteins Are L Stereoisomers

TABLE 3–1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r	pK_1 (—COOH)	pK_2 (—NH$_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)[†]
Nonpolar, aliphatic R groups								
Glycine	Gly G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8
Proline	Pro P	115	1.99	10.96		6.48	1.6	5.2
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups								
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups								
Serine	Ser S	105	2.21	9.15		5.68	−0.8	6.8
Threonine	Thr T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups								
Lysine	Lys K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups								
Aspartate	Asp D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	−3.5	6.3

*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 11. From Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.

[†]Average occurrence in more than 1,150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed.), pp. 599–623, Plenum Press, New York.

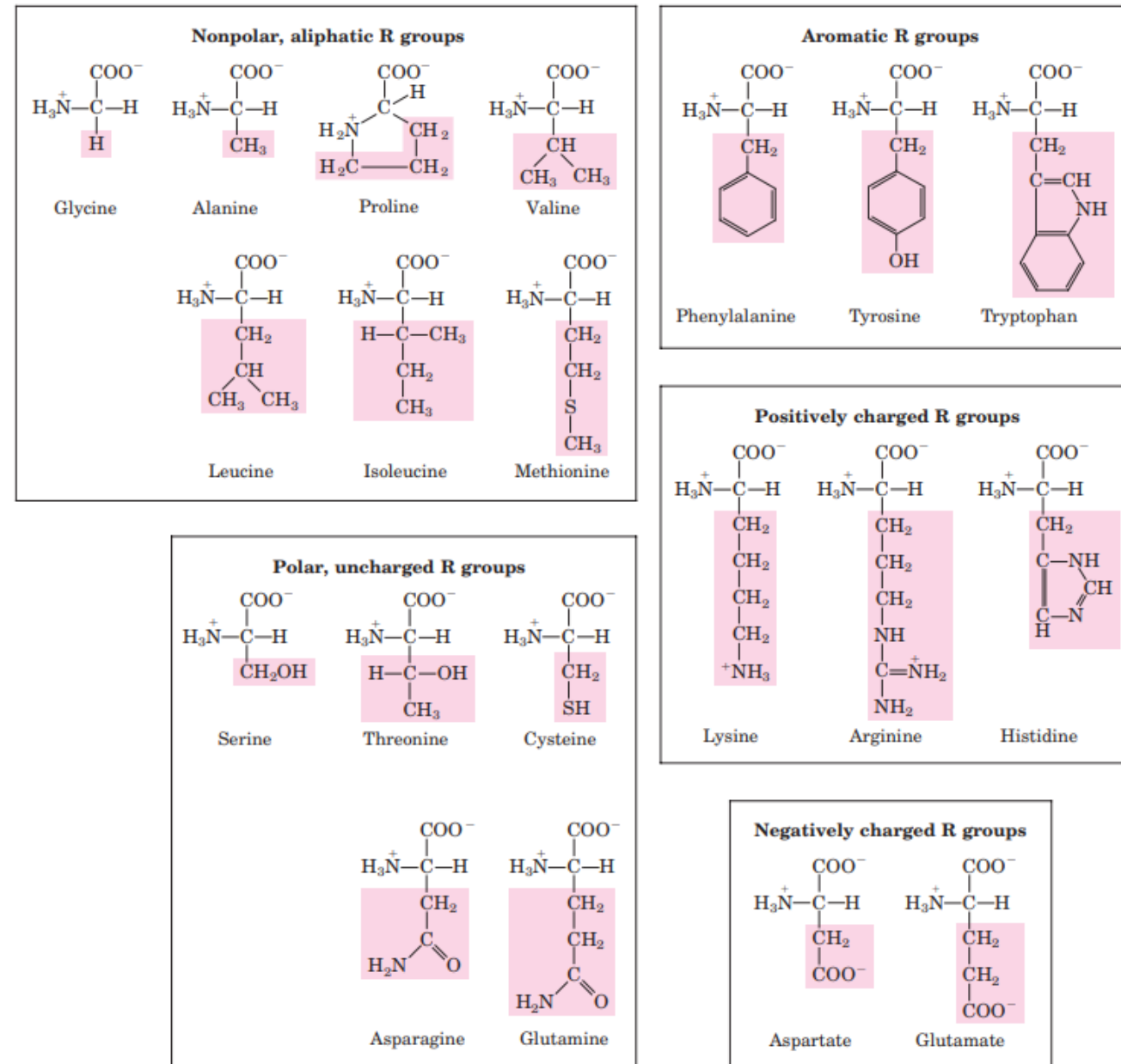# Amino Acids Can Be Classified by R Group

FIGURE 3–5 The 20 common amino acids of proteins. The structural formulas show the state of ionization that would predominate at pH 7.0. The unshaded portions are those common to all the amino acids; the portions shaded in red are the R groups. Although the R group of histidine is shown uncharged, its pK_a (see Table 3–1) is such that a small but significant fraction of these groups are positively charged at pH 7.0.
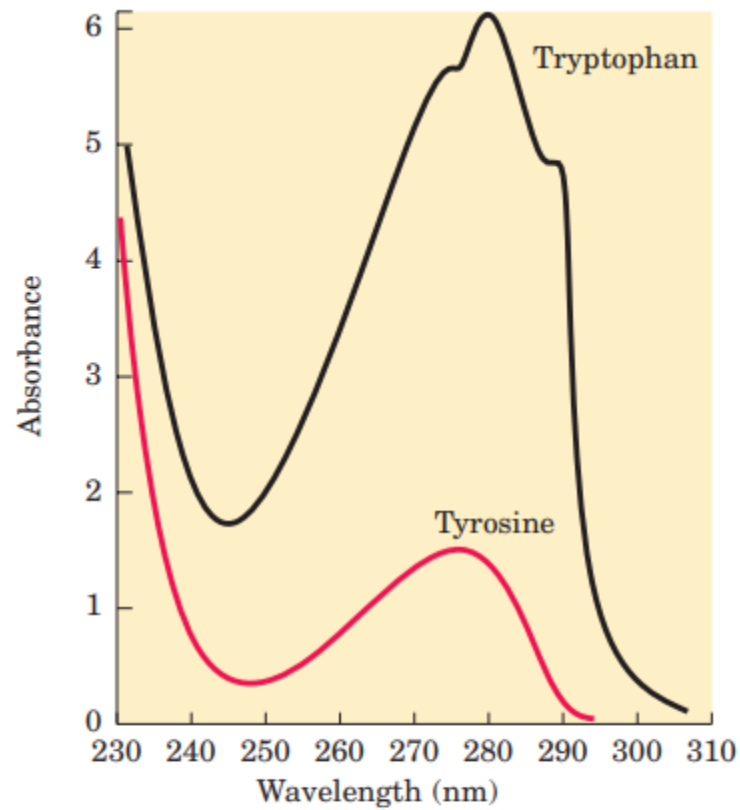
6

