Biological Chemistry
Chapter 2

Protein Composition and Structure

Learning objectives:
- structure and properties of amino acids
- properties of peptide bond
- secondary structure
- tertiary structure
- quaternary structure
- protein folding

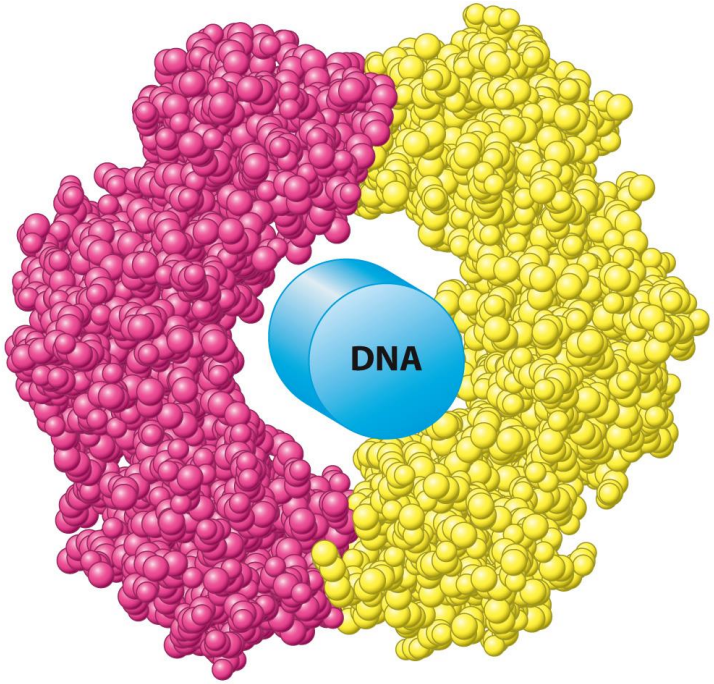


Figure 2.1
Biochemistry, Eighth Edition
© 2015 Macmillan Education

E. coli DNA polymerase III holoenzyme: a sliding DNA clamp.

Proteins:

Greek world *proteos*
versatile bio-macromolecules
carry- out many functions

A) <u>linear polypetides formed by AA</u>

AA are linked through the covalent bond and fold spontaneously into tertiary or quaternary structure

B) <u>Contain a wide range of functional groups</u>

alcohols, thiols, carboxylic acid
heme, flavine adenine dinucleotide,
carbohydrates

C) <u>Interacts with other bio-macromolecules</u>

forming large complexes (ribozome, muscle fibers, chromatin)

A) Some proteins are rigid some are flexible

structural proteins versus signaling proteins

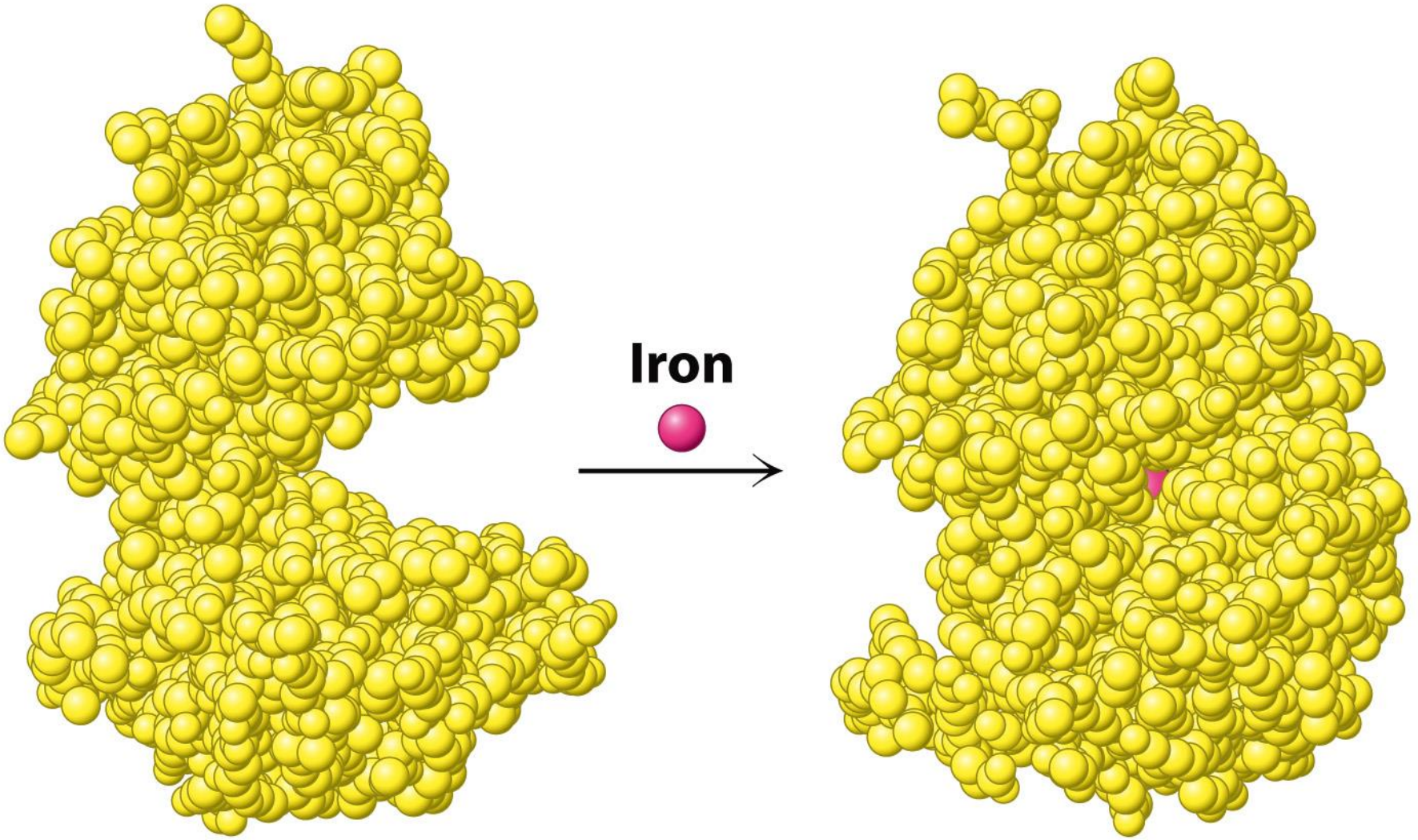


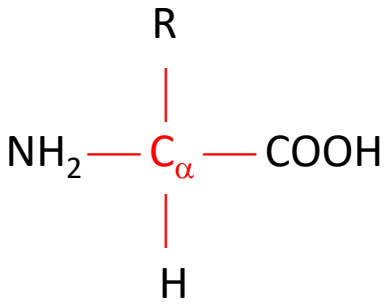
Figure 2.3
Biochemistry, Eighth Edition

Lactoferrin- iron transferring protein
undergoes a substantial change in conformation that allows other molecules to distinguish between the iron-free and the iron-bound forms

Amino acid residues

- Building blocks of proteins
- All proteins are composed of 20 “standard” amino acid residues

General structure of amino acid residue



R group = side chain
varies among AA
α-amino acids (primary amino group is located on Cα)
α-amino acids are chiral, four different groups are connected to Cα

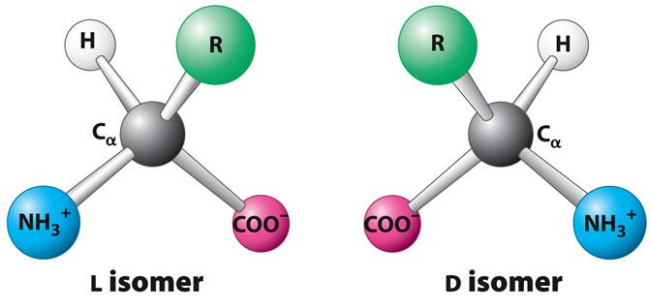


Figure 2.4
Biochemistry, Eighth Edition
© 2015 Macmillan Education

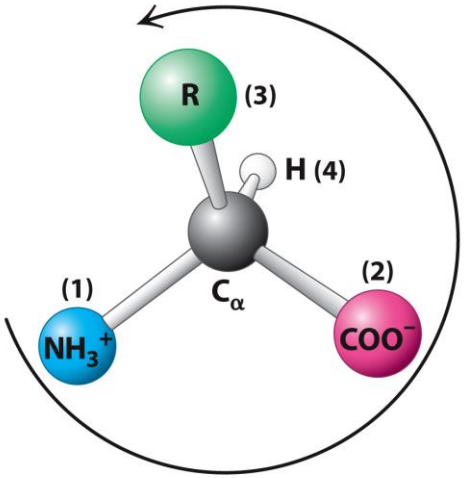


Figure 2.5
Biochemistry, Eighth Edition
© 2015 Macmillan Education

L- isomers are found in proteins
Majority of AA have S absolute configuration

Ionization forms of amino acid residues

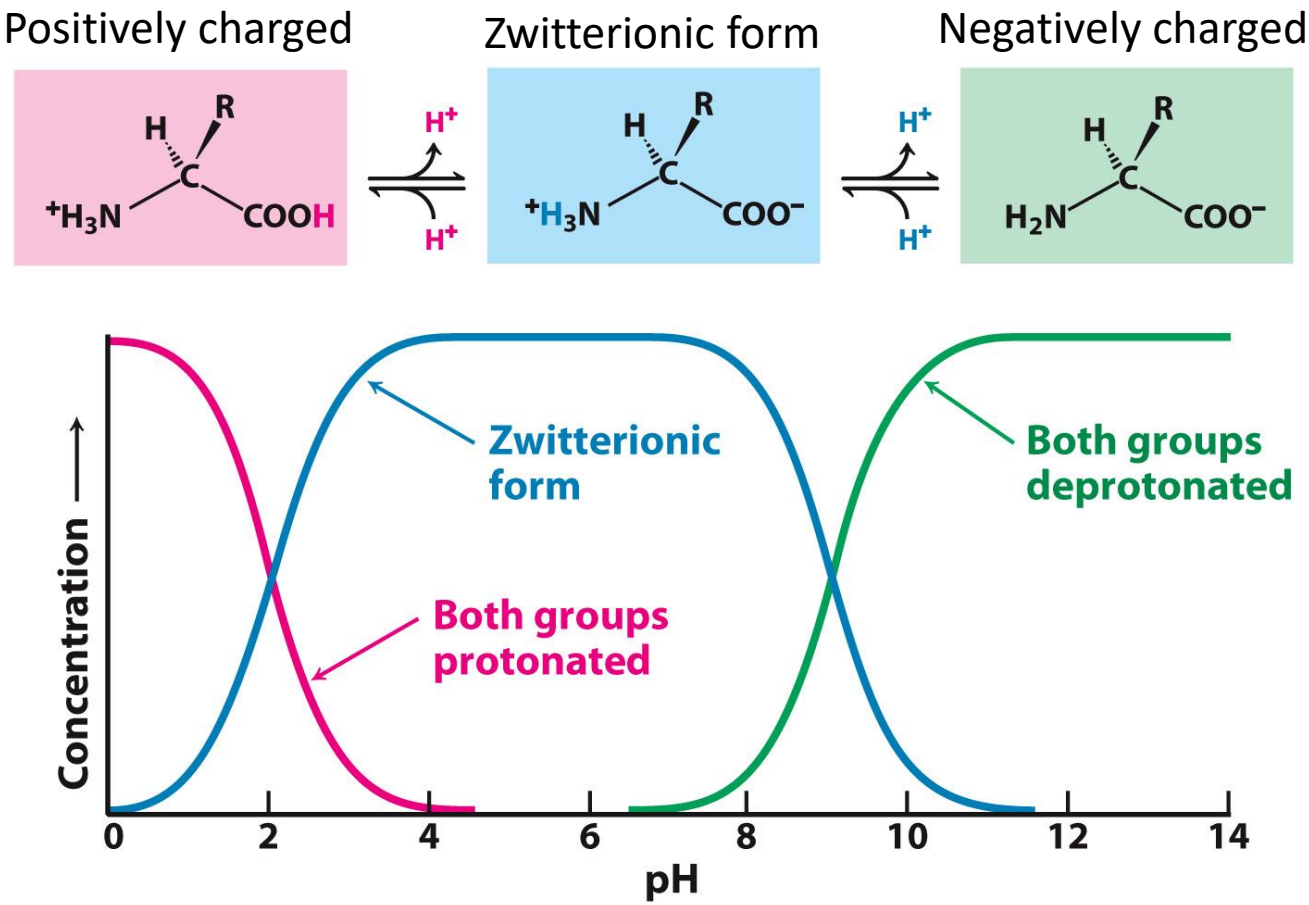


Figure 2.6
Biochemistry, Eighth Edition

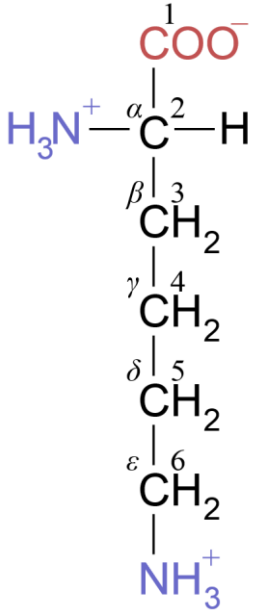
The twenty amino acids found in proteins contain unique side-chains that vary in size, shape, charge, hydrogen-bonding capacity, hydrophobic character, and chemical reactivity.

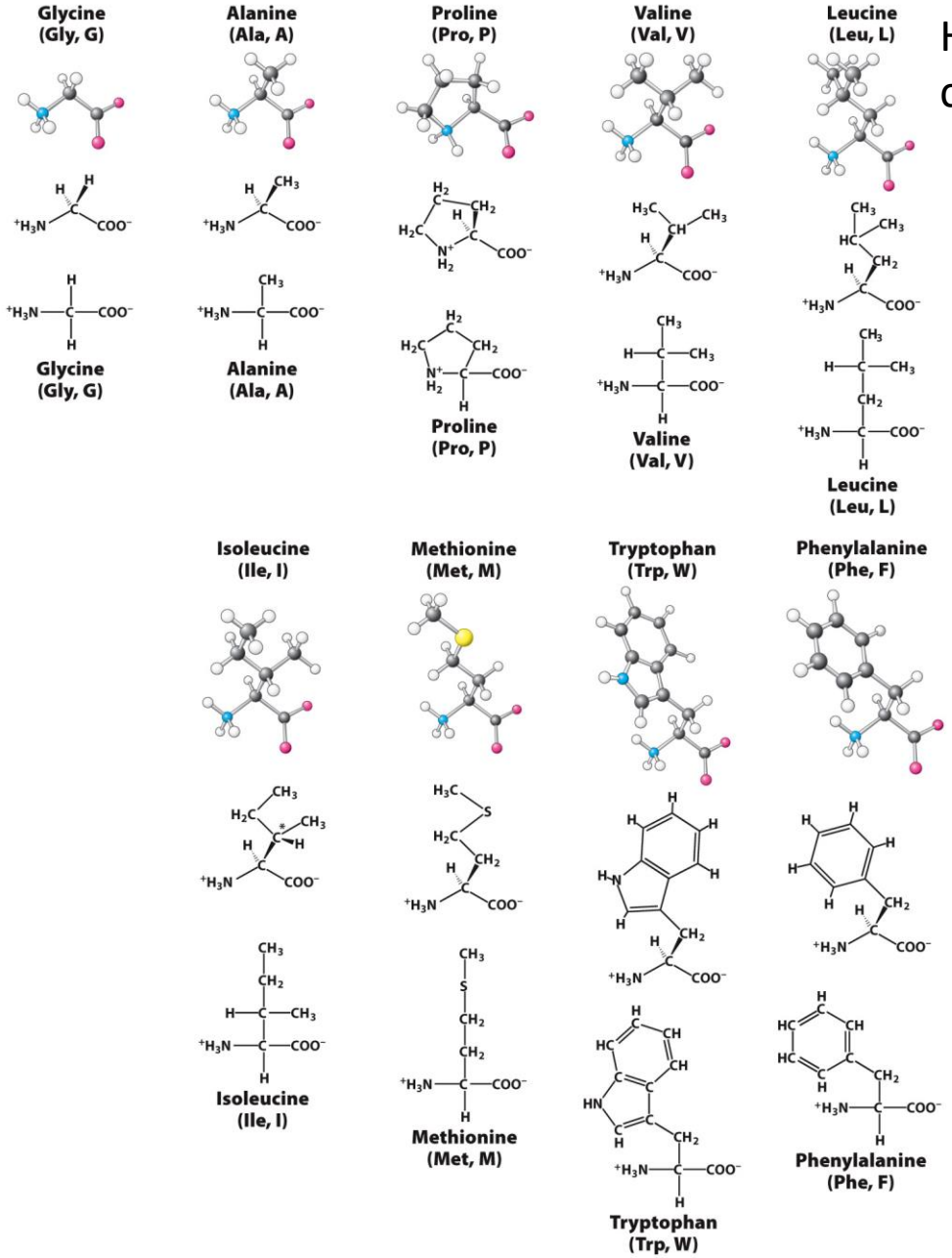
Four classes of amino acids:

hydrophobic (small side chain and large side chain)
polar,
positively charged
negatively charged.

Nomenclature of amino acids
full name (glycin, alanine, glutamin and tyrosin)
three-letter abbreviations code (Gly, Ala, Gln and Tyr);
first three letters of the name
and one-letter symbols (G, A, Q, Y)

Greek nomenclature for amino acids:





Hydrophobic amino acids with nonpolar side chains.

