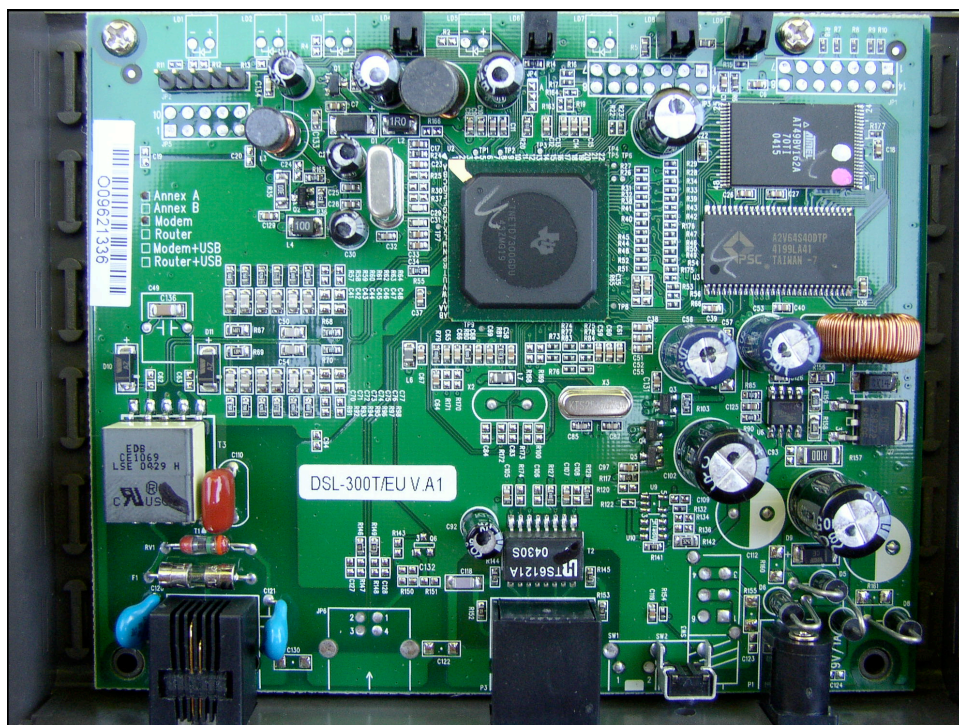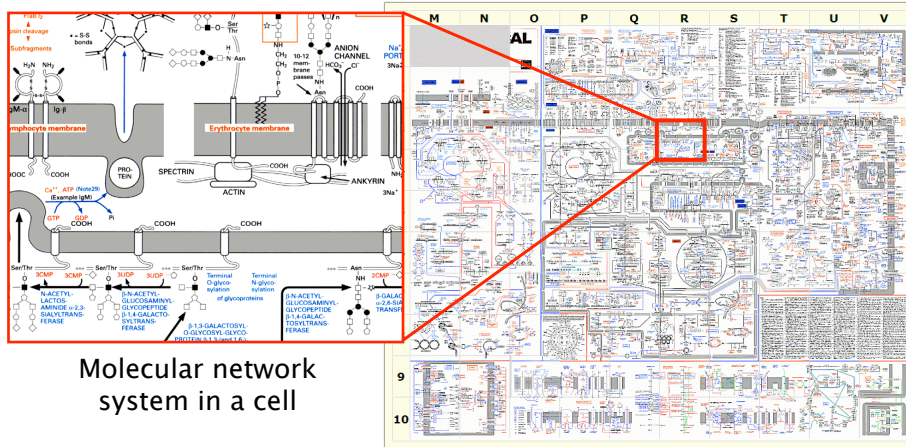# Amino Acids, Peptides & Proteins

Amino Acids
Zwitterion or ampholyte
Peptides
Proteins
Structure of Proteins
Separation & purification
Protein Sequencing
Protein Synthesis

# Our Life Is Maintained by Molecular Network Systems

Molecular network system in a cell

Proteins Play Key Roles in a Living System

- Three examples of protein functions

    - Catalysis:
      Almost all chemical reactions in a living cell are catalyzed by protein enzymes.