FIGURE 3–6 Absorption of ultraviolet light by aromatic amino acids.
Comparison of the light absorption spectra of the aromatic amino acids tryptophan and tyrosine at pH 6.0. The amino acids are present in equimolar amounts (10^{-3} M) under identical conditions. The measured absorbance of tryptophan is as much as four times that of tyrosine. Note that the maximum light absorption for both tryptophan and tyrosine occurs near a wavelength of 280 nm. Light absorption by the third aromatic amino acid, phenylalanine (not shown), generally contributes little to the spectroscopic properties of proteins.
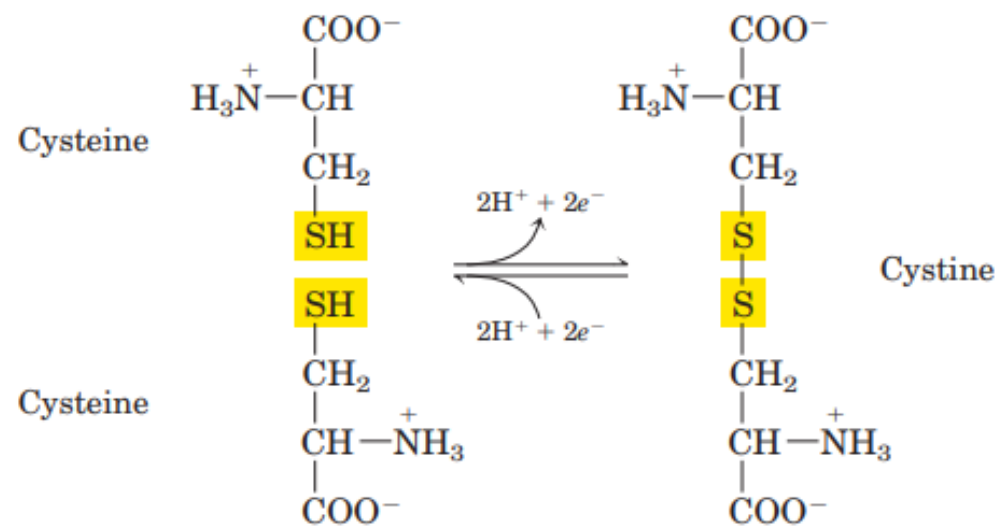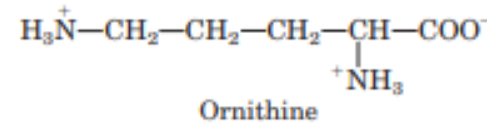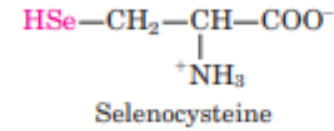
FIGURE 3–7 Reversible formation of a disulfide bond by the oxidation of two molecules of cysteine. Disulfide bonds between Cys residues stabilize the structures of many proteins.

4-Hydroxyproline

5-Hydroxylysine

6-N-Methyllysine

γ-Carboxyglutamate

Desmosine
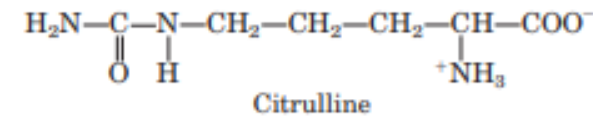
(a) Selenocysteine
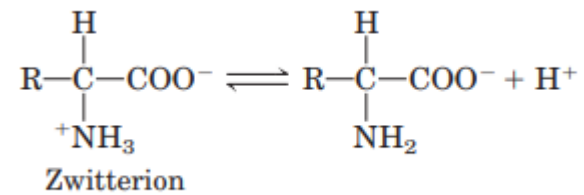
(b)

Ornithine

Citrulline
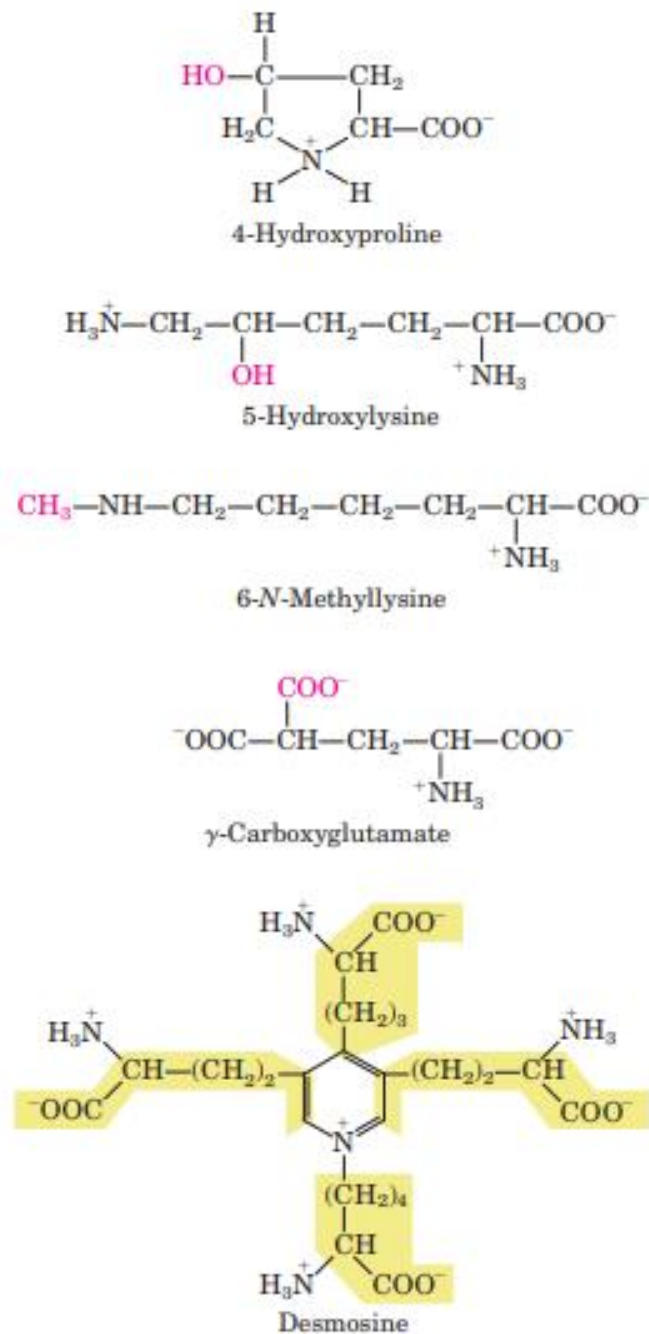
FIGURE 3–8 Uncommon amino acids. (a) Some uncommon amino acids found in proteins. All are derived from common amino acids. Extra functional groups added by modification reactions are shown in red. Desmosine is formed from four Lys residues (the four carbon backbones are shaded in yellow). Note the use of either numbers or Greek letters to identify the carbon atoms in these structures. **(b)** Ornithine and citrulline, which are not found in proteins, are intermediates in the biosynthesis of arginine and in the urea cycle.
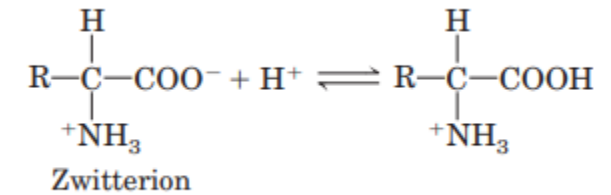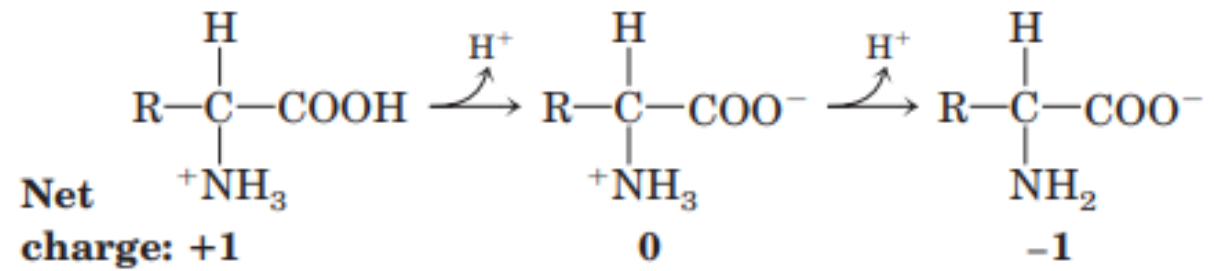
9