Glycine:	smallest AA, achiral
Methionine:	contains thioether group
Isoleucine:	additional chiral center
Proline:	pyrrolidine group more conformational restricted
Phenylalanine:	aromatic sidechain
Tryptophan	aromatic sidechain

Figure 2.7
Biochemistry, Eighth Edition

Polar amino acid residues – uncharged

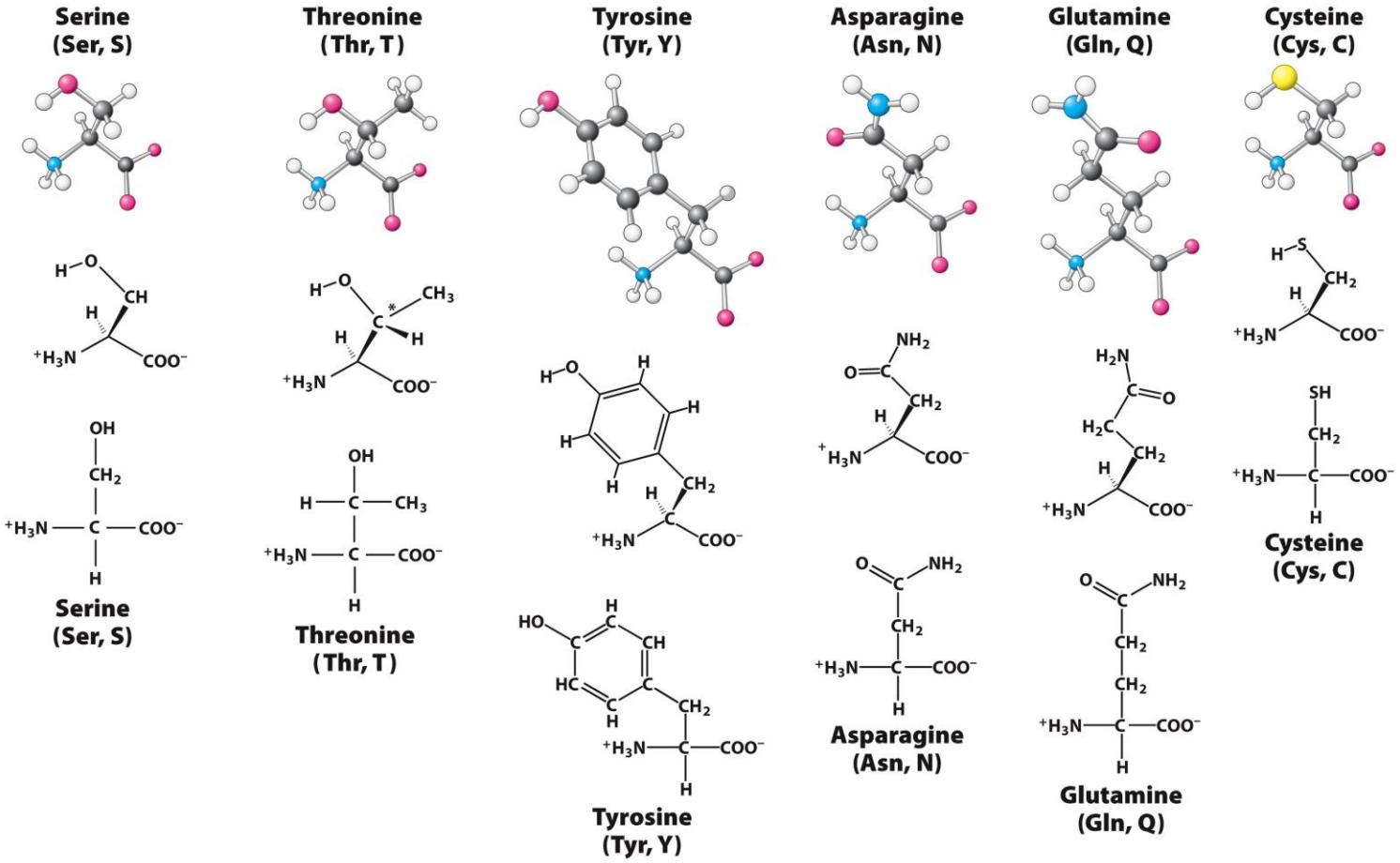


Figure 2.8
Biochemistry, Eighth Edition

Serine; threonine and tyrosine – contains hydroxy group (reactive)
Threonine – 2 chiral centers
Cysteine – thiol group involved in disulfide bridges
Asparagine and glutamine – terminal carboxamide

Positively charged amino acids:

R groups have a positive charge at physiological pH.

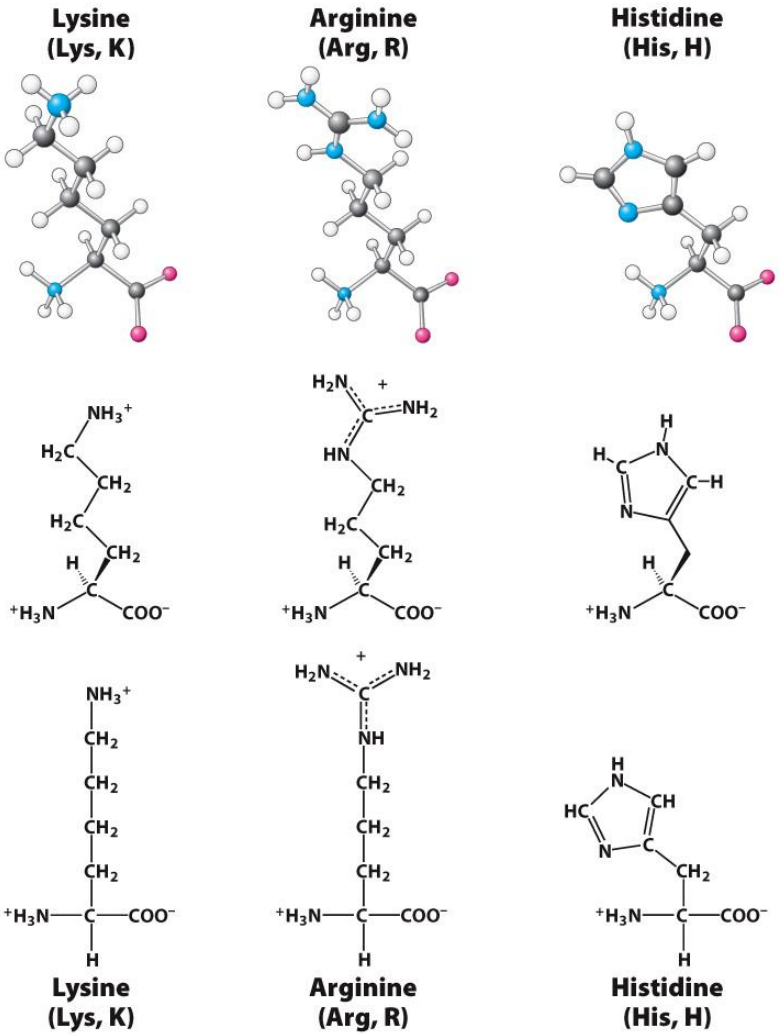


Figure 2.9
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Lysine : primary amino group
Arginine: guanidinium group
Histidine: imidazole group with pKa of side chain ~ 6.0

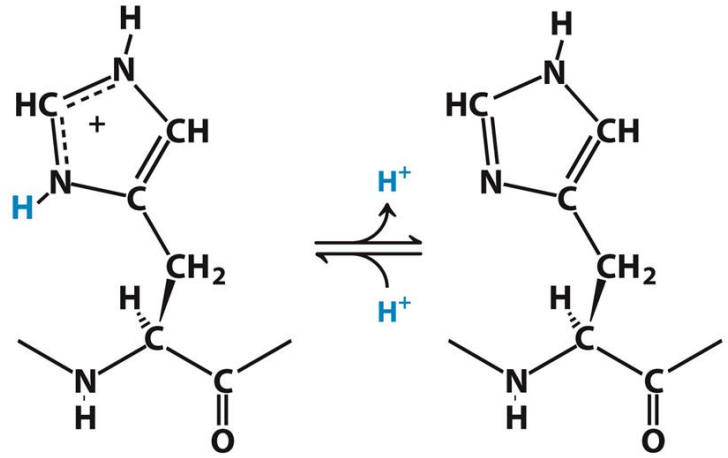


Figure 2.10
Biochemistry, Eighth Edition
© 2015 Macmillan Education

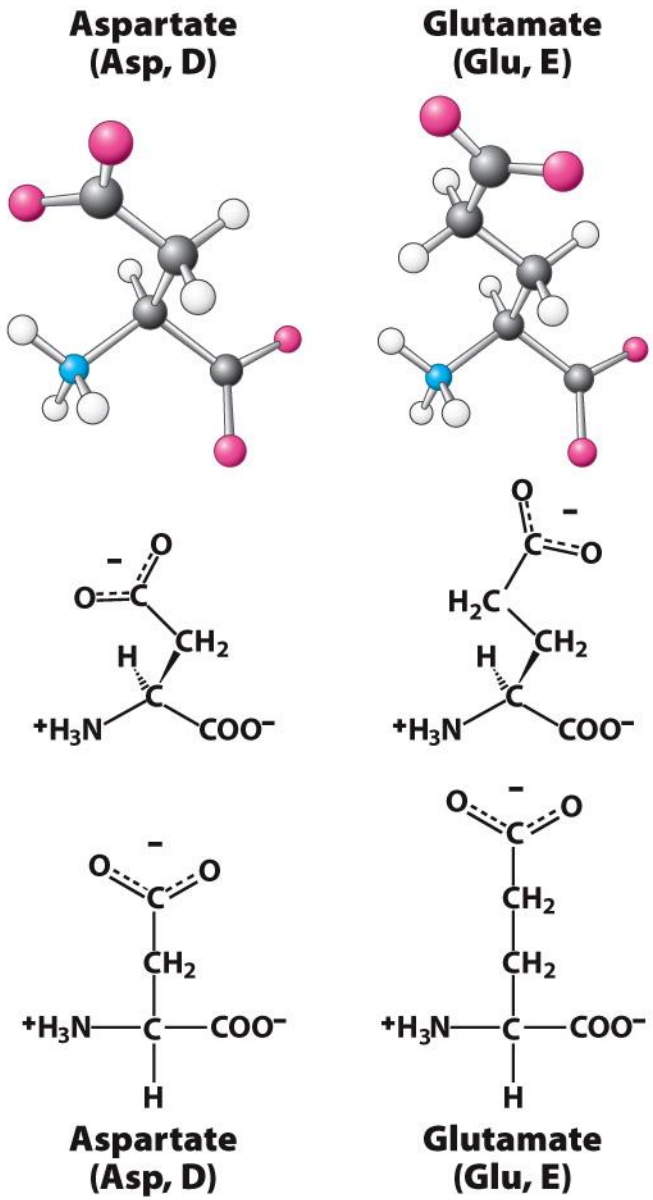


Figure 2.11
Biochemistry, Eighth Edition

Negatively charged amino acid residues

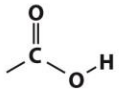
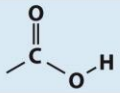
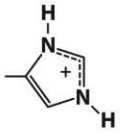
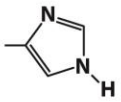
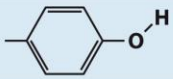
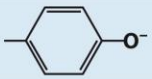
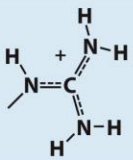
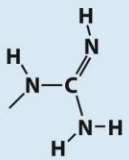
Carboxy group is deprotonated at neutral pH

In some proteins these group can accept protons
(proton acceptable groups in bacterial reaction centers)

Participation in salt bridge

Protonable amino acid residues .

TABLE 2.1 Typical pK_a values of ionizable groups in proteins

Group	Acid		Base	Typical pK_a*
Terminal α-carboxyl group		⇌		3.1
Aspartic acid Glutamic acid		⇌		4.1
Histidine		⇌		6.0
Terminal α-amino group		⇌		8.0
Cysteine		⇌		8.3
Tyrosine		⇌		10.9
Lysine		⇌		10.8
Arginine		⇌		12.5

*pK_a values depend on temperature, ionic strength, and the microenvironment of the ionizable group

Table 2.1
Biochemistry, Eighth Edition

The pKa values can vary significantly dependent on environment and proteins

TABLE 2.2 Abbreviations for amino acids

Amino acid	Three-letter abbreviation	One-letter abbreviation	Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A	Methionine	Met	M
Arginine	Arg	R	Phenylalanine	Phe	F
Asparagine	Asn	N	Proline	Pro	P
Aspartic acid	Asp	D	Serine	Ser	S
Cysteine	Cys	C	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	Trp	W
Glutamic acid	Glu	E	Tyrosine	Tyr	Y
Glycine	Gly	G	Valine	Val	V
Histidine	His	H	Asparagine or aspartic acid	Asx	B
Isoleucine	Ile	I	Glutamine or glutamic acid	Glx	Z
Leucine	Leu	L			
Lysine	Lys	K			

Table 2.2
Biochemistry, Eighth Edition

Why particular 20 amino acid are found in proteins:

1. These amino acids provide chemical versatility.
2. They may have been available for prebiotic reactions.
3. Larger amino acids may be too reactive.
4. Additional two amino acid residues are note encoded by DNA but synthesized by enzymes:
   selenocysteine (in 25 proteins in humans) and pyrrolysine
5. D- amino acid residues are found in bacteria, gramicidine

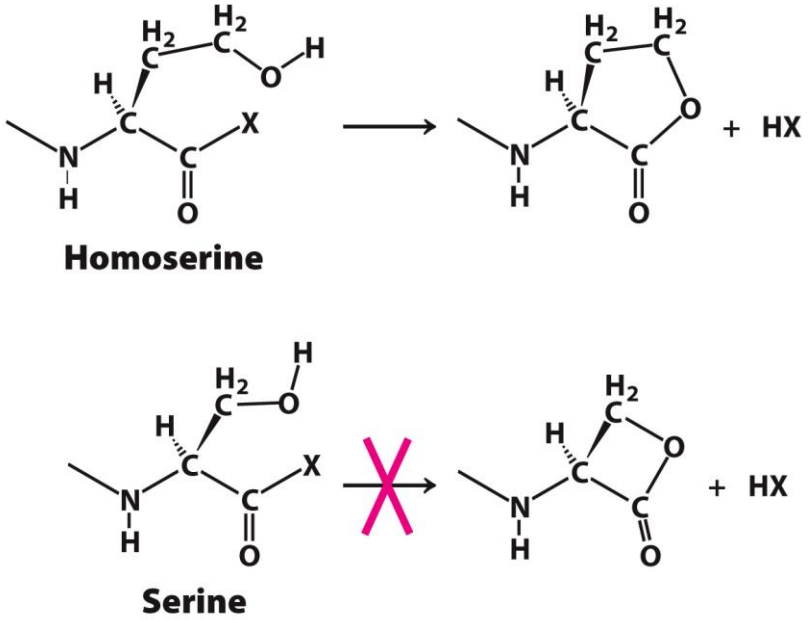


Figure 2.12
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Primary structure of proteins

Primary sequence is the amino acid sequence of the polypeptide chain

Linear polymers formed by linking of the carboxy group from one AA residue and amino group from the second residue.

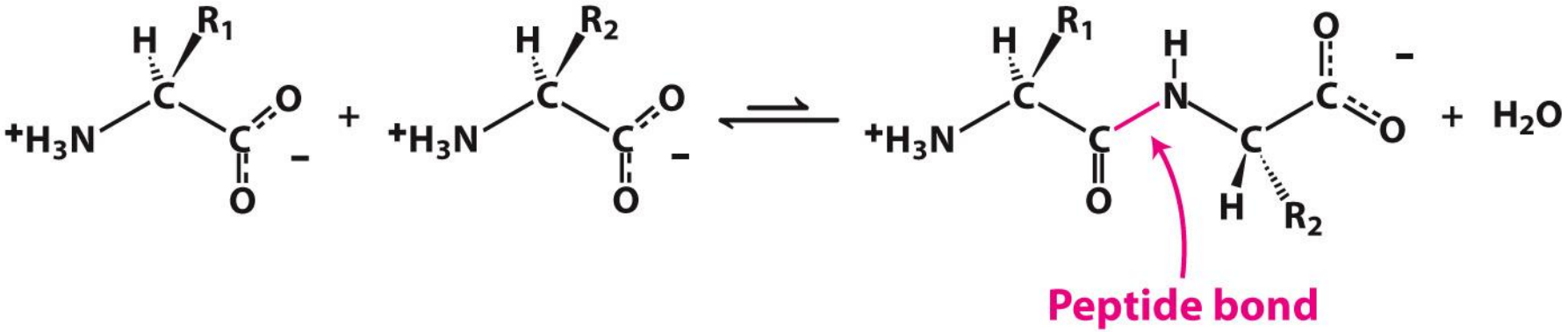


Figure 2.13
Biochemistry, Eighth Edition
© 2015 Macmillan Education

In aqueous solution, the hydrolysis is energetically favorable
Peptide bond synthesis requires input of energy
Hydrolysis is kinetically unfavorable; peptide bond lifetime is 1000 years.