    - Transport:
      Some proteins transports various substances, such as oxygen, ions, and so on.

    - Information transfer:
      For example, hormones.

Alcohol dehydrogenase oxidizes alcohols to aldehydes or ketones
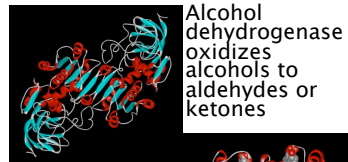
Haemoglobin carries oxygen
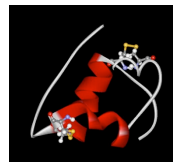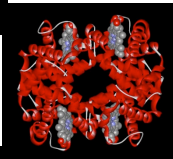
Insulin controls the amount of sugar in the blood

Estimated Functional Roles (by % of Proteins) of the Proteome in a Complex Organism

Protein Functions

- Transport
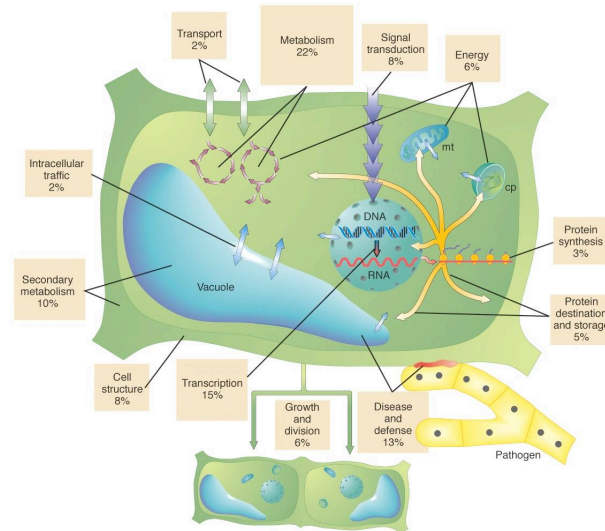- Regulatory
- Motor
- Fibrous Protein
- Enzyme
- Immunoglobulin

$$\text{H}_3\overset{+}{\text{N}} - \underset{\underset{\text{R}}{|}}{\overset{\overset{\text{COO}^-}{|}}{\text{C}}} - \text{H}$$

Figure 3-2
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

(a) L-Alanine   D-Alanine

(b) L-Alanine   D-Alanine

(c) L-Alanine   D-Alanine
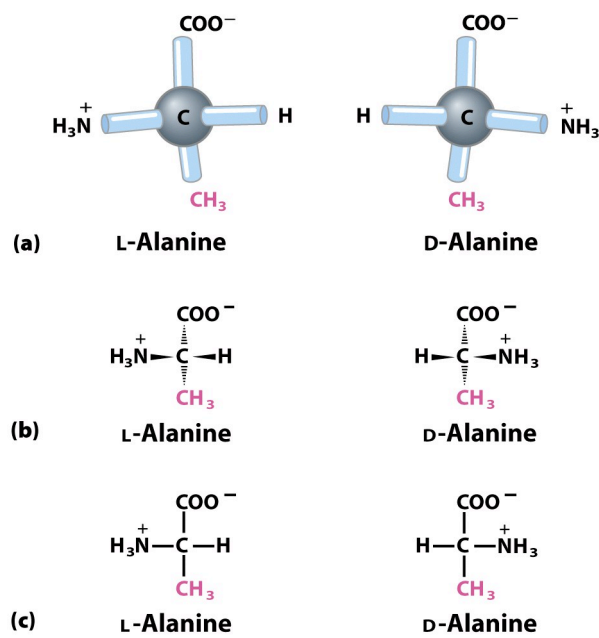
Figure 3-3
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

20 Amino acids

Glycine (G) · Alanine (A) · Valine (V) · Isoleucine (I) · Leucine (L)
Proline (P) · Methionine (M) · Phenylalanine (F) · Tryptophan (W) · Asparagine (N)
Glutamine (Q) · Serine (S) · Threonine (T) · Tyrosine (Y) · Cysteine (C)
Asparatic acid (D) · Glutamic acid (E) · Lysine (K) · Arginine (R) · Histidine (H)