Amino Acids Can Act as Acids and Bases

FIGURE 3–9 Nonionic and zwitterionic forms of amino acids. The nonionic form does not occur in significant amounts in aqueous solutions. The zwitterion predominates at neutral pH.

Proton Donor

Proton Acceptor

$$R - \underset{\underset{+NH_3}{|}}{\overset{\overset{H}{|}}{C}} - COOH \xrightarrow{H^+} R - \underset{\underset{+NH_3}{|}}{\overset{\overset{H}{|}}{C}} - COO^- \xrightarrow{H^+} R - \underset{\underset{NH_2}{|}}{\overset{\overset{H}{|}}{C}} - COO^-$$

Net charge: +1   0   −1

Substances having this dual nature are **amphoteric** and are often called **ampholytes**

Amino Acids Have Characteristic Titration Curves

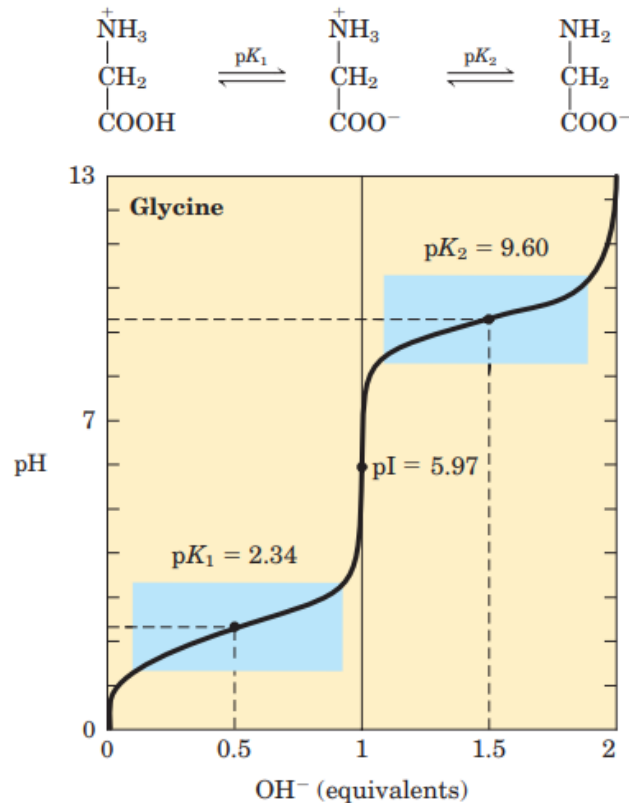Titration Curves Predict the Electric Charge of Amino Acids

The characteristic pH at which the net electric charge is zero is called the isoelectric point or isoelectric pH, designated pI.

For glycine, which has no ionizable group in its side chain, the isoelectric point is simply the arithmetic mean of the two pKa values:

$$pI = \frac{1}{2}(pK_1 + pK_2) = \frac{1}{2}(2.34 + 9.60) = 5.97$$

FIGURE 3–10 **Titration of an amino acid.** Shown here is the titration curve of 0.1 M glycine at 25 °C. The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centered at about $pK_1 = 2.34$ and $pK_2 = 9.60$, indicate the regions of greatest buffering power.
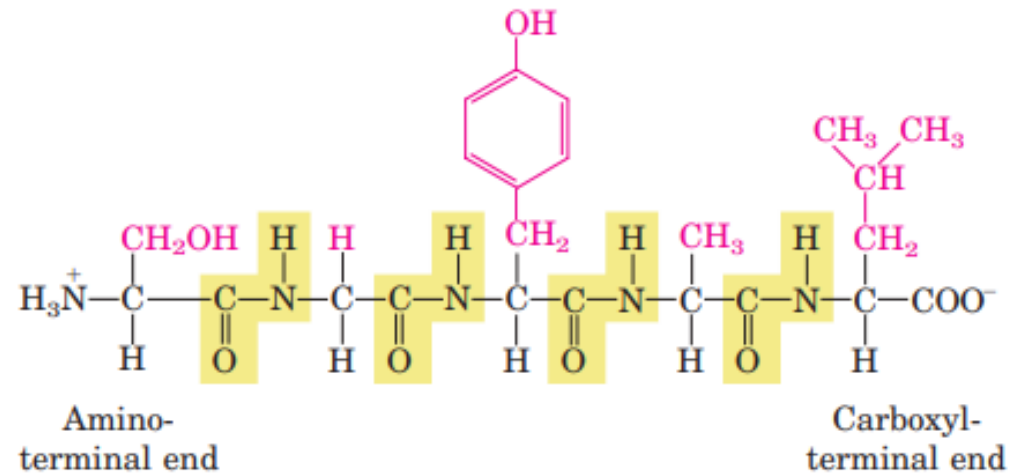
Peptides