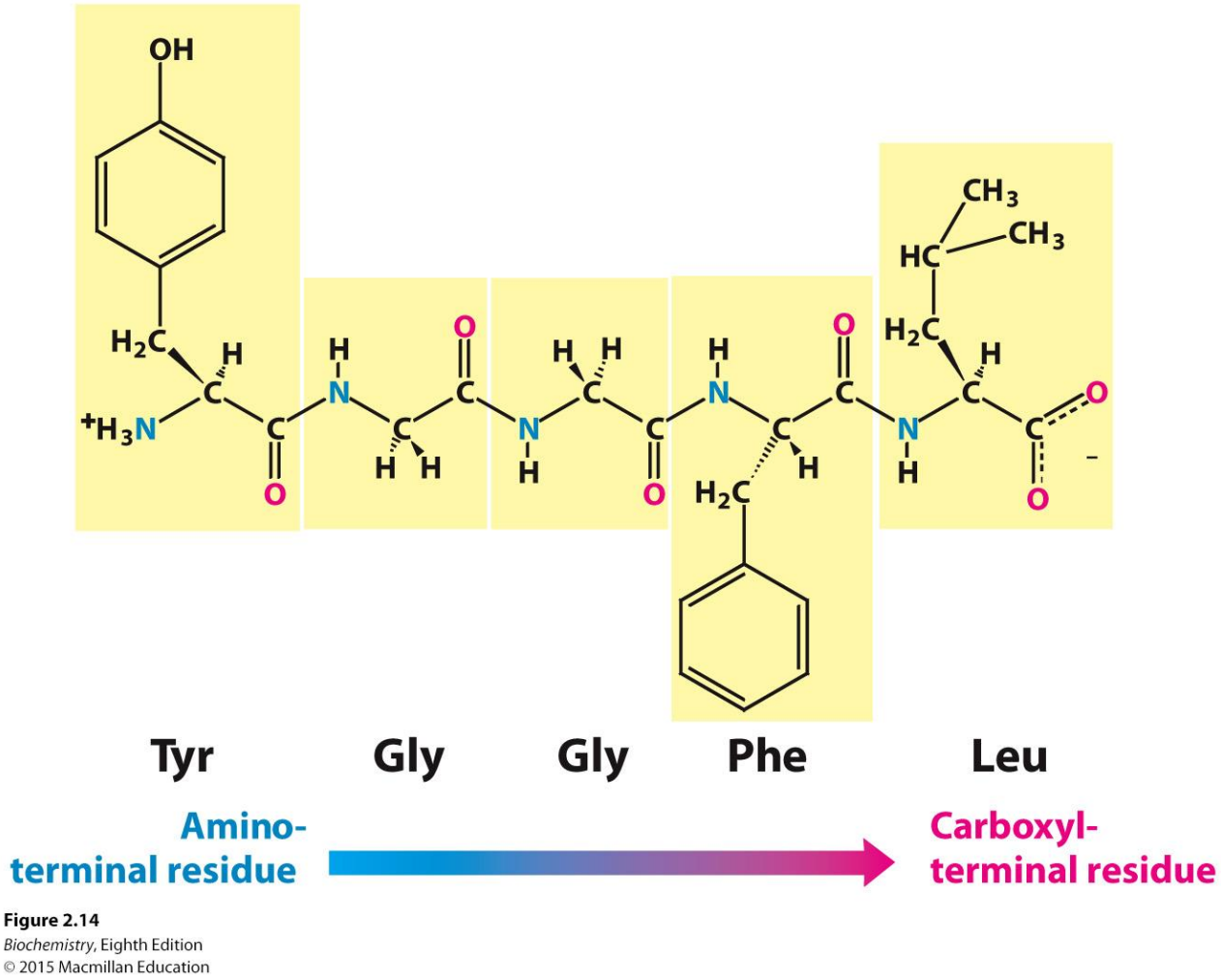


Figure 2.14
Biochemistry, Eighth Edition

N-terminal and C-terminal , amino acids are called residues
Peptide chain has polarity
Sequence is written from N-terminal to C-terminal: YGGFL
Sequence LFGGY represent a different polypeptide

Proteins (polypeptides) have more than 40 amino acid residues
Shorter polypeptides are called peptides
Largest known protein titin (in muscle tissue) has ~ 30,000 amino acid residues
Usually proteins have 100 – 1000 residues

Molecular mass of proteins is in units of Dalton

1Da = 1 g mol^{-1}
Mr (titin) = 2,990 kDa
Mr (myoglobin) = 16 kDa

Protein of 100 amino acid residues:
20^{100} possible sequences ; $1.27x10^{130}$ possible sequences

Restriction on the compositing is done by the fact that no all amino acid residues happen with the same probability:
Trp; Cys; Met and His: rare residues

The polypeptide consists of a repeating part called the **main chain** or **backbone** and a variable part consisting of the distinctive amino acid side chains.

The backbone has hydrogen bonding potential because of the carbonyl groups and hydrogen atoms that are bonded to the nitrogen of the amine group.

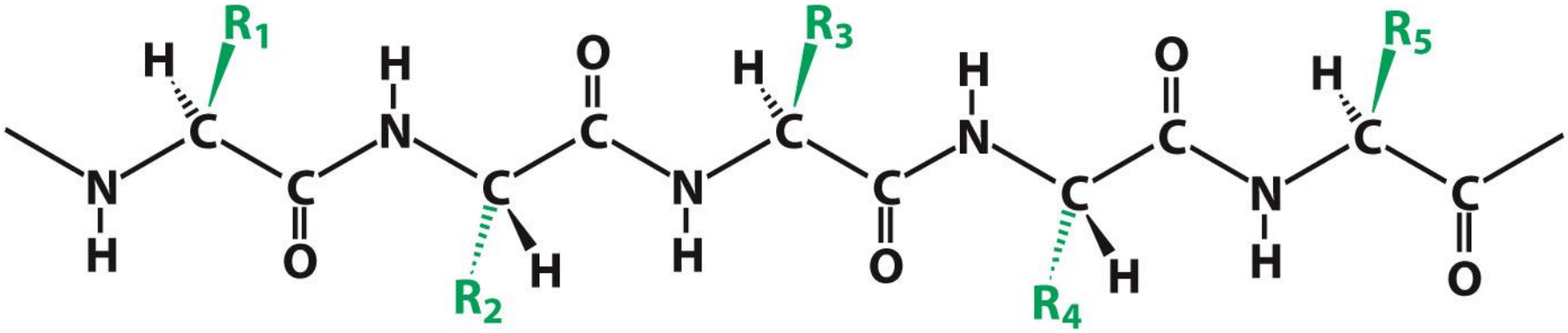


Figure 2.15
Biochemistry, Eighth Edition

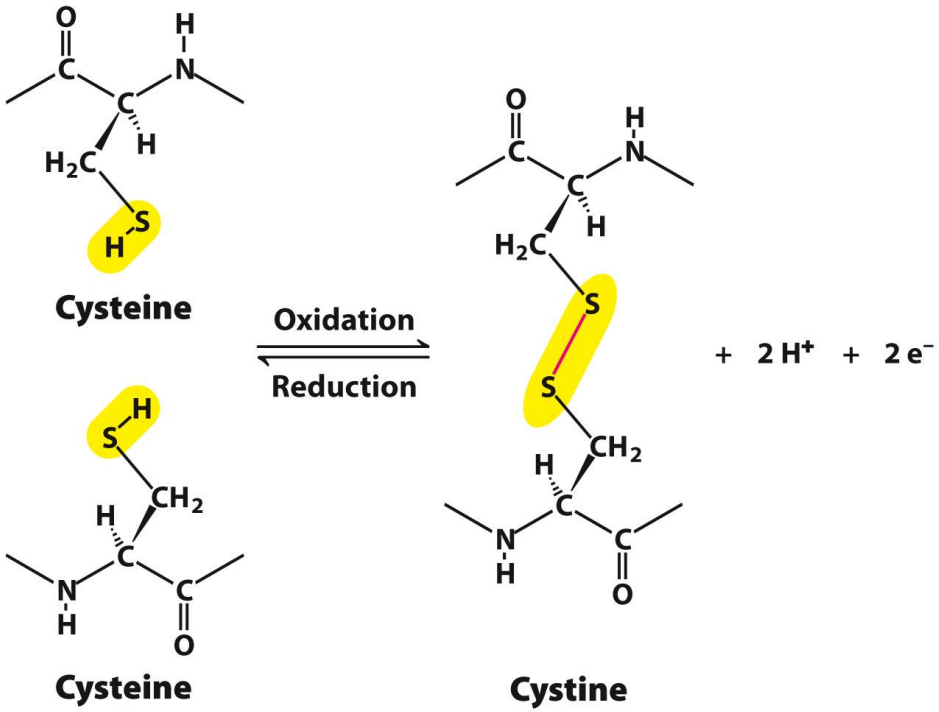


Figure 2.16
Biochemistry, Eighth Edition

In some proteins, the polypeptide chain can be cross-linked by disulfide bonds.

Disulfide bonds form by the oxidation of two cysteines.

The cross-linked cysteines are called cystine.

Example: insulin (two chains are connected by the disulfide bridges).

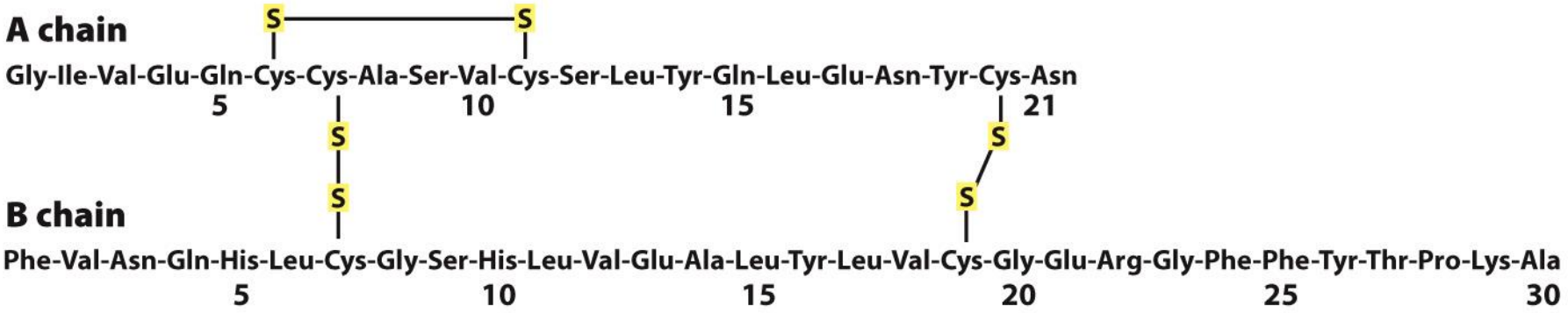
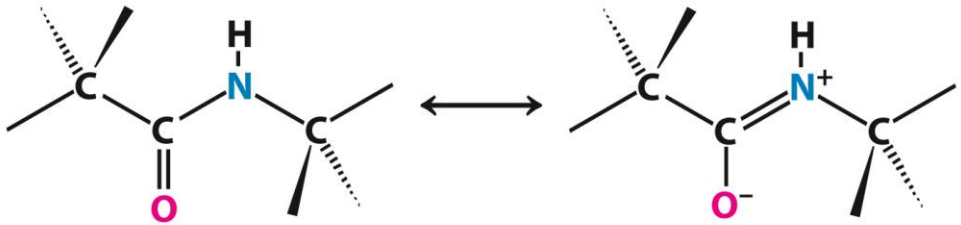


Figure 2.17
Biochemistry, Eighth Edition

Properties of the peptide bond

The peptide bond is essentially planar. Six atoms (C_α, C, O, N, H, and C_α) lie in a plane.

The peptide bond has partial double bond character because of resonance, and thus rotation about the bond is prohibited.



Peptide-bond resonance structures

Unnumbered 2 p38
Biochemistry, Eighth Edition

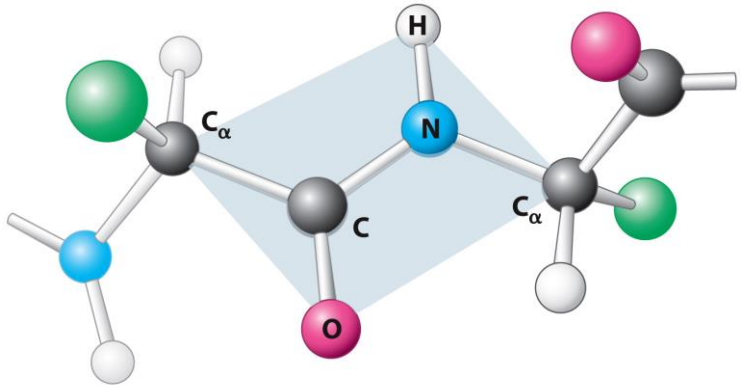


Figure 2.18
Biochemistry, Eighth Edition

The peptide bond is uncharged.
The peptide bond has 40 % double bond character

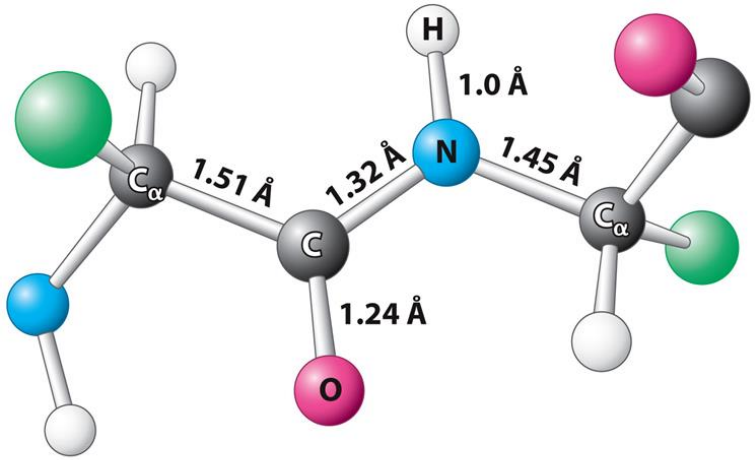


Figure 2.19
Biochemistry, Eighth Edition

Peptide bound is in trans conformation

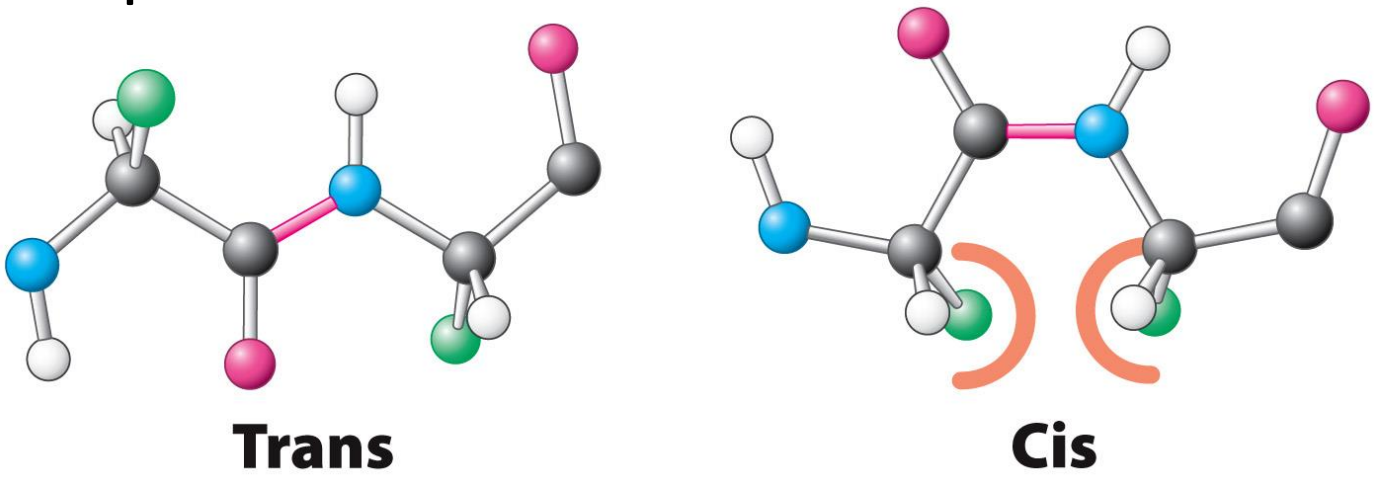
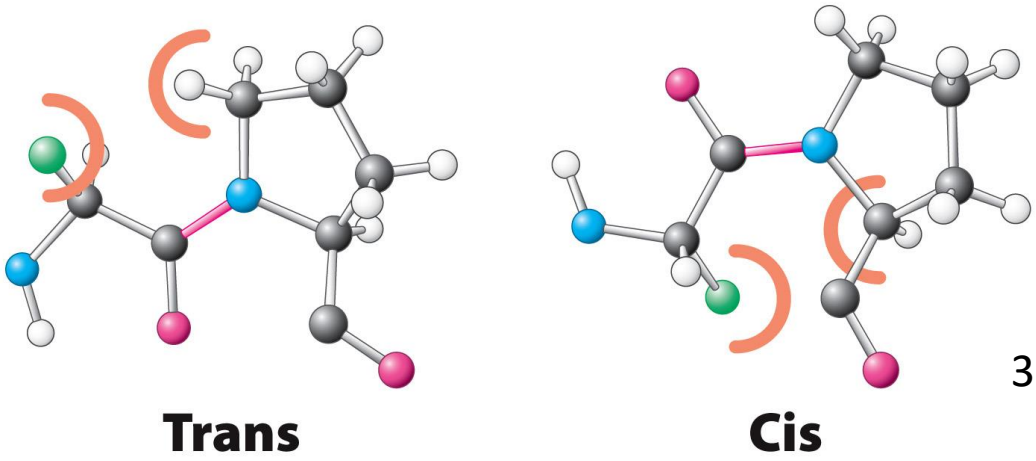


Figure 2.20
Biochemistry, Eighth Edition

Exception: X-Pro linkage



30% in cis conformation

Figure 2.21
Biochemistry, Eighth Edition

Rotation around torsion angles (dihedral angles)

Rotation is permitted about the N-C_α bond (the phi (ϕ)bond) and about the C_α-carbonyl bond (the psi (ψ) bond.)

The rotation about the Φ and ψ bonds, called the torsion angle, determines the path of the polypeptide chain. Not all torsion angles are permitted.

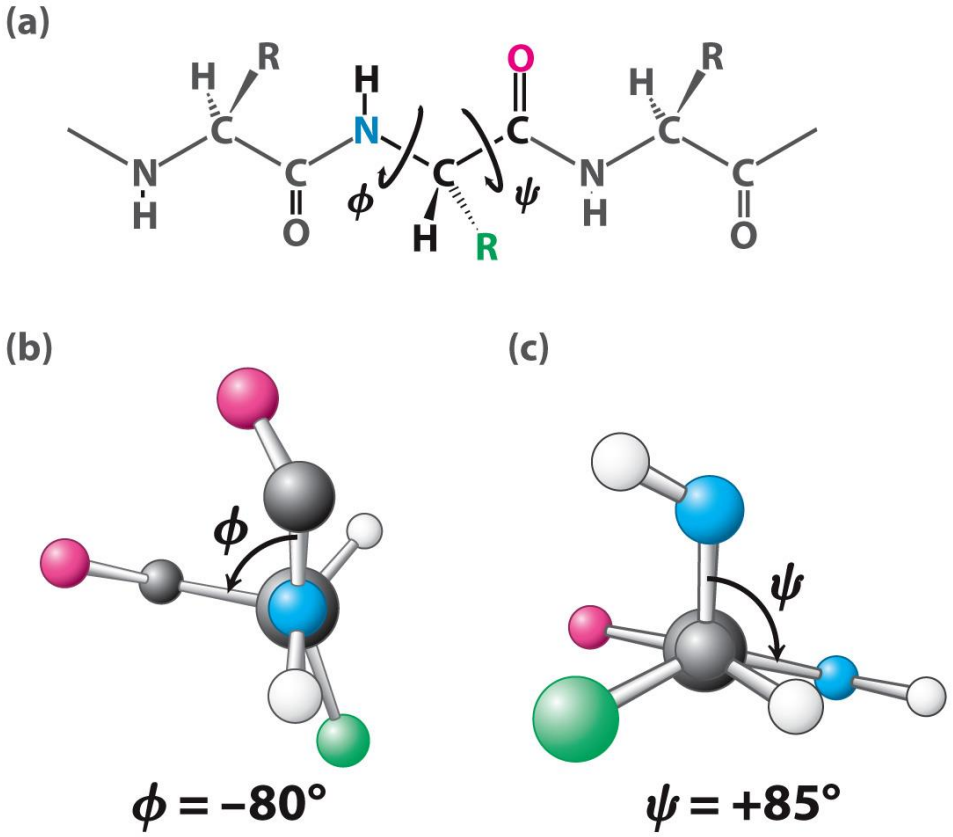


Figure 2.22
Biochemistry, Eighth Edition

The Ramachandran plot illustrates the φ and ψ angles that are favorable because there is no steric hindrance.