White: Hydrophobic, Green: Hydrophilic, Red: Acidic, Blue: Basic

α-amino aicds
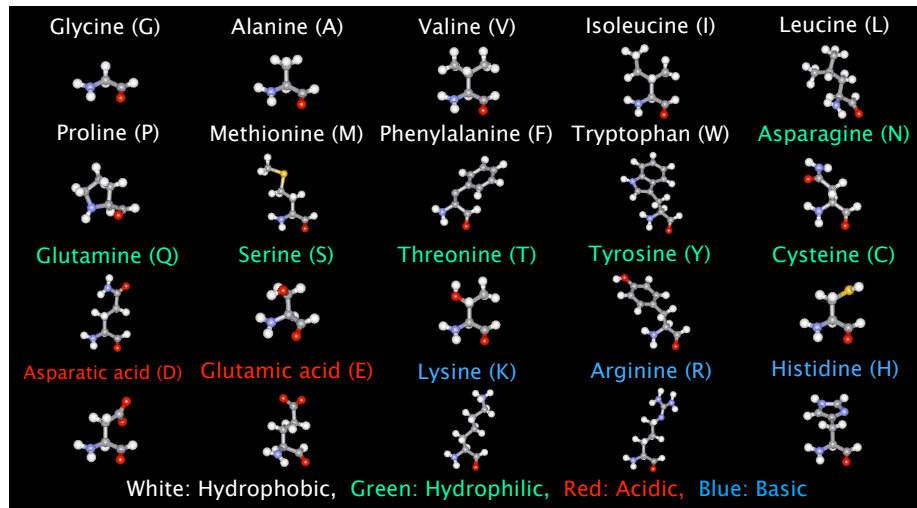
Ø All proteins are composed of amino acids
Ø Twenty common amino acids
Ø All amino acids are primary amino acids except for proline
Ø A primary amine group is attached to the α-carbon of carboxyl group
Ø Except for Glycine, all other amino acids have at least one chiral centers
Ø All chiral amino acids are belong to L-amino acids

^{1}CHO
^{1}CHO

HO $—$ ^{2}C $◀—$ H
H $—$ C $◀—$ OH

^{3}CH$_2$OH
CH$_2$OH

L-Glyceraldehyde
D-Glyceraldehyde

COO$^-$
COO$^-$
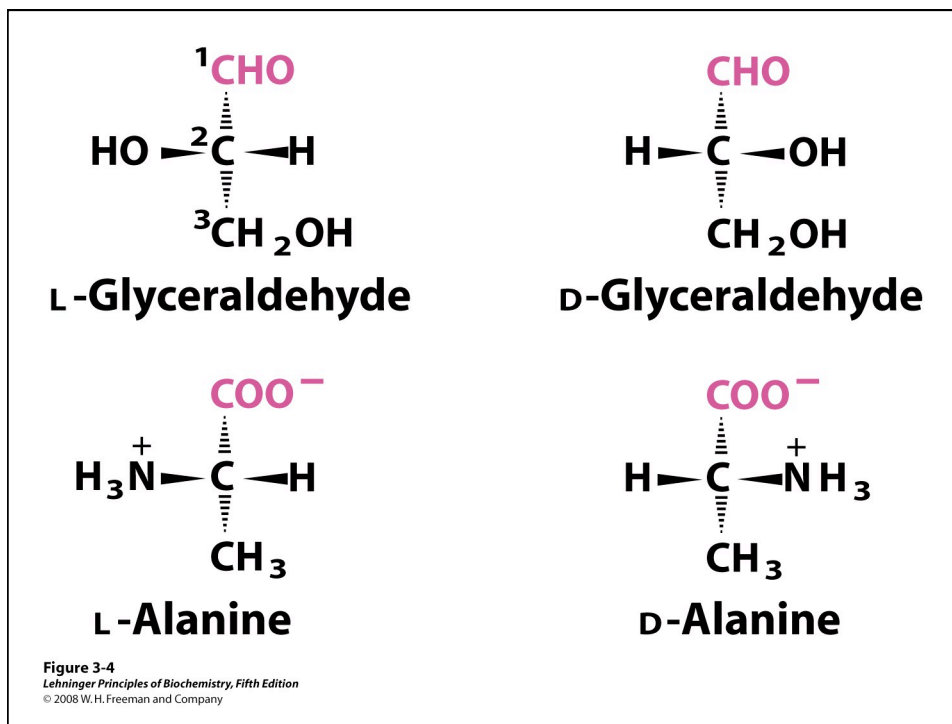
H$_3$$\overset{+}{N}$ $—$ C $◀—$ H
H $—$ C $◀—$ $\overset{+}{N}$H$_3$