FIGURE 3-14 The pentapeptide serylglycyltyrosylalanylleucine, or Ser–Gly–Tyr–Ala–Leu. Peptides are named beginning with the amino-terminal residue, which by convention is placed at the left. The peptide bonds are shaded in yellow; the R groups are in red.
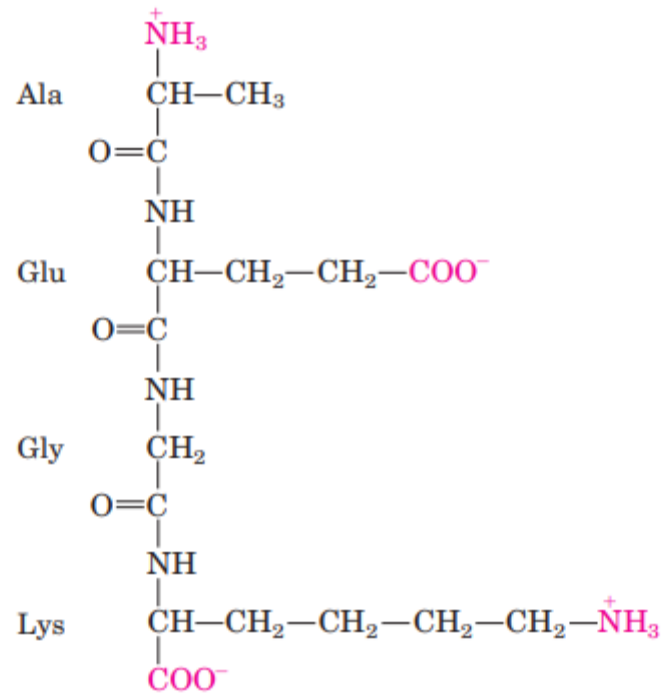
Ala $\overset{+}{N}H_3$ — CH — CH$_3$

$O{=}C$

NH

Glu   CH — CH$_2$ — CH$_2$ — COO$^-$

$O{=}C$

NH

Gly   CH$_2$

$O{=}C$

NH

Lys   CH — CH$_2$ — CH$_2$ — CH$_2$ — CH$_2$ — $\overset{+}{N}H_3$

COO$^-$

FIGURE 3–15 Alanylglutamylglycyllysine. This tetrapeptide has one free α-amino group, one free α-carboxyl group, and two ionizable R groups. The groups ionized at pH 7.0 are in red.

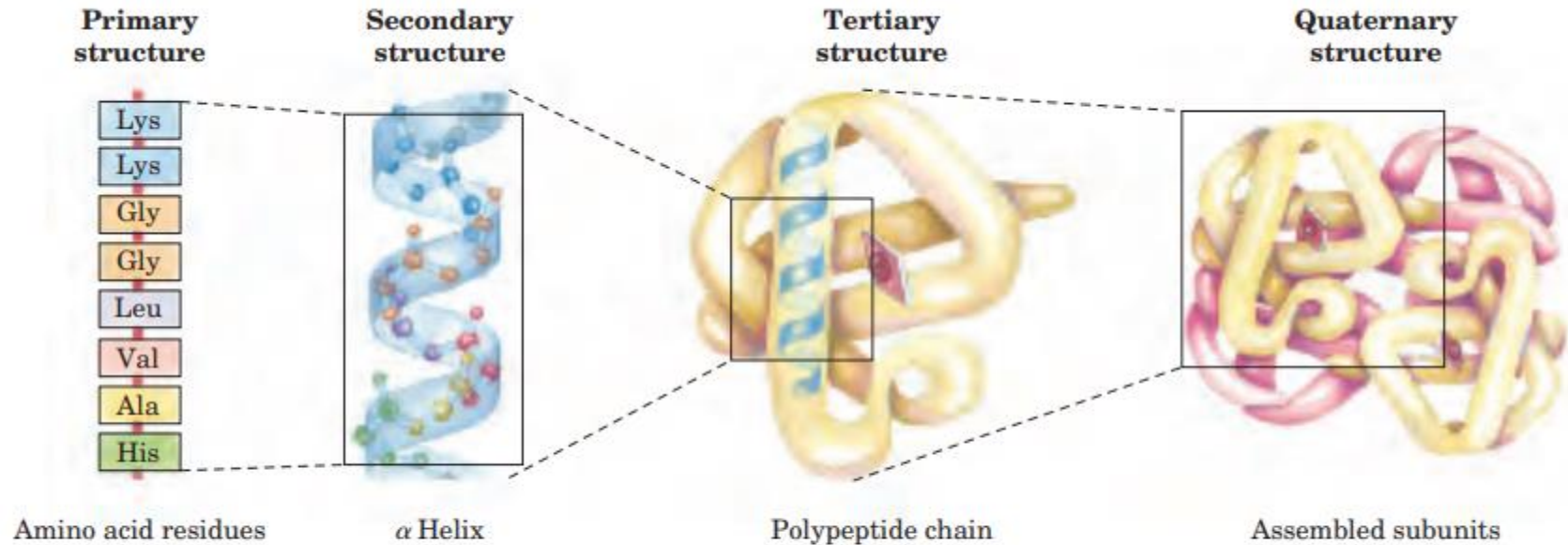| Primary structure	Secondary structure	Tertiary structure	Quaternary structure
Lys Lys Gly Gly Leu Val Ala His			
Amino acid residues	α Helix	Polypeptide chain	Assembled subunits

FIGURE 3–16 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by peptide bonds and includes any disulfide bonds. The resulting polypeptide can be coiled into units of *secondary structure*, such as an α helix. The he-lix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multisubunit protein, in this case hemoglobin.
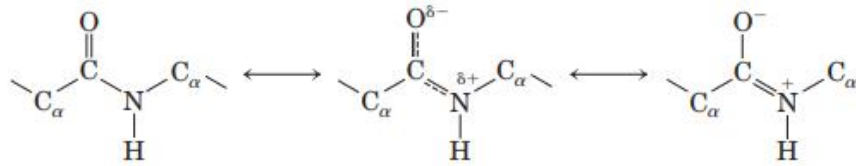
The carbonyl oxygen has a partial negative charge and the amide nitrogen a partial positive charge, setting up a small electric dipole. Virtually all peptide bonds in proteins occur in this trans configuration; an exception is noted in Figure 4–8b.