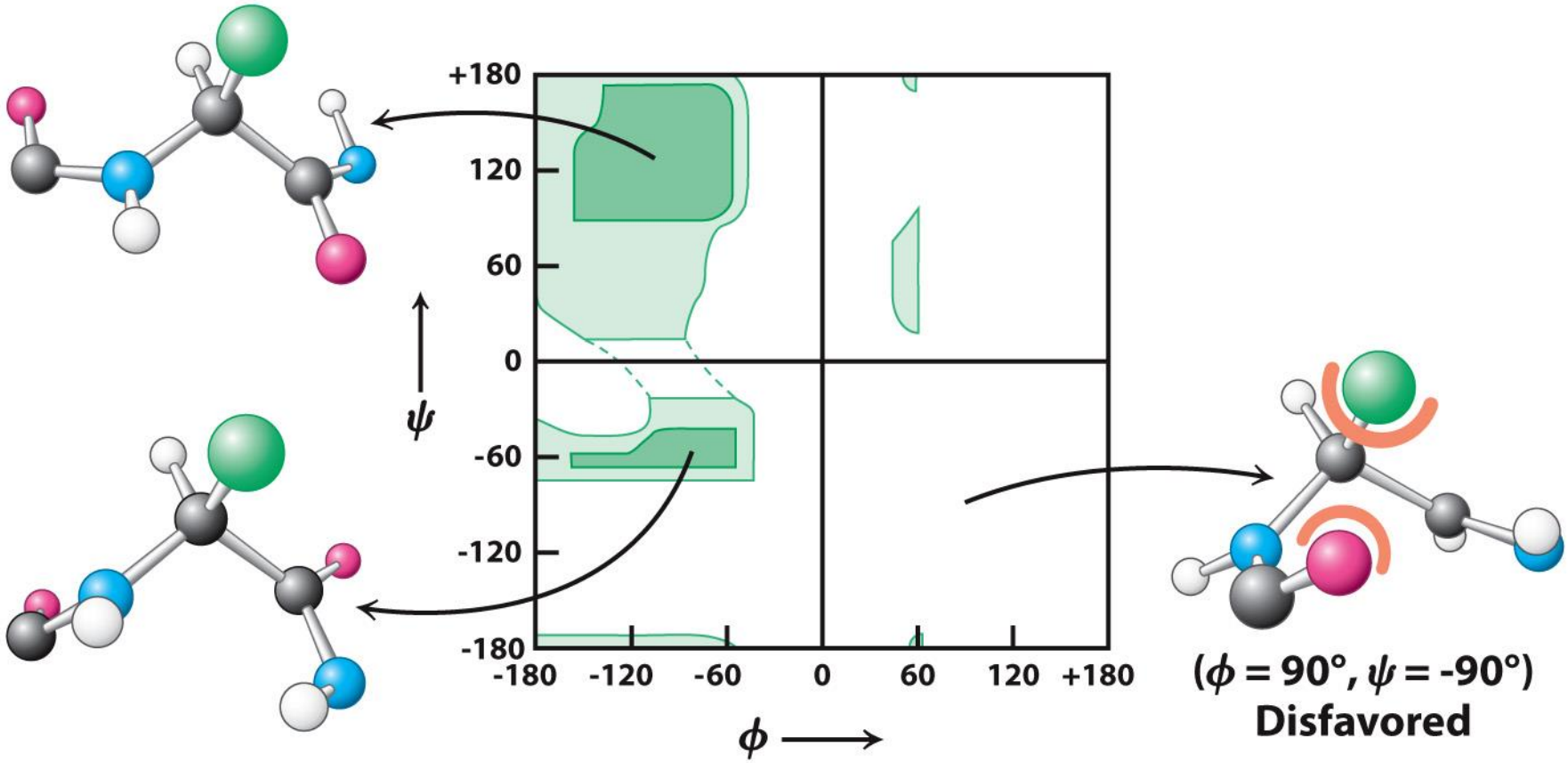


Figure 2.23
Biochemistry, Eighth Edition

Secondary structure elements

α- helices (globins)

β - sheets (for example GFP)

Ω- loops (important for protein protein recofnitions)

β- turns (connect beta sheets)

Random coil

α- helices

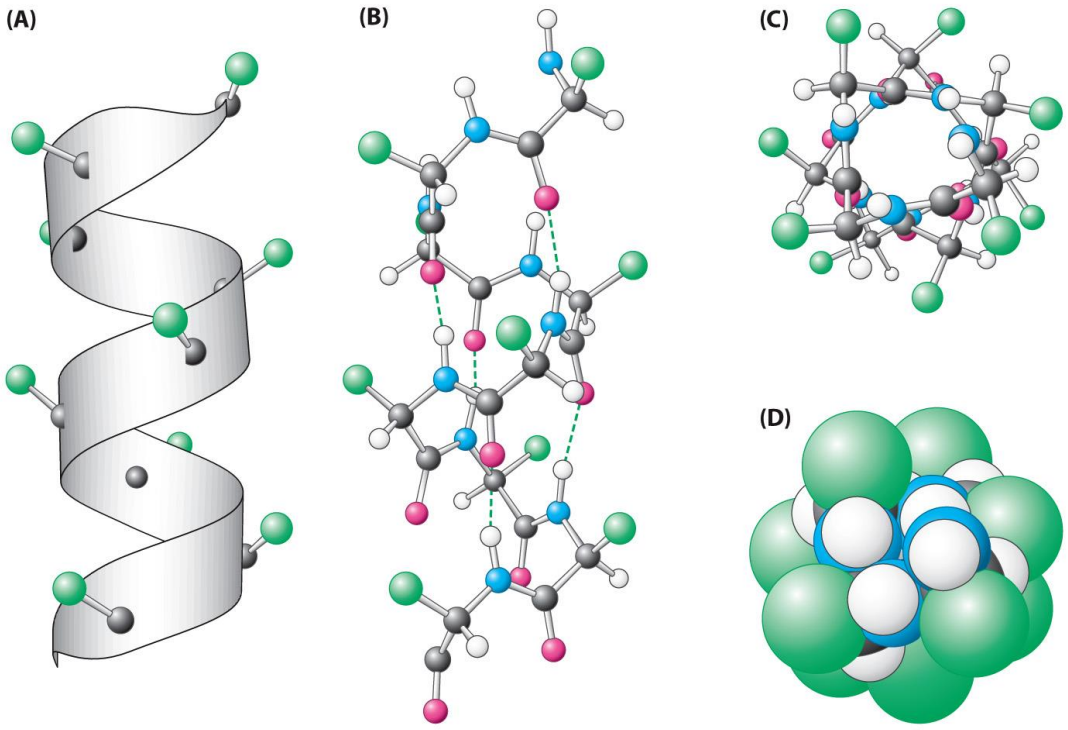


Figure 2.24
Biochemistry, Eighth Edition

-rod like structure
Backbone forms the inner part of alfa helic
-sidechains are exposed towards the solvent
-structure is stabilized by hydrogen bonding

3.6 AA residues per turn
Rise per amino acid residue is 1.5A
Pitch (rise per turn) 5.4 A
Right handed helices are sterically favorable (less sterical hindrance between sidechains)
(Val Ile Thr Ser Asp Asn nonfavorable)

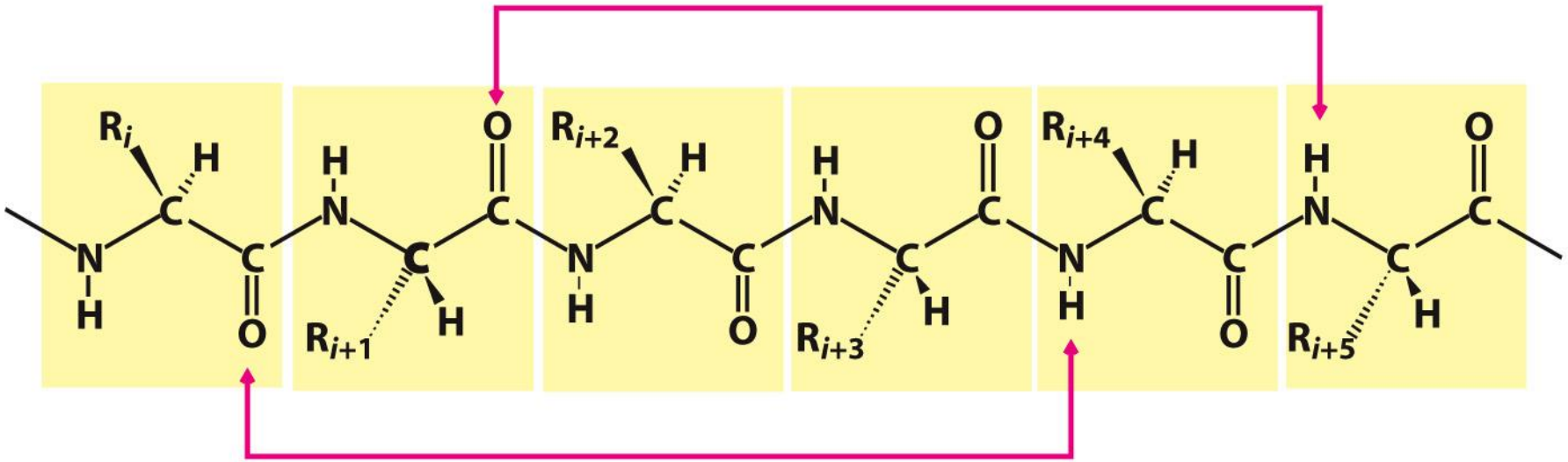


Figure 2.25
Biochemistry, Eighth Edition

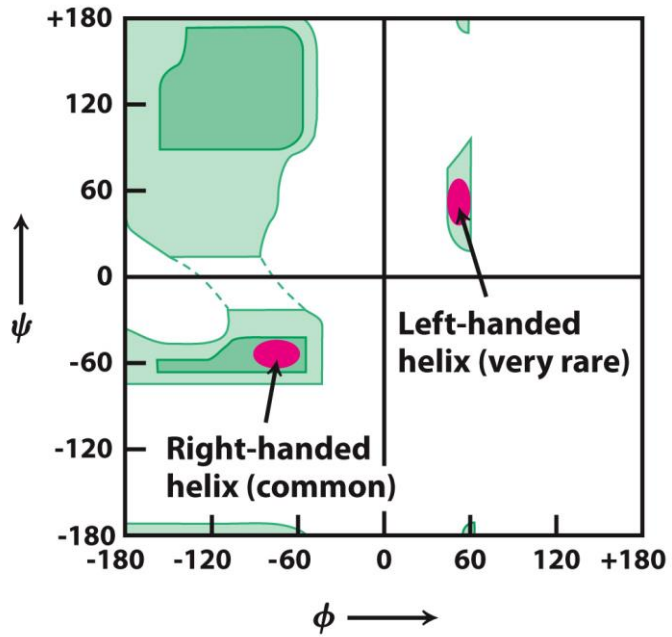


Figure 2.26
Biochemistry, Eighth Edition

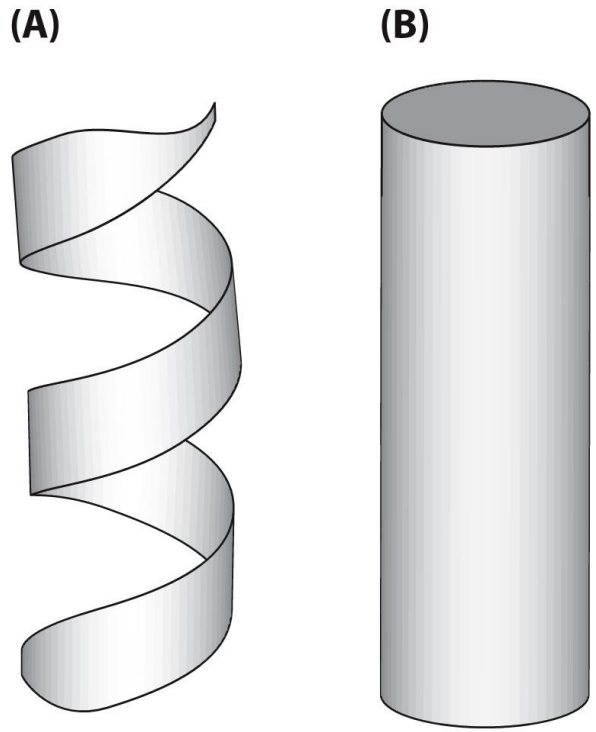


Figure 2.27
Biochemistry, Eighth Edition

The β-strand is another common form of secondary structure.

Beta sheets are formed by adjacent β-strands.

In contrast to an α-helix, the polypeptide in a β-strand is fully extended.
The side-chains are below and above the plane of the chain

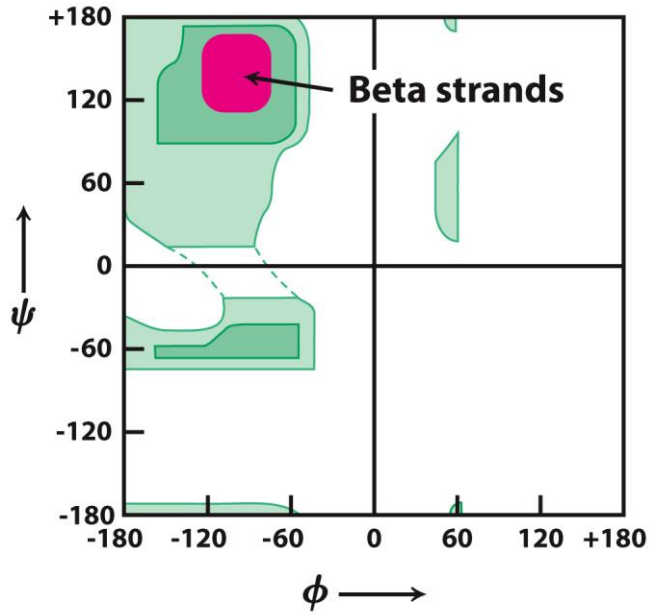


Figure 2.29
Biochemistry, Eighth Edition
© 2015 Macmillan Education

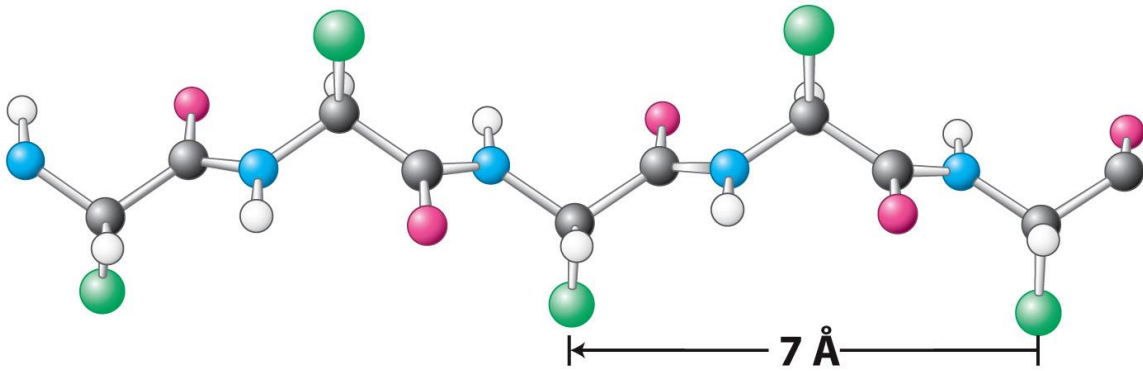


Figure 2.30
Biochemistry, Eighth Edition
© 2015 Macmillan Education

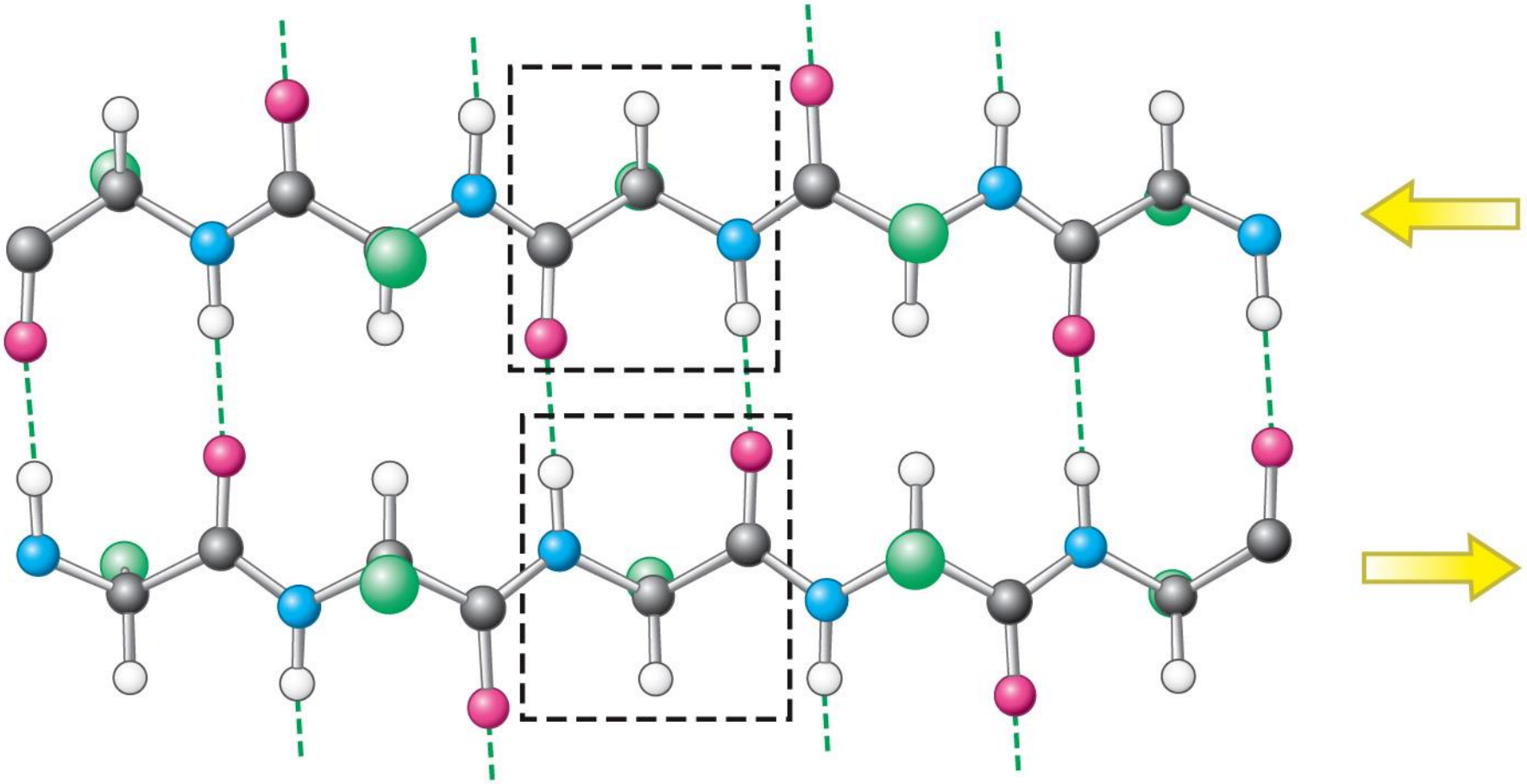


Figure 2.31
Biochemistry, Eighth Edition

β- sheet is formed by two beta strand running in the same (PARALEL) of opposite direction (ANTIPARALEL) beta sheets