CH$_3$
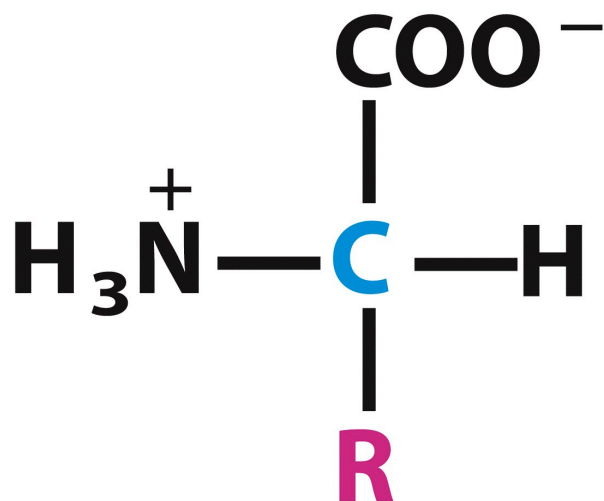CH$_3$

L-Alanine
D-Alanine

Figure 3-4
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Amino acids vary in

Size

Structure

Electric charge

Solubility in water

Classification of amino acids

Classified by polarity of side chains

> hydrophobic: water fearing, non-polar side chains
>
> hydrophilic: water loving, polar neutral
>
> positively charged
>
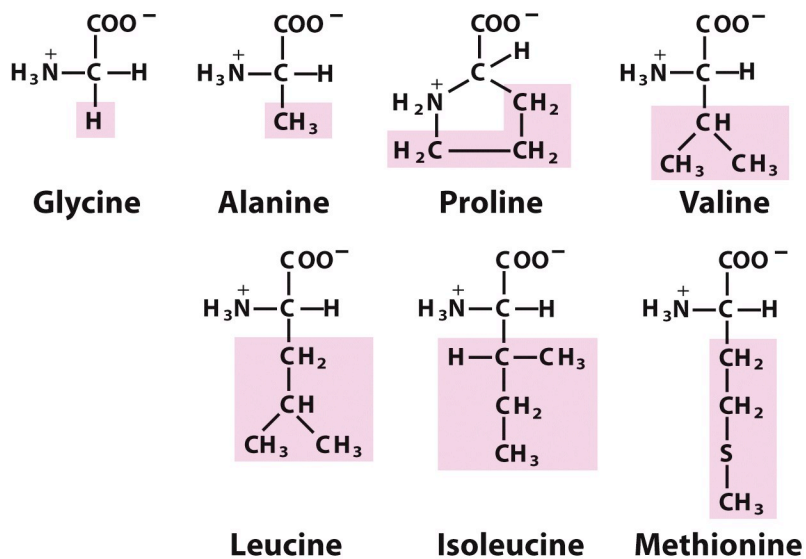> negatively charged
>
> aromatic

Nonpolar, aliphatic R groups

Glycine   Alanine   Proline   Valine

Leucine   Isoleucine   Methionine

Figure 3-5 part 1
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Aromatic R groups

Phenylalanine **Tyrosine** **Tryptophan**

Figure 3-5 part 2
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Polar, uncharged R groups

Serine **Threonine** **Cysteine**

Asparagine **Glutamine**

Figure 3-5 part 3
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Positively charged R groups

Lysine Arginine Histidine