(a)
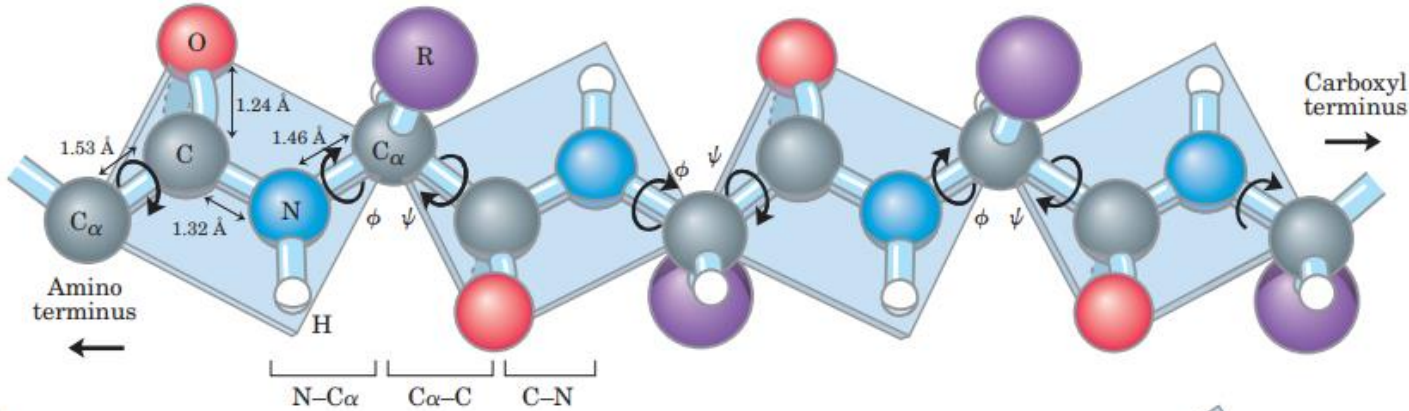
(b)

FIGURE 4–2 **The planar peptide group.** (a) Each peptide bond has some double-bond character due to resonance and cannot rotate. (b) Three bonds separate sequential α carbons in a polypeptide chain. The N—C_α and C_α—C bonds can rotate, with bond angles designated ϕ and ψ, respectively. The peptide C—N bond is not free to rotate. Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. In the conformation shown, ϕ and ψ are 180° (or $-180°$). As one looks out from the α carbon, the ψ and ϕ angles increase as the carbonyl or amide nitrogens (respectively) rotate clockwise. (c) By convention, both ϕ and ψ are defined as 0° when the two peptide bonds flanking that α carbon are in the same plane and positioned as shown. In a protein, this conformation is prohibited by steric overlap between an α-carbonyl oxygen and an α-amino hydrogen atom. To illustrate the bonds between atoms, the balls representing each atom are smaller than the van der Waals radii for this scale. 1 Å = 0.1 nm.

(c)