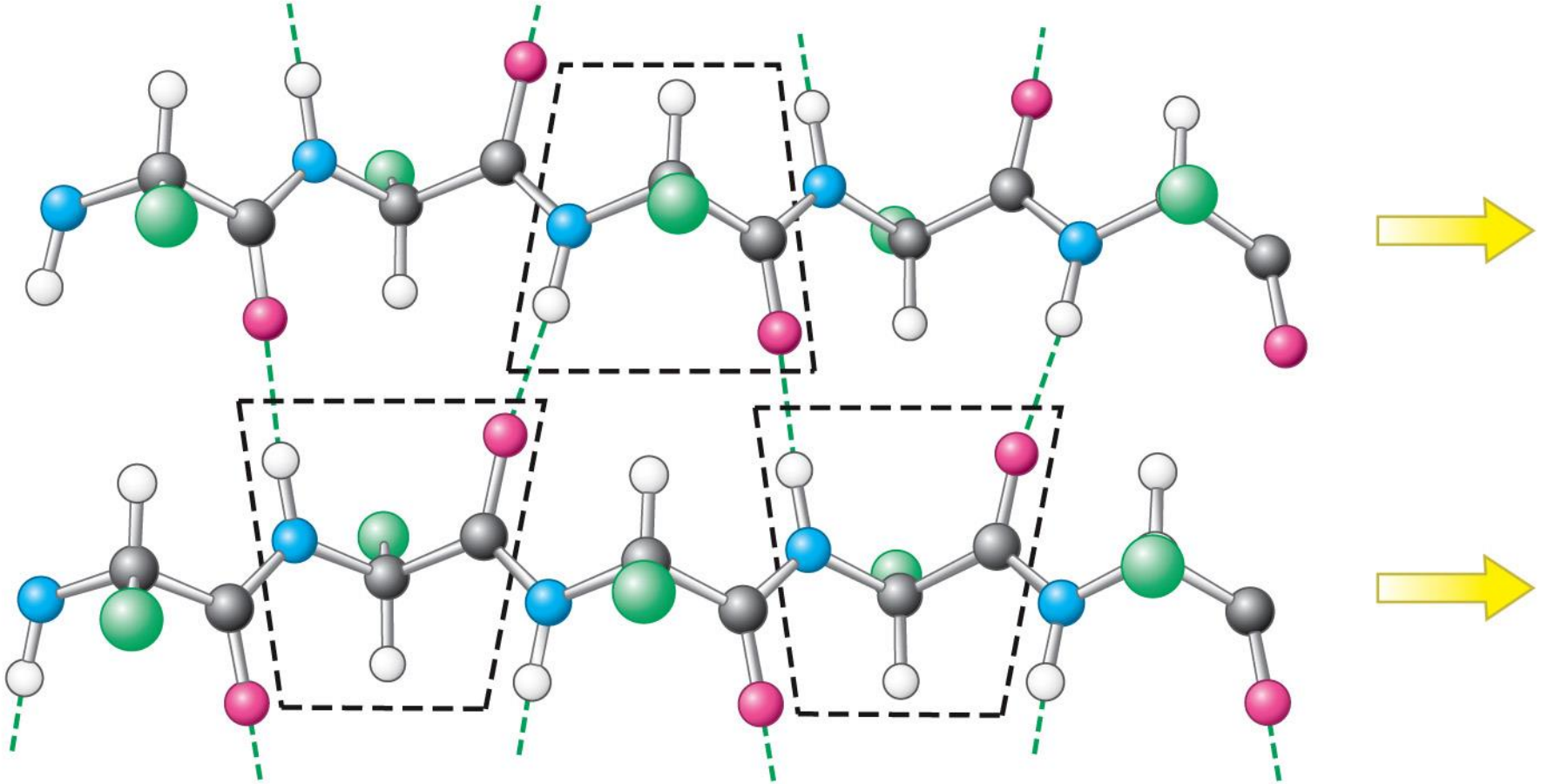


Figure 2.32
Biochemistry, Eighth Edition

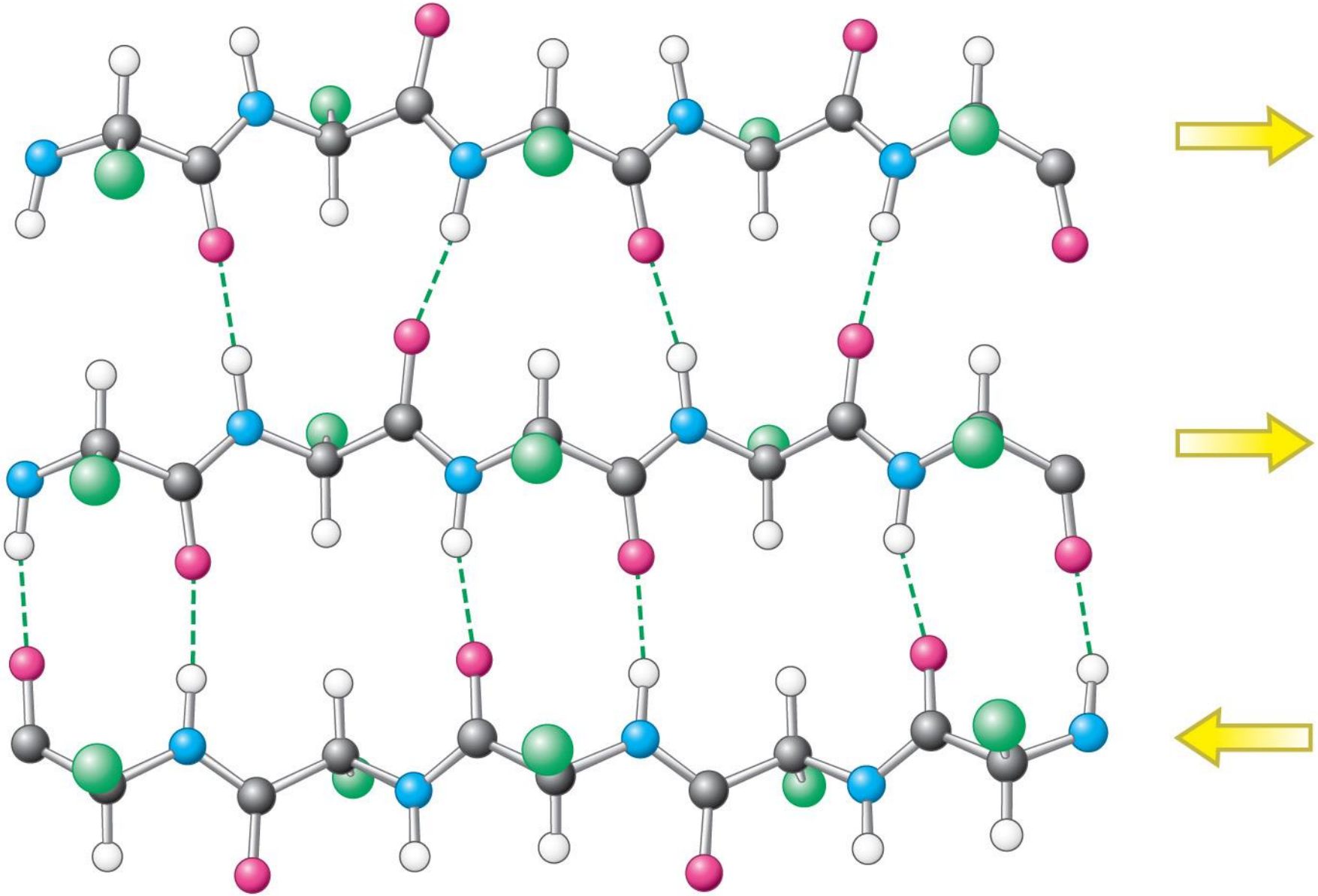


Figure 2.33
Biochemistry, Eighth Edition

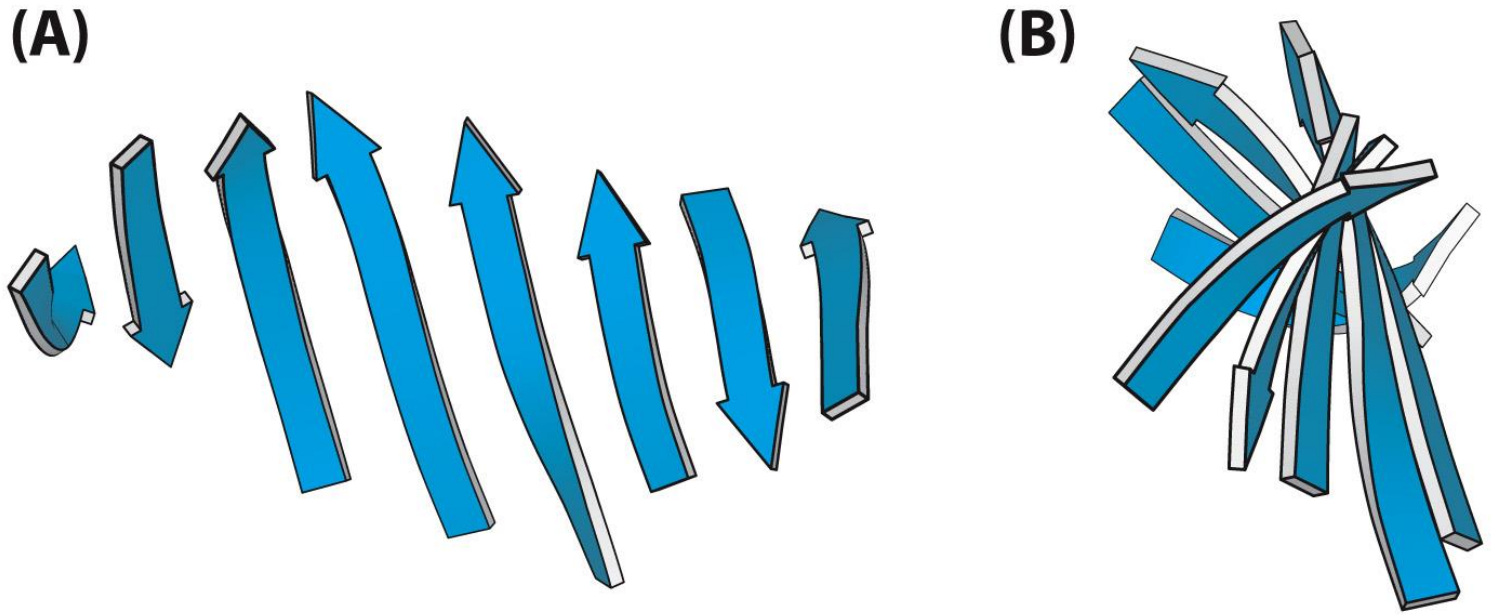


Figure 2.34
Biochemistry, Eighth Edition

Individual beta strands are arranged such way that they are twisted

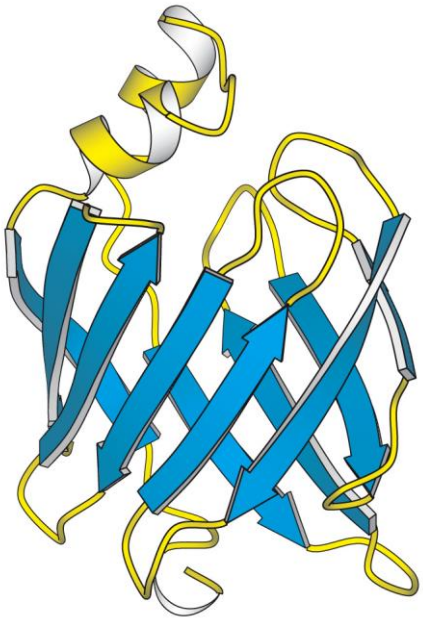


Figure 2.35
Biochemistry, Eighth Edition

Fatty acid binding protein

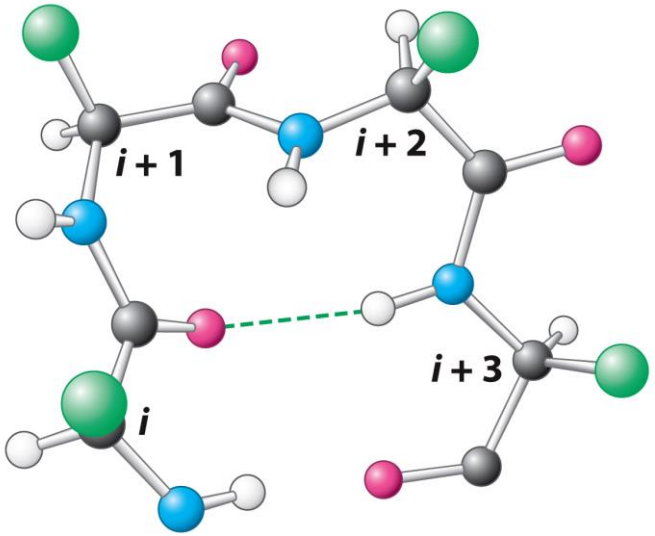


Figure 2.36
Biochemistry, Eighth Edition

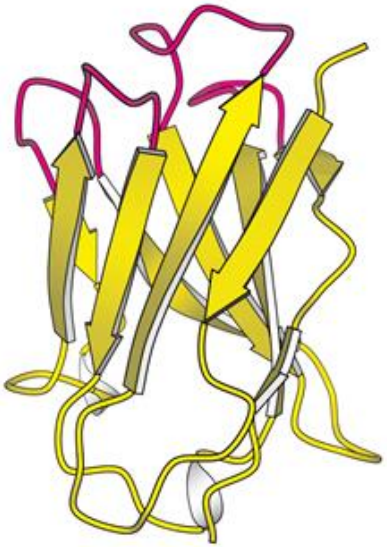


Figure 2.37
Biochemistry, Eighth Edition

Antibody bonding molecule

Fibrous proteins provide structural support for cells and tissues

α-Keratin, a structural protein found in wool and hair, is composed of two right-handed α-helices intertwined to form a left-handed super helix called a coiled-coil. The helices interact with ionic bonds or van der Waals interactions.

α-Keratin is a member of a superfamily of structural proteins called coiled-coil proteins.

Other members of the family include some cytoskeleton proteins and muscle proteins.

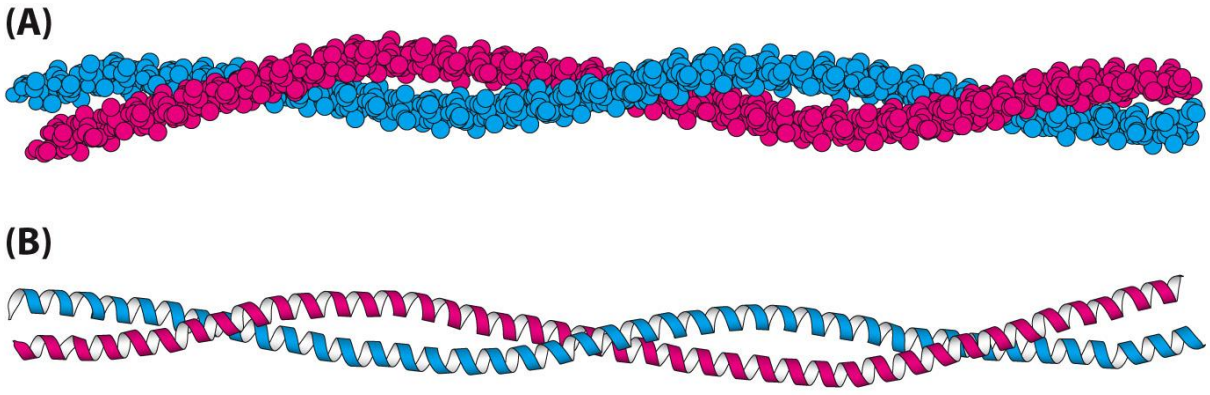


Figure 2.38
Biochemistry, Eighth Edition

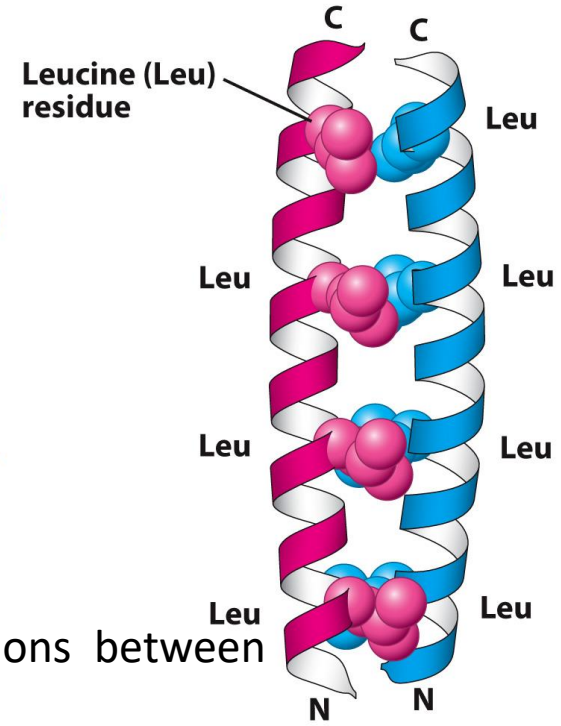


Figure 2.39
Biochemistry, Eighth Edition

Coiled coled structures are stabilized by hydrophobic interactions between each seventh residue from two helices

Collagen

Collagen is a structural protein that is a component of skin, bone, tendons, cartilage and teeth.