16
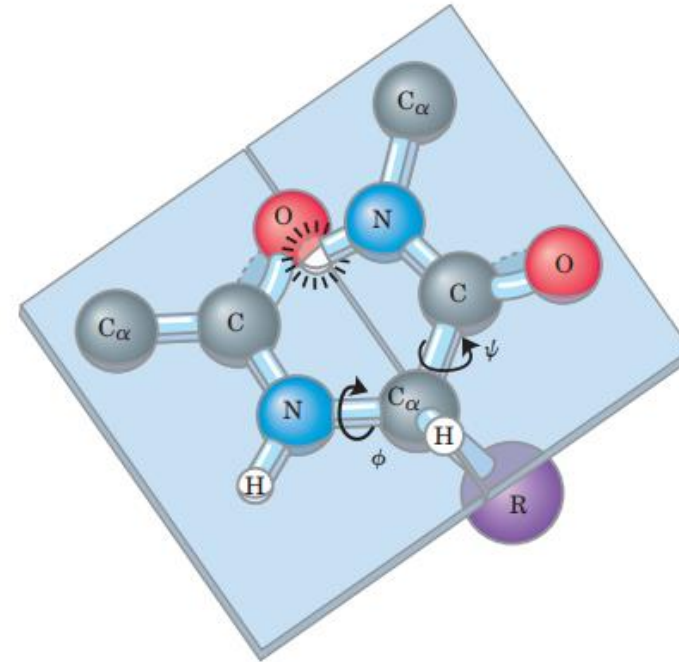
FIGURE 4–3 Ramachandran plot for L-Ala residues. The conformations of peptides are defined by the values of ϕ and ψ. Conformations deemed possible are those that involve little or no steric interference, based on calculations using known van der Waals radii and bond angles. The areas shaded dark blue reflect conformations that involve no steric overlap and thus are fully allowed; medium blue indicates conformations allowed at the extreme limits for unfavorable atomic contacts; the lightest blue area reflects conformations that are permissible if a little flexibility is allowed in the bond angles. The asymmetry of the plot results from the L stereochemistry of the amino acid residues. The plots for other L-amino acid residues with unbranched side chains are nearly identical. The allowed ranges for branched amino acid residues such as Val, Ile, and Thr are somewhat smaller than for Ala. The Gly residue, which is less sterically hindered, exhibits a much broader range of allowed conformations. The range for Pro residues is greatly restricted because ϕ is limited by the cyclic side chain to the range of $-35°$ to $-85°$.
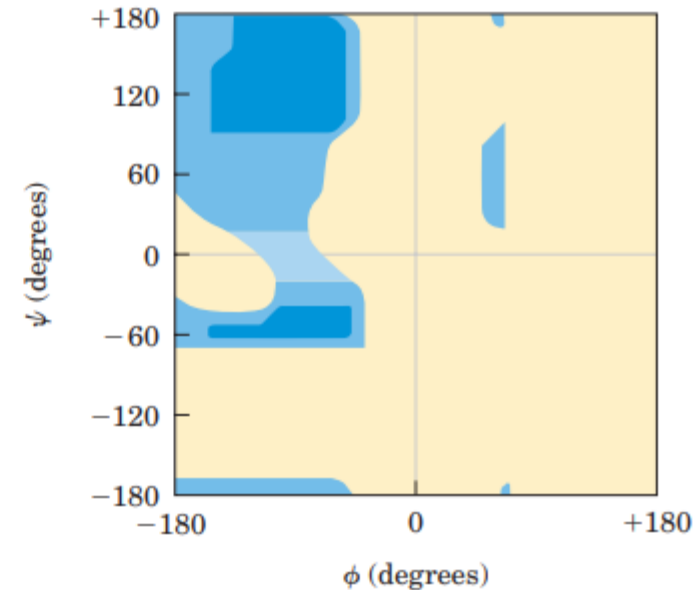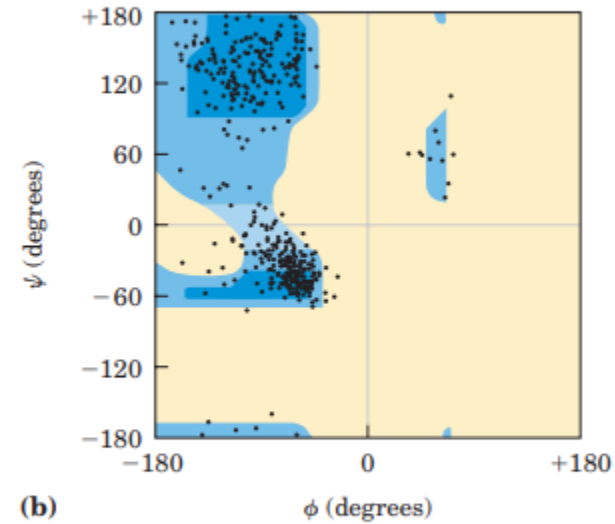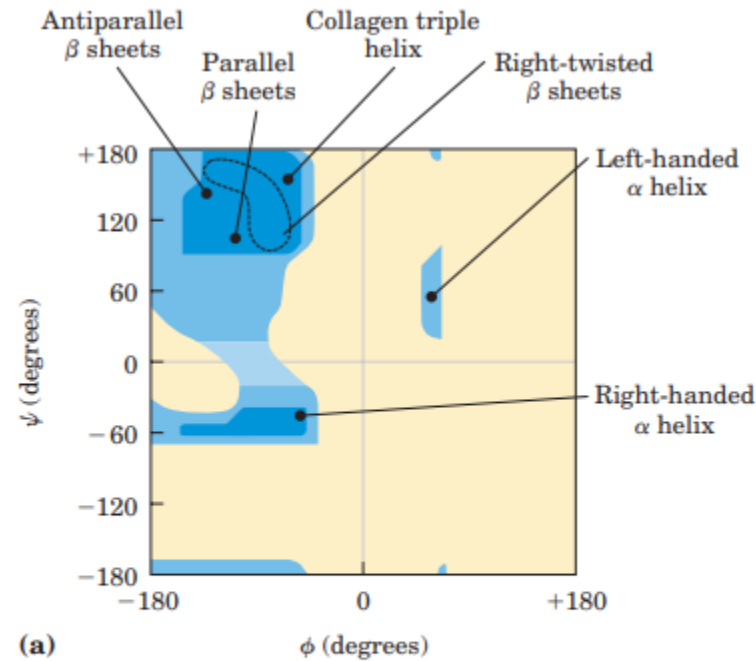
FIGURE 4–9 Ramachandran plots for a variety of structures. (a) The values of ϕ and ψ for various allowed secondary structures are overlaid on the plot from Figure 4–3. Although left-handed α helices extending over several amino acid residues are theoretically possible, they have not been observed in proteins. **(b)** The values of ϕ and ψ for all the amino acid residues except Gly in the enzyme pyruvate kinase (isolated from rabbit) are overlaid on the plot of theoretically allowed conformations (Fig. 4–3). The small, flexible Gly residues were excluded because they frequently fall outside the expected ranges (blue).
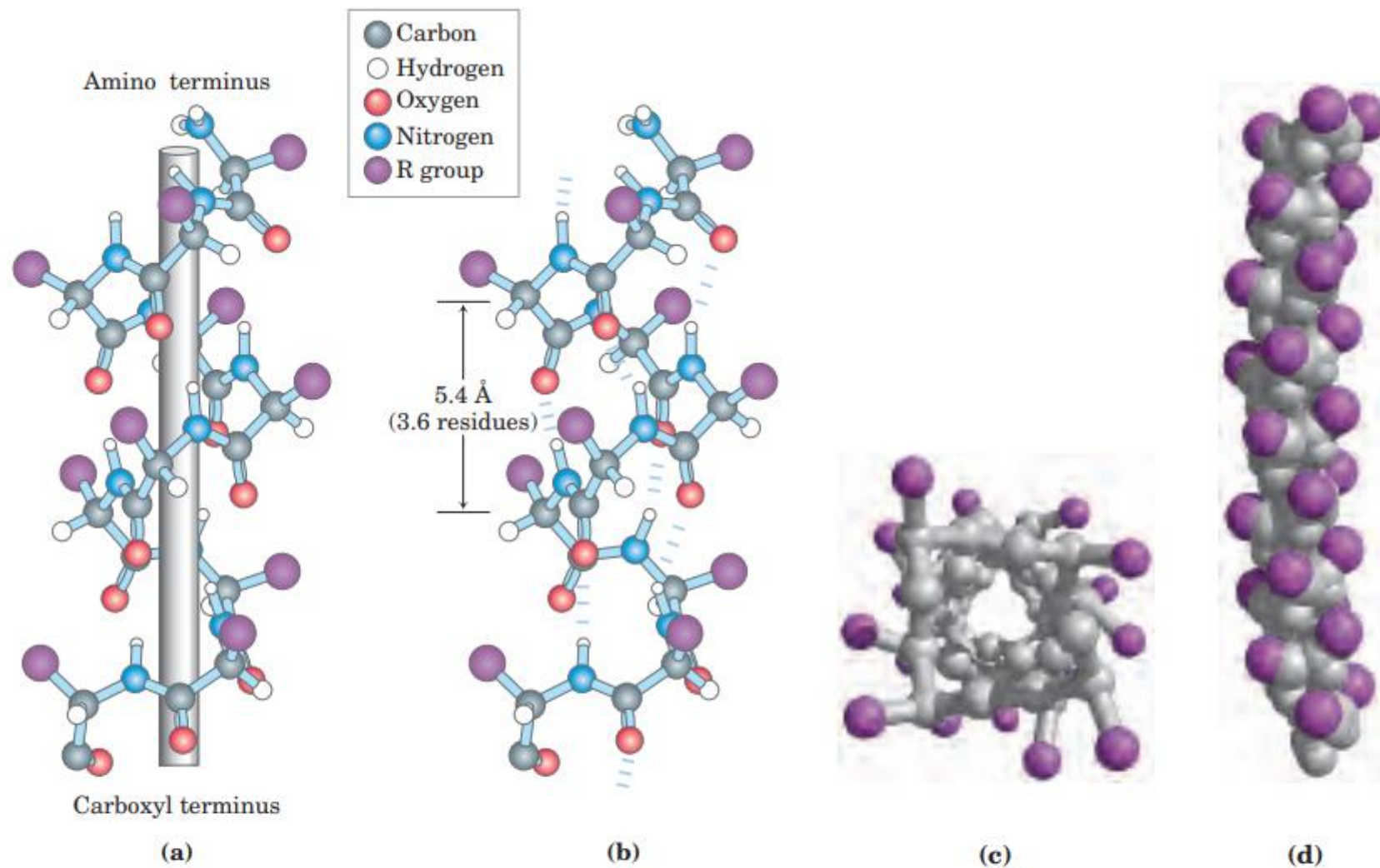
18

Carbon
Hydrogen
Oxygen
Nitrogen
R group

Amino terminus

5.4 Å
(3.6 residues)

Carboxyl terminus

(a)

(b)

(c)

(d)

FIGURE 4–4 Four models of the α helix, showing different aspects of its structure. (a) Formation of a right-handed α helix. The planes of the rigid peptide bonds are parallel to the long axis of the helix, depicted here as a vertical rod. **(b)** Ball-and-stick model of a right-handed α helix, showing the intrachain hydrogen bonds. The repeat unit is a single turn of the helix, 3.6 residues. **(c)** The α helix as viewed from one end, looking down the longitudinal axis (derived from PDB ID 4TNC). Note the positions of the R groups, represented by purple spheres. This ball-and-stick model, used to emphasize the helical arrangement, gives the false impression that the helix is hollow, because the balls do not represent the van der Waals radii of the individual atoms. As the space-filling model **(d)** shows, the atoms in the center of the α helix are in very close contact.

BOX 4–1 WORKING IN BIOCHEMISTRY

Knowing the Right Hand from the Left

There is a simple method for determining whether a helical structure is right-handed or left-handed. Make fists of your two hands with thumbs outstretched and pointing straight up. Looking at your right hand, think of a helix spiraling up your right thumb in the direction in which the other four fingers are curled as shown (counterclockwise). The resulting helix is right-handed. Your left hand will demonstrate a left-handed helix, which rotates in the clockwise direction as it spirals up your thumb.
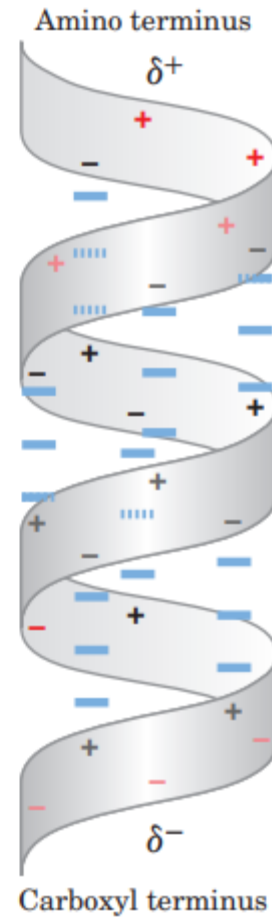
FIGURE 4–6 Helix dipole. The electric dipole of a peptide bond (see Fig. 4–2a) is transmitted along an α-helical segment through the intrachain hydrogen bonds, resulting in an overall helix dipole. In this illustration, the amino and carbonyl constituents of each peptide bond are indicated by $+$ and $-$ symbols, respectively. Non-hydrogen-bonded amino and carbonyl constituents in the peptide bonds near each end of the α-helical region are shown in red.

Keratin α helix

Two-chain coiled coil

Protofilament — 20–30 Å
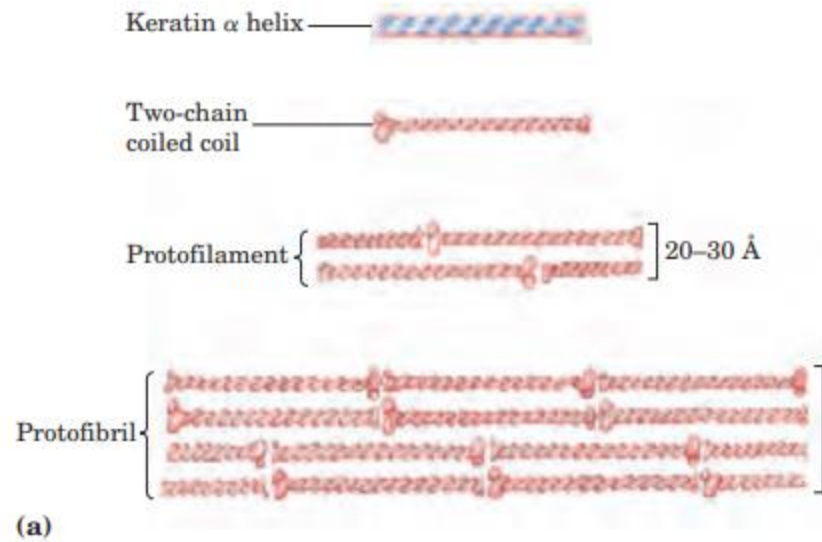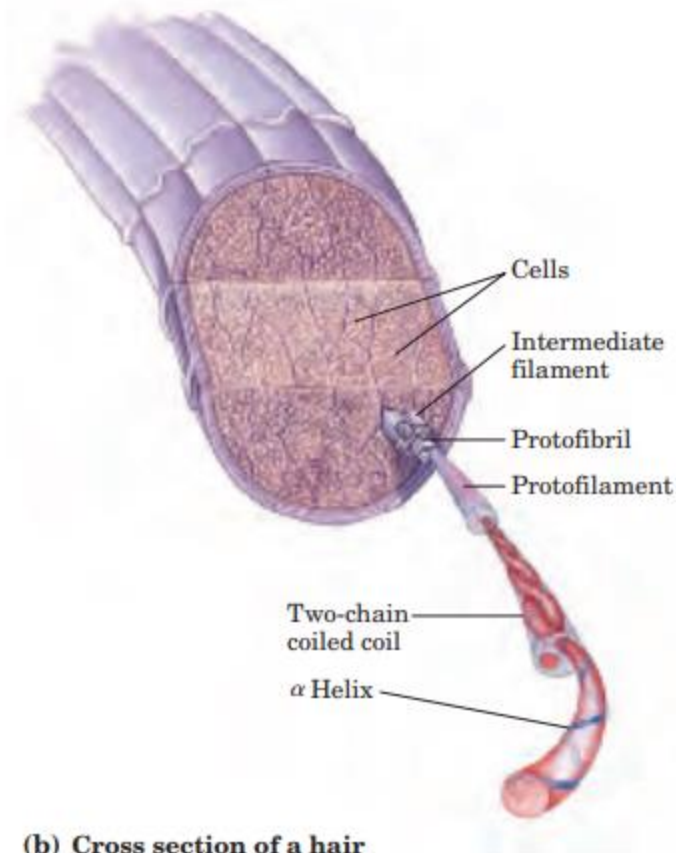
Protofibril

(a)

FIGURE 4–11 Structure of hair. (a) Hair α-keratin is an elongated α helix with somewhat thicker elements near the amino and carboxyl termini. Pairs of these helices are interwound in a left-handed sense to form two-chain coiled coils. These then combine in higher-order structures called protofilaments and protofibrils. About four protofibrils—32 strands of α-keratin altogether—combine to form an intermediate filament. The individual two-chain coiled coils in the various substructures also appear to be interwound, but the handedness of the interwinding and other structural details are unknown. **(b)** A hair is an array of many α-keratin filaments, made up of the substructures shown in **(a).**