Collagen consists of three intertwined helical polypeptide chains that form a superhelical cable. The helical polypeptide chains of collagen are not α-helices.

Glycine appears at every third residue and the sequences Gly-Pro-Hyp and Gly-Pro-Pro are common.

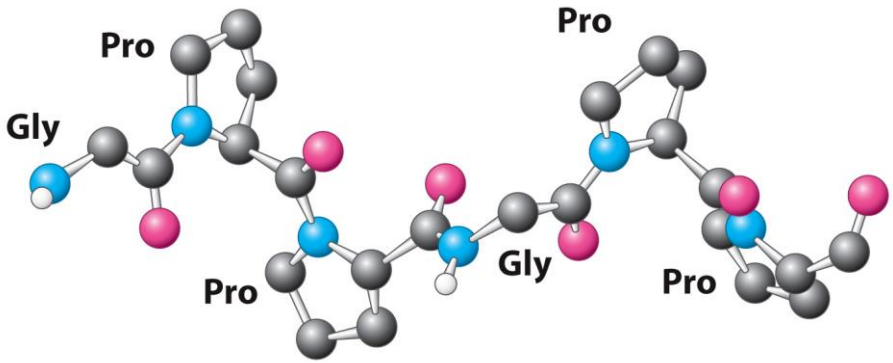


Figure 2.41
Biochemistry, Eighth Edition
© 2015 Macmillan Education

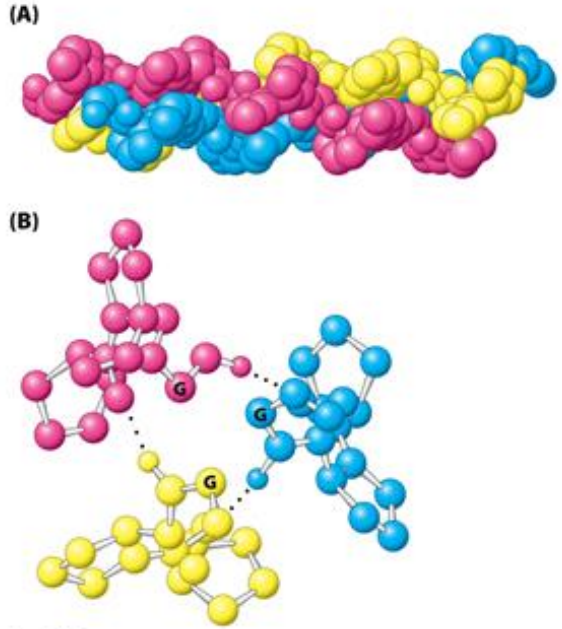


Figure 2.42
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Tertiary structure

Tertiary structure refers to the spatial arrangement of amino acids that are far apart in the primary structure and to the pattern of disulfide bond formation.

Globular proteins, such as myoglobin, form complicated three-dimensional structures.

Globular proteins are very compact. There is little or no empty space in the interior of globular proteins.

The interior of globular proteins consists mainly of hydrophobic amino acids.

The exterior of globular proteins consists of charged and polar amino acids.

Membrane proteins have the reverse distribution of hydrophilic and hydrophobic amino acids.

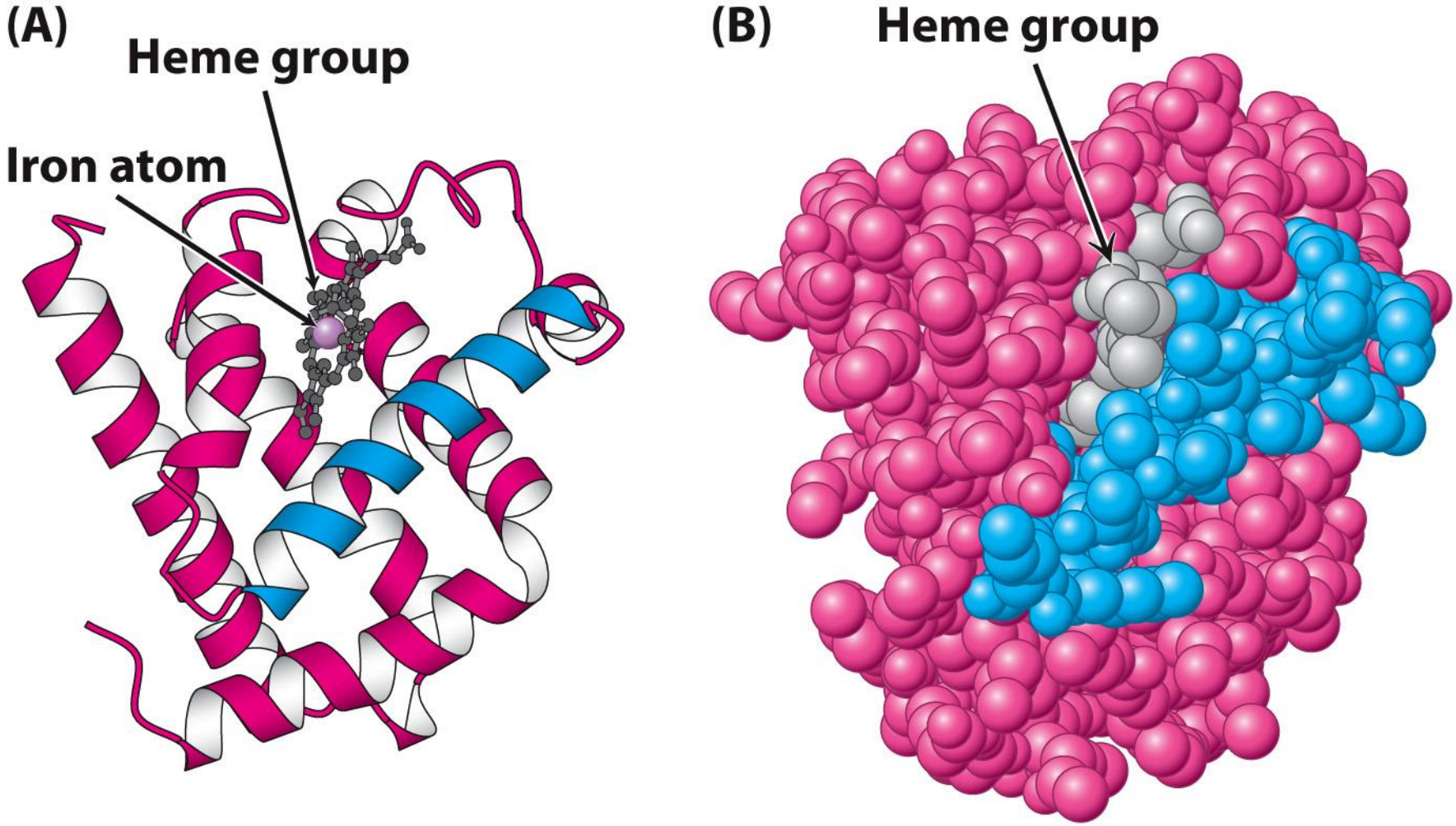


Figure 2.43
Biochemistry, Eighth Edition

Distribution of amino acid residues in porin

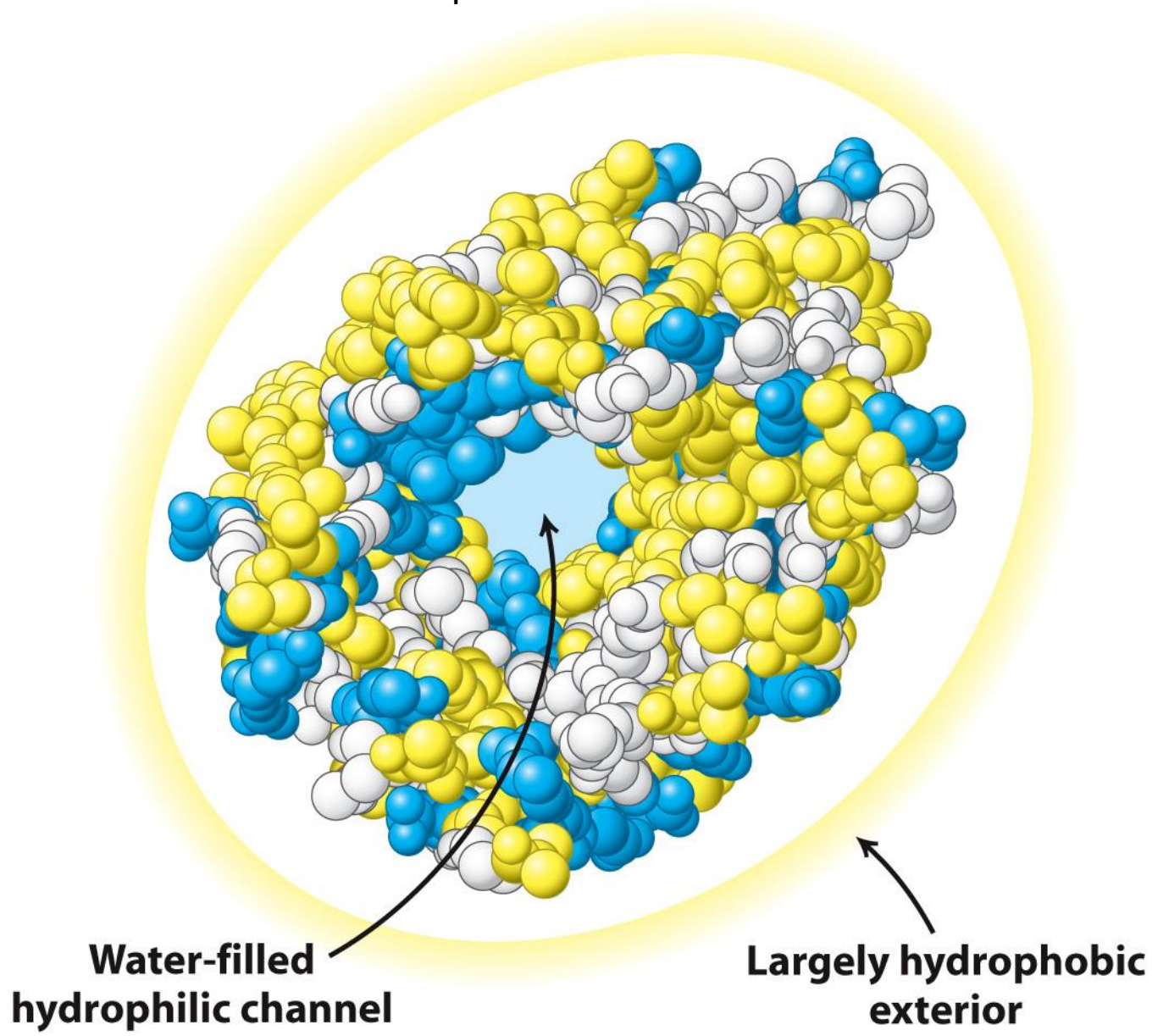


Figure 2.45
Biochemistry, Eighth Edition

Motifs, or **supersecondary structures**, are combinations of secondary structure that are found in many proteins.

Some proteins have two or more similar or identical compact structures called domains.

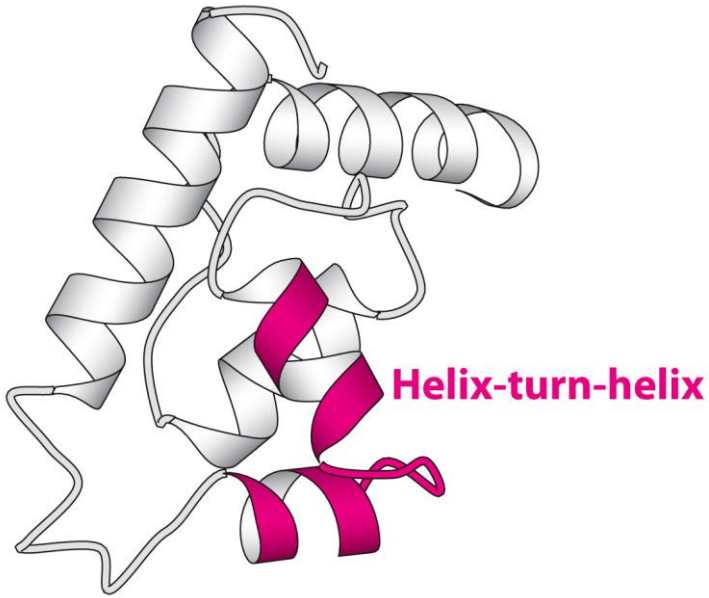
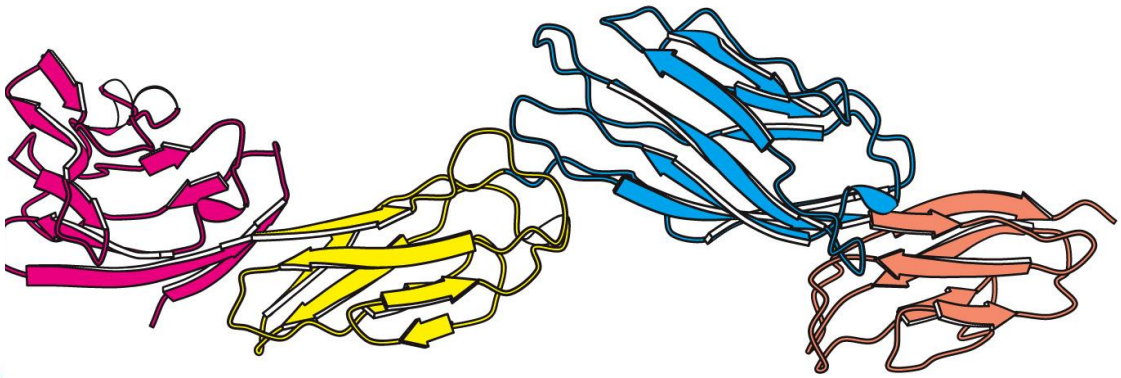


Figure 2.46
Biochemistry, Eighth Edition
© 2015 Macmillan Education



ure 2.47
chemistry, Eighth Edition
2015 Macmillan Education

Cell surface protein CD4; individual domains are colored in red, yellow, blue and orange

Many proteins are composed of multiple polypeptide chains called subunits. Such proteins are said to display quaternary structure.

Quaternary structure can be as simple as two identical polypeptide chains or as complex as dozens of different polypeptide chains.