Cells

Intermediate filament

Protofibril

Protofilament

Two-chain coiled coil

α Helix

(b) Cross section of a hair

(a) Antiparallel

Top view

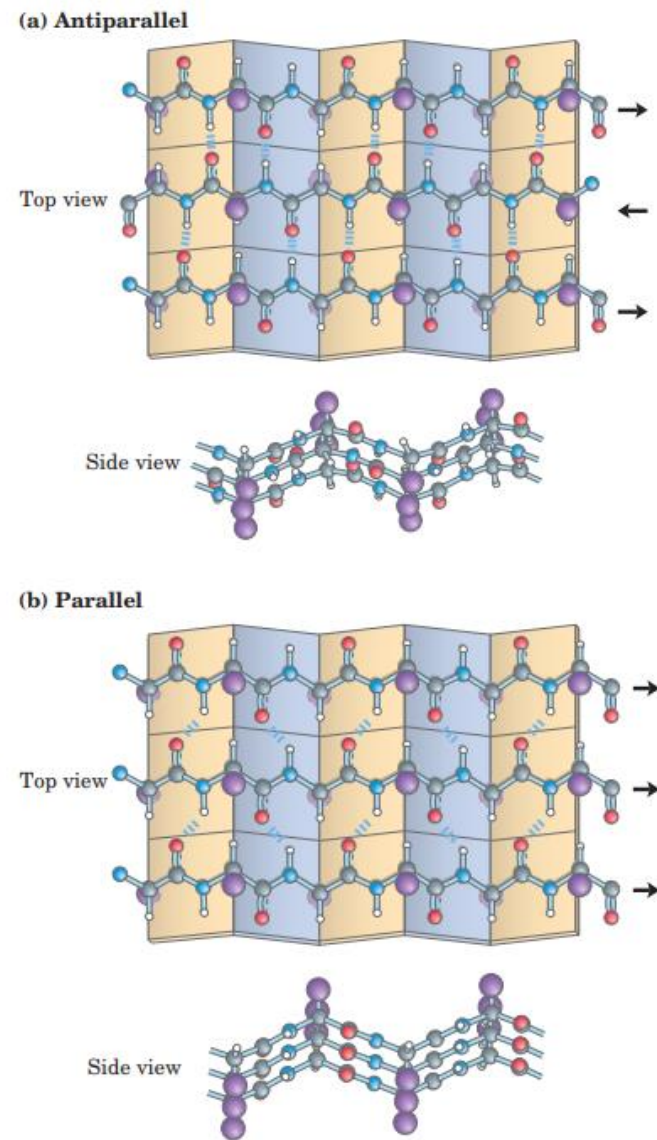Side view

(b) Parallel

Top view

Side view

FIGURE 4–7 The β conformation of polypeptide chains. These top and side views reveal the R groups extending out from the β sheet and emphasize the pleated shape described by the planes of the peptide bonds. (An alternative name for this structure is β-pleated sheet.) Hydrogen-bond cross-links between adjacent chains are also shown. **(a)** Antiparallel β sheet, in which the amino-terminal to carboxyl-terminal orientation of adjacent chains (arrows) is inverse. **(b)** Parallel β sheet.
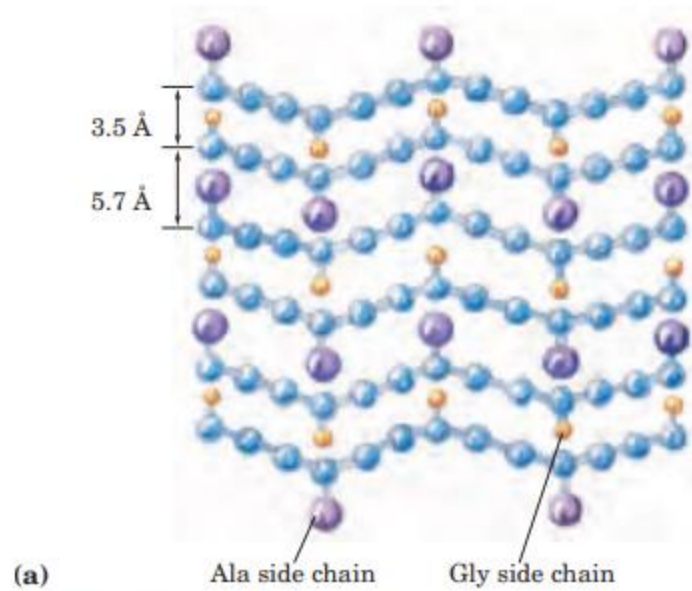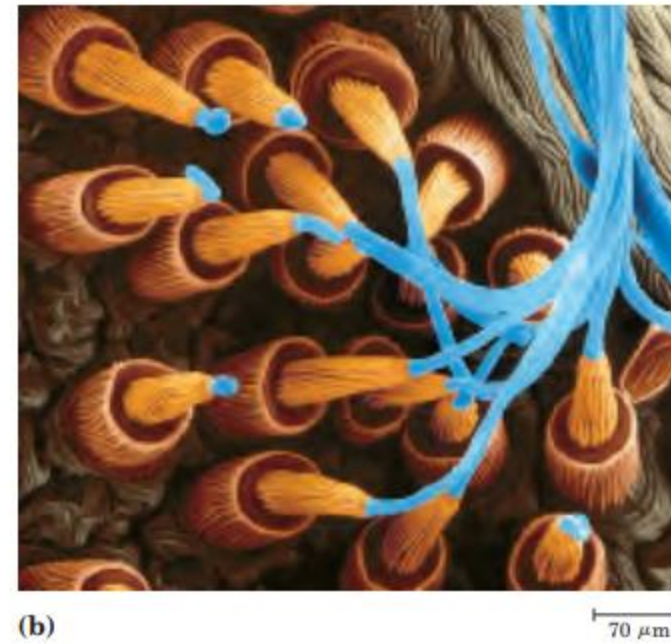
23

(a)   Ala side chain   Gly side chain

3.5 Å

5.7 Å

(b)   70 μm

FIGURE 4–14 Structure of silk. The fibers used to make silk cloth or a spider web are made up of the protein fibroin. **(a)** Fibroin consists of layers of antiparallel β sheets rich in Ala (purple) and Gly (yellow) residues. The small side chains interdigitate and allow close packing of each layered sheet, as shown in this side view. **(b)** Strands of fibroin (blue) emerge from the spinnerets of a spider in this colorized electron micrograph.