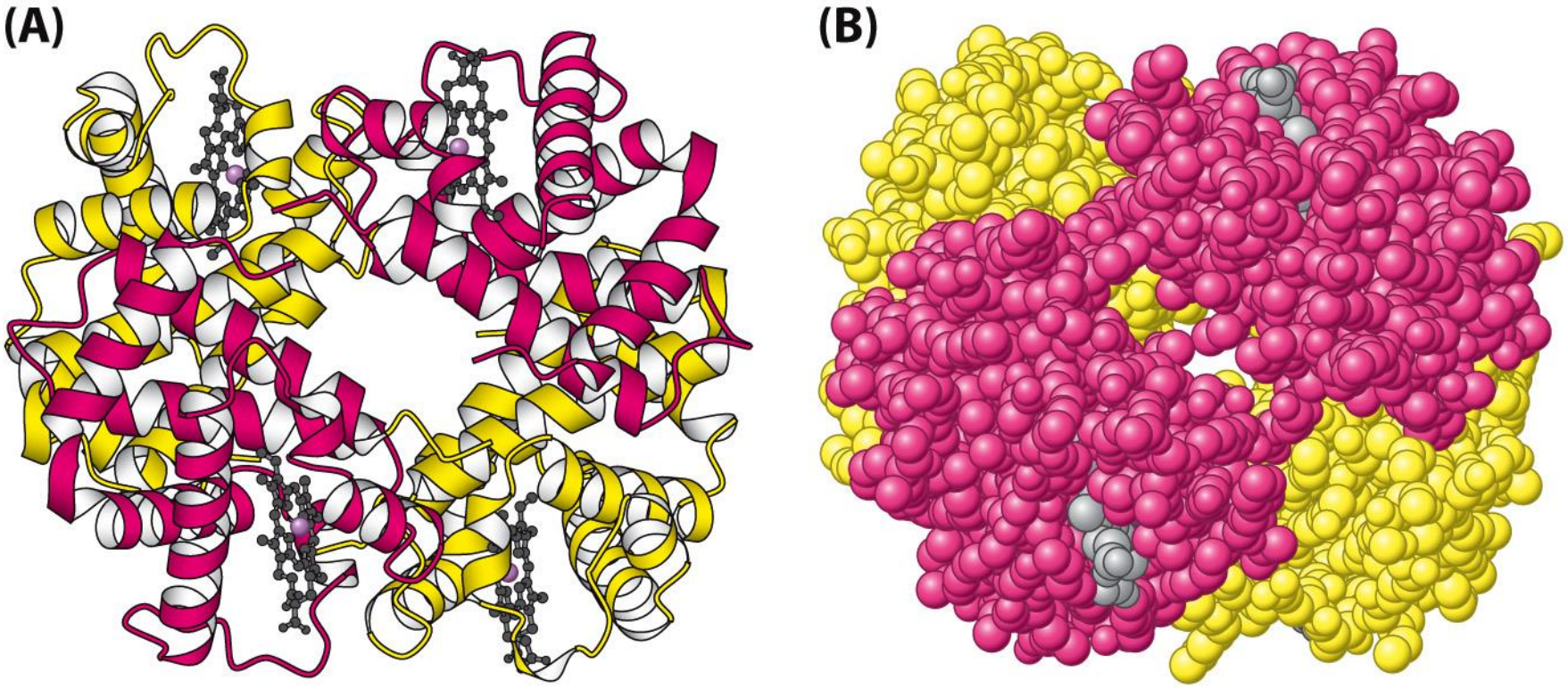


Figure 2.49
Biochemistry, Eighth Edition

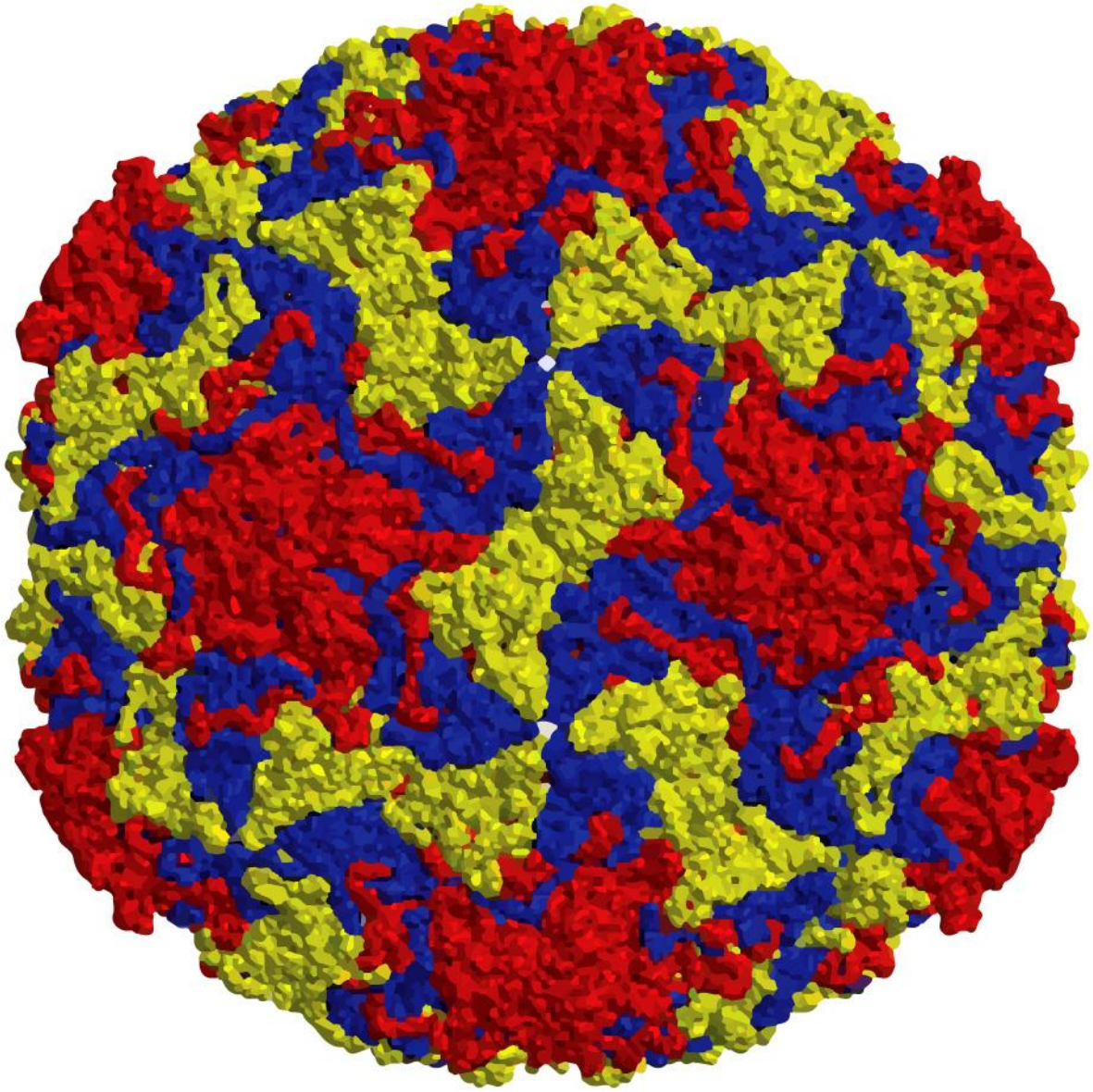


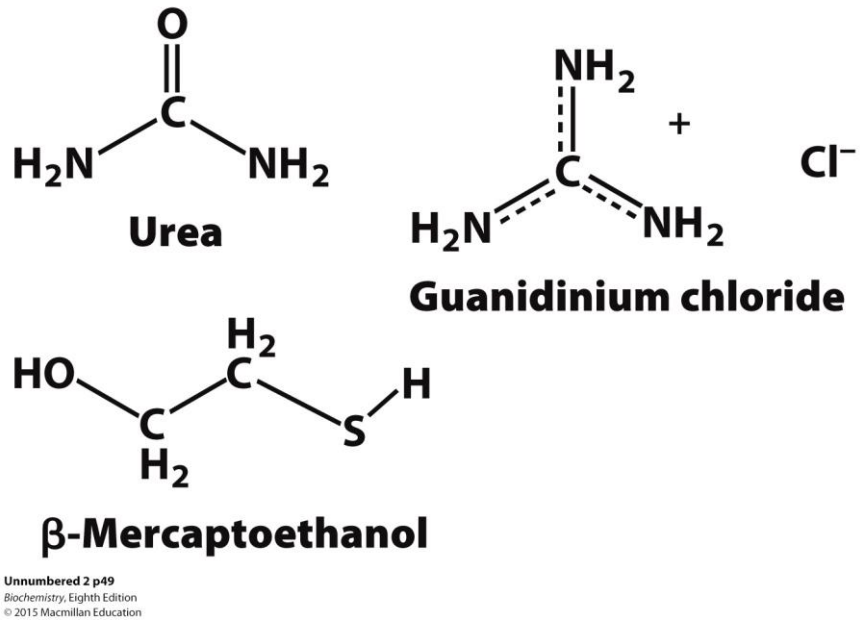
Figure 2.50
Biochemistry, Eighth Edition

Coat of human rhinovirus is formed of 60 copies of each of 4 subunits

Protein folding

Christian Anfinsen unfolded ribonuclease, in a solution containing urea or guanidinium chloride and β-mercaptoethanol. Urea or guanidinium chloride destroyed all **noncovalent bonds**, while the β-mercaptoethanol destroyed the **disulfide bonds**. The enzyme displayed no enzymatic activity and existed only as a random coil. The ribonuclease was denatured.

When the urea and β-mercaptoethanol were slowly removed, the enzyme regained its structure and its activity. Ribonuclease was renatured and attained its normal or native state.



These results demonstrated that the **information required for a polypeptide chain to fold into a functional protein with a defined three-dimensional structure is inherent in the primary structure.**

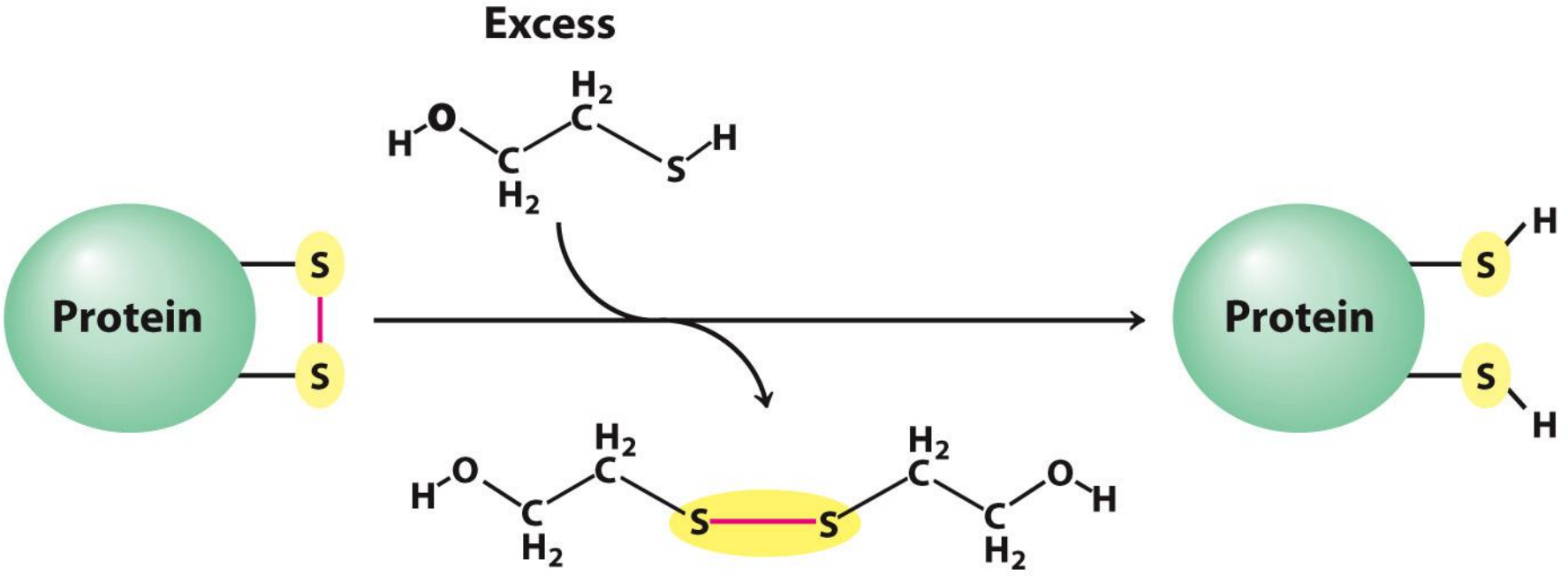


Figure 2.52
Biochemistry, Eighth Edition

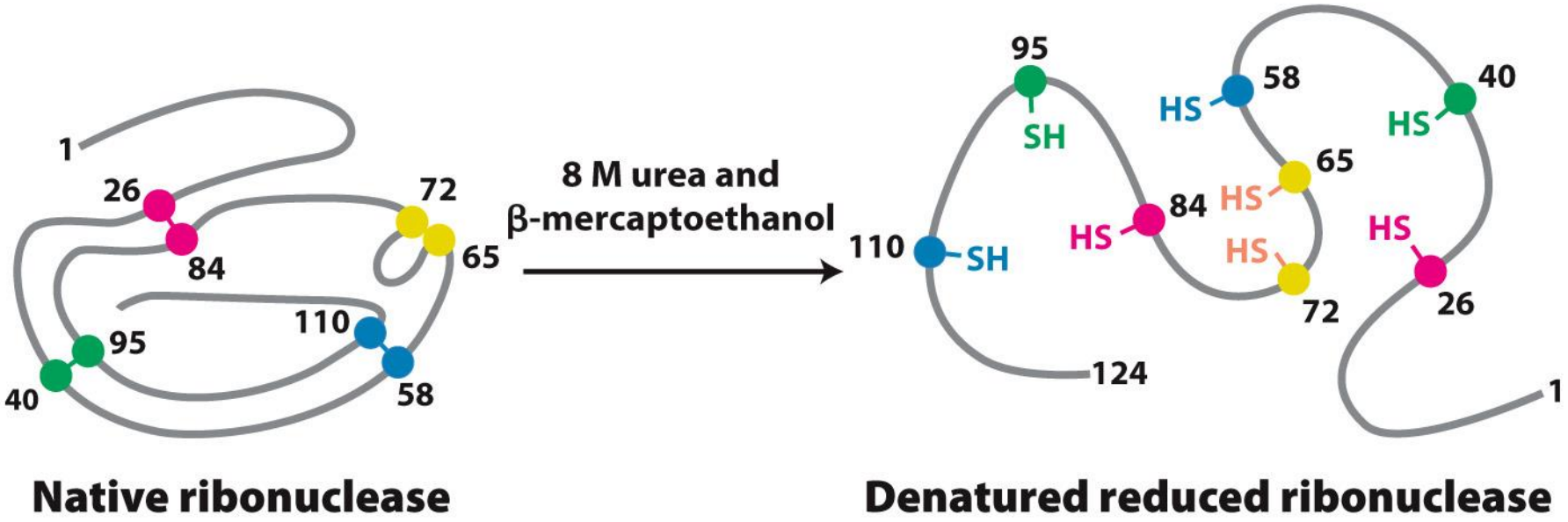
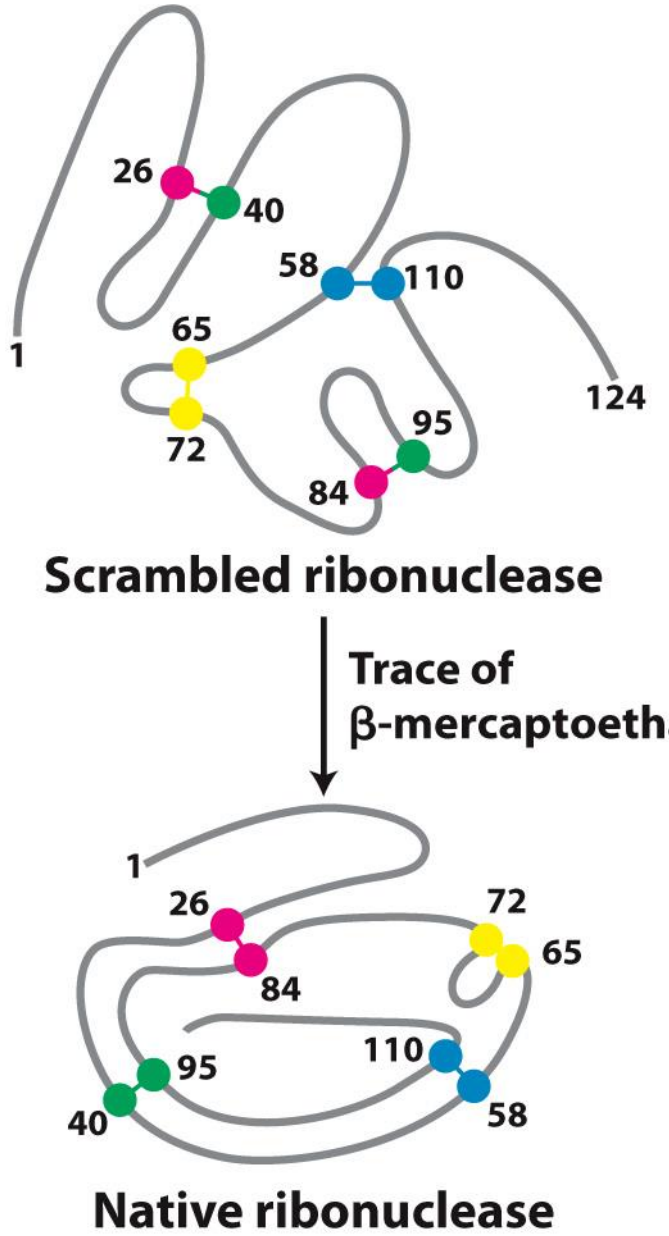


Figure 2.53
Biochemistry, Eighth Edition



Denaturated protein with uncorectly formed disulfide bonds

Slow refolding in the presence of low concentration of mercaptoethanol allows for correct formation of native disulfide bonds

Figure 2.54
Biochemistry, Eighth Edition

Propensity of amino acid residues to form secondary structure elements

Conformational preferences of amino acids are not strong, suggesting that the majority of amino acids will be found in helices, strands or turns.

TABLE 2.3 Relative frequencies of amino acid residues in secondary structures

Amino acid	α helix	β sheet	Reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.96
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

Note: The amino acids are grouped according to their preference for α helices (top group), β sheets (middle group), or turns (bottom group).

Source: T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2d ed. (W. H. Freeman and Company, 1992), p. 256.

Table 2.3
Biochemistry, Eighth Edition

Protein folding

Protein folding is a highly cooperative process

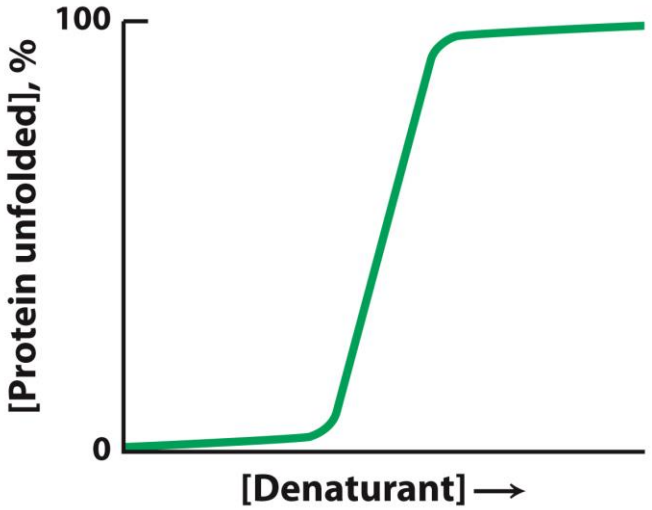


Figure 2.56
Biochemistry, Eighth Edition
© 2015 Macmillan Education

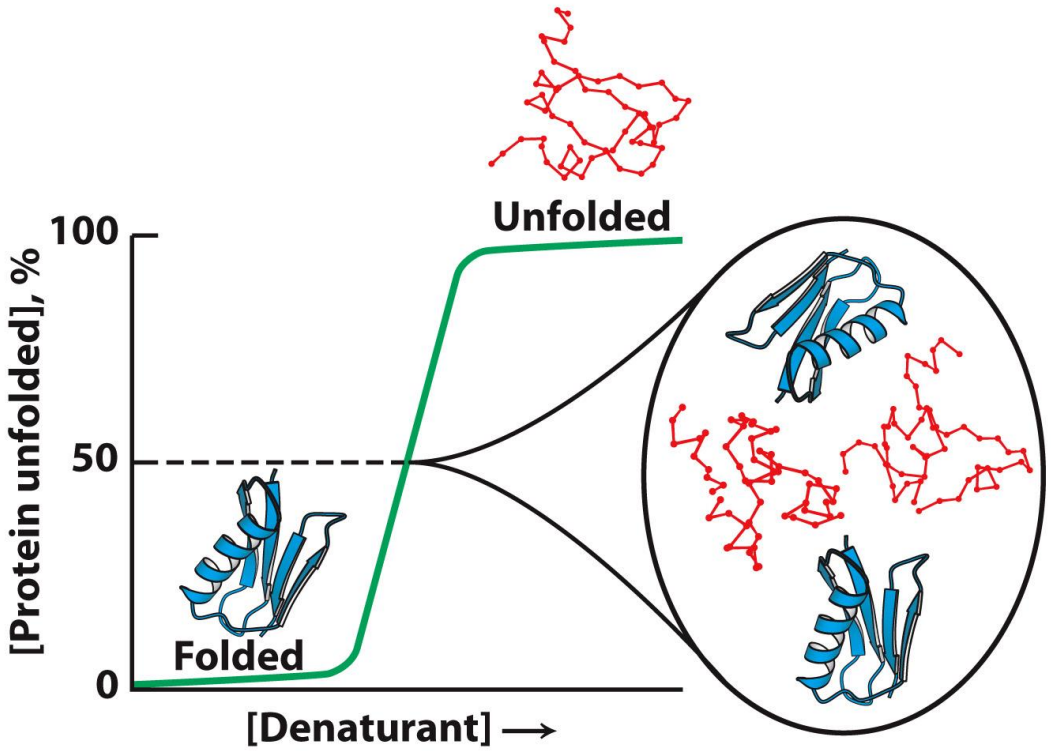


Figure 2.57
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Protein folding is often represented as a folding funnel. The protein has maximum entropy and minimal structure at the top of the funnel. The folded protein exists at the bottom of the funnel.

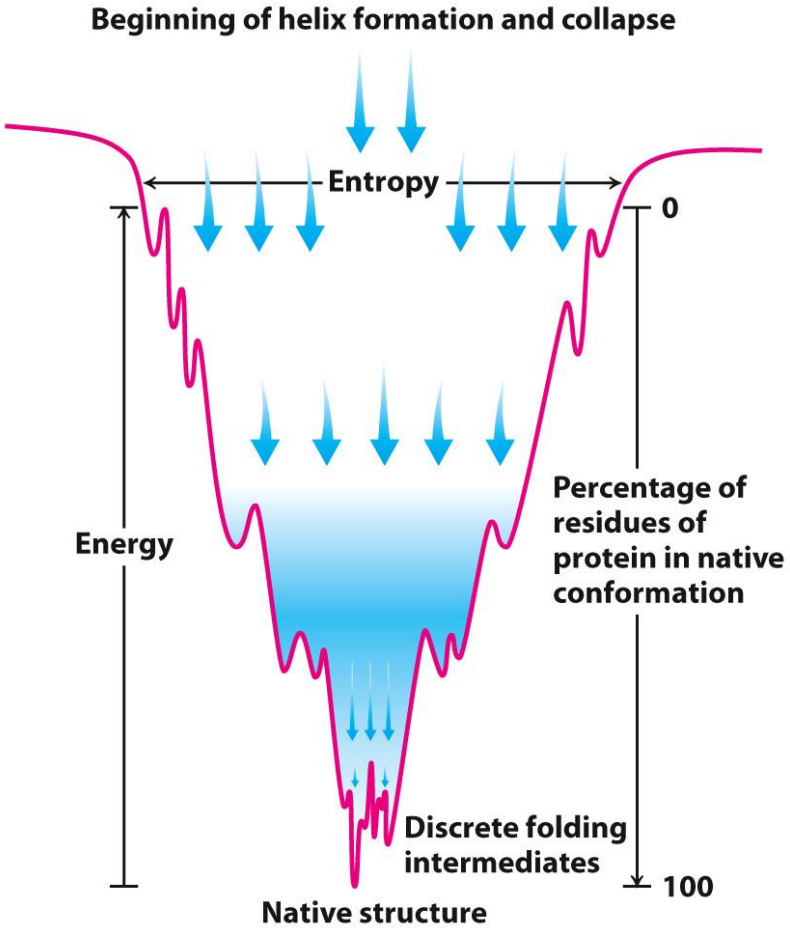


Figure 2.60
Biochemistry, Eighth Edition

Protein folds by progressive stabilization of intermediates rather than by random search

Protein folding also occurs by cumulative selection. Partly correct folding intermediates are retained because they are slightly more stable than unfolded regions.

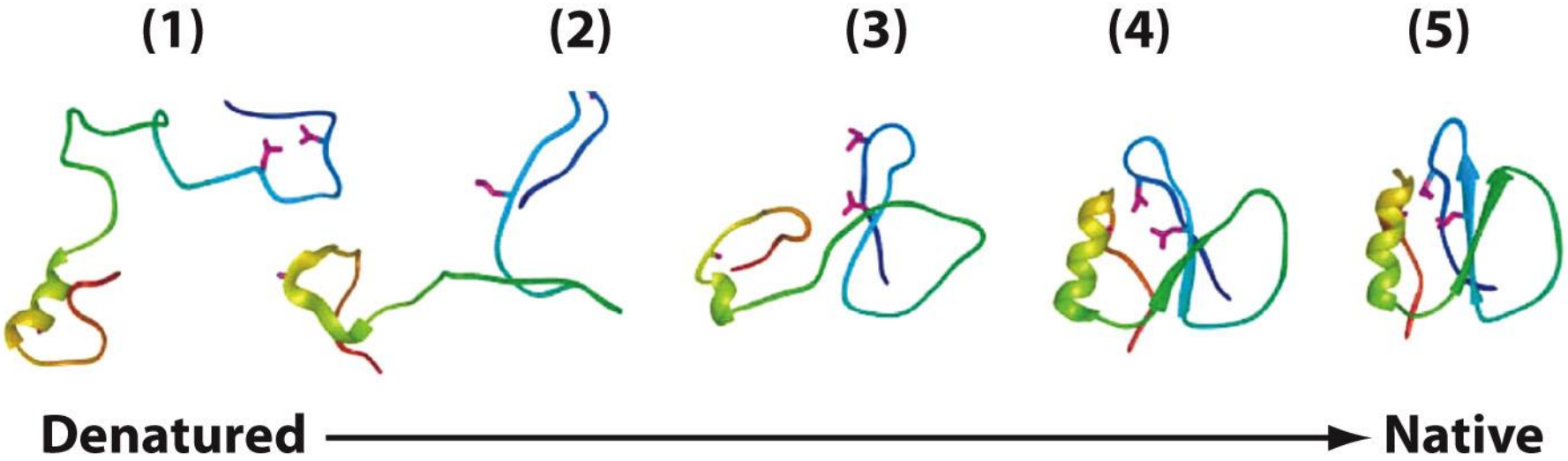


Figure 2.59
Biochemistry, Eighth Edition
From A. R. Fersht and V. Daggett. *Cell* 108:573–582, 2002; with permission from Elsevier

Intrinsically unstructured proteins (IUP) do not have a defined structure under physiological conditions until they interact with other molecules.

Metamorphic proteins exist in an ensemble of structures of approximately equal energies that are in equilibrium.

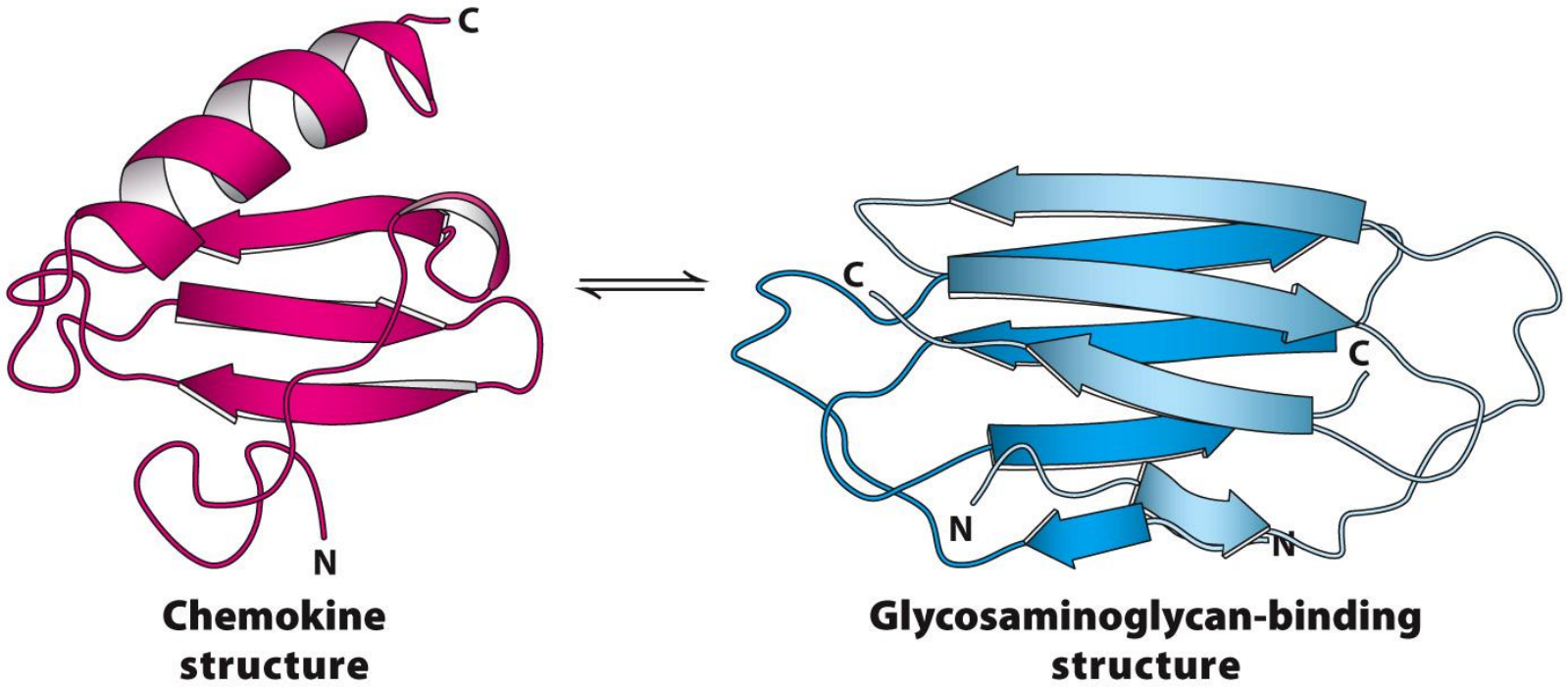


Figure 2.61
Biochemistry, Eighth Edition

Lymphotactic protein

Protein misfolding and aggregation are associated with some neurological diseases

Amyloidoses are diseases that result from the formation of protein aggregates, called amyloid fibrils or plaques (Alzheimer, Creutzfeldt-Jakob-Disease , scrapie)

Normal protein conformations can exist in forms rich in β sheet, which are prone to aggregate.

An abnormally folded aggregate serves as a nucleus to recruit more proteins.

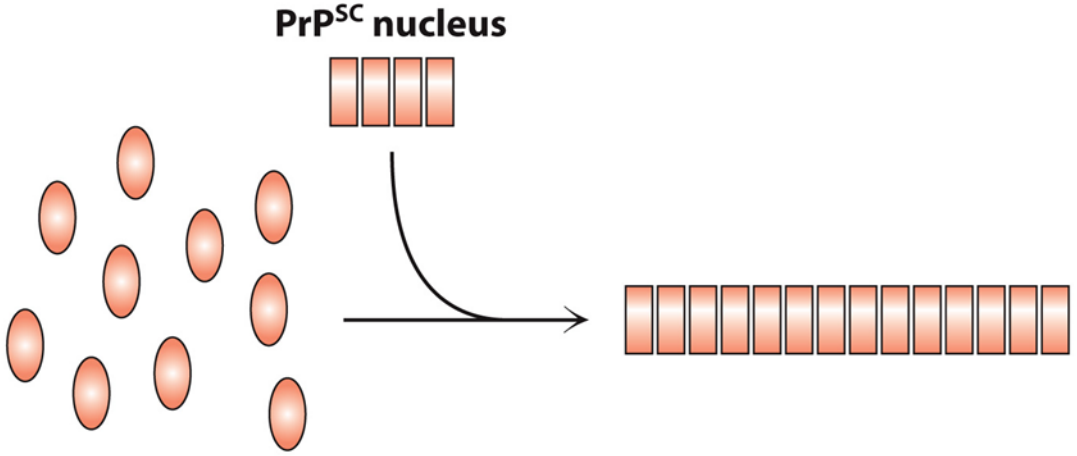


Figure 2.63
Biochemistry, Eighth Edition
© 2015 Macmillan Education

PrP prion protein (Stanley B. Prusiner)
SC scrapie

Protein modification and cleavage confer new capabilities

Lack of appropriate protein modification can result in pathological conditions.

Lack of vitamin C prevents hydroxylation of proline in collagen, which results in scurvy.

If vitamin K is missing, clotting proteins are not carboxylated and hemorrhaging results.

Green fluorescent protein (GFP), from the jellyfish *Aequorea victoria*, can be attached to cellular proteins. GFP fluoresces green when exposed to blue light, allowing the determination of the cellular location of the attached protein.

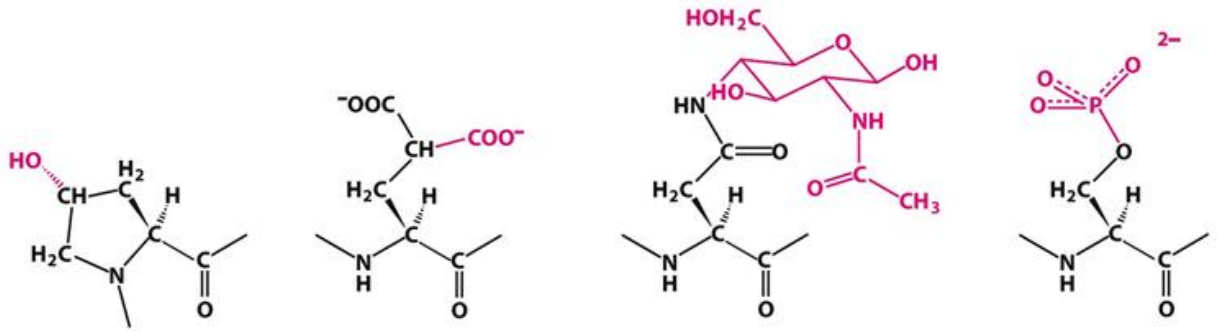


Figure 2.64
***Biochemistry*, Seventh Edition**

Green fluorescent protein

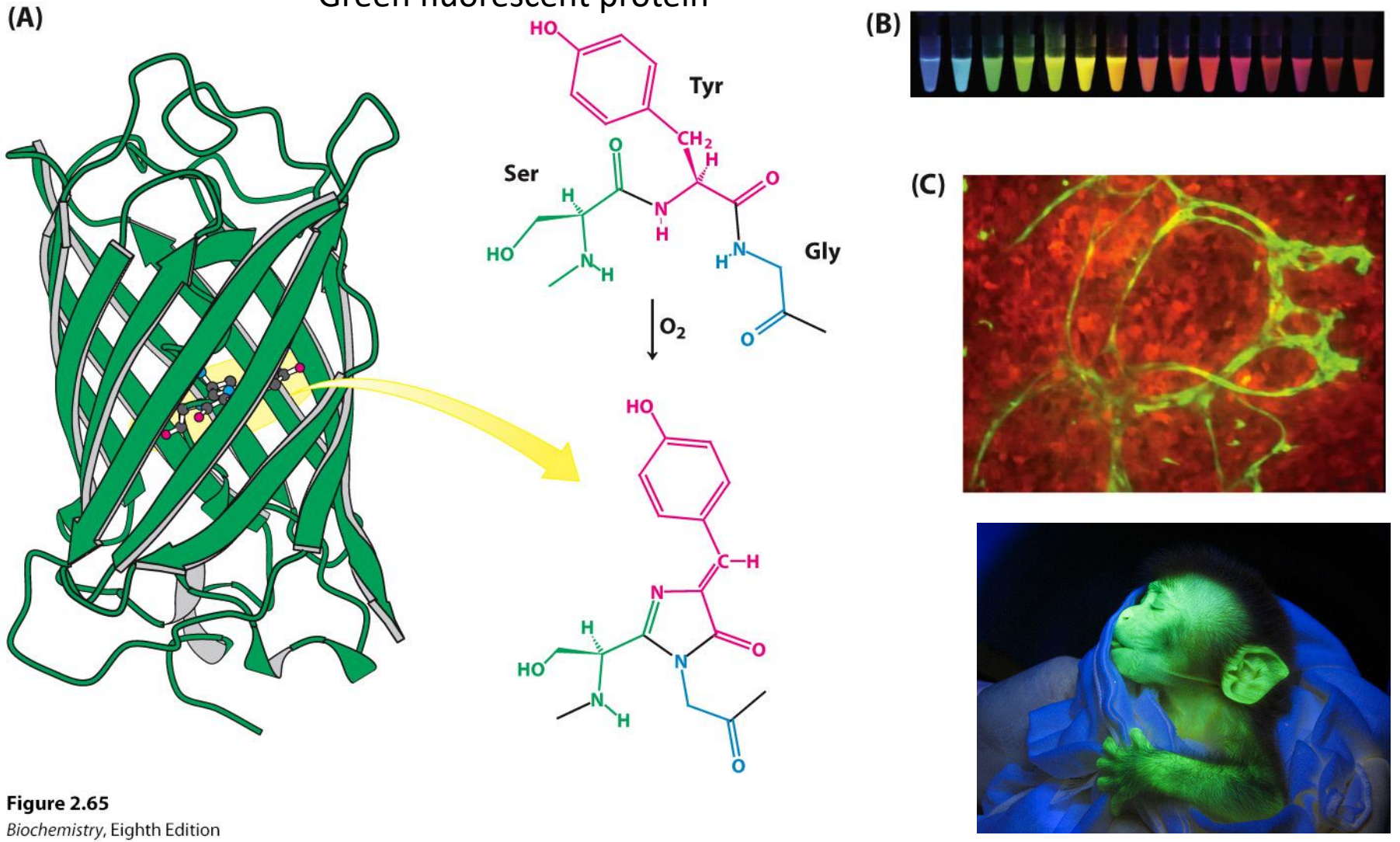


Figure 2.65

Biochemistry, Eighth Edition

 [Photos: (B) R.Y. Tsien. *Integr. Biol.* 2:77–93, 2010, Fig. 12; (C) M. Yang, et al. *Proc. Natl. Acad. Sci. U.S.A.* 100:14259–14262, 2003, Fig. 2B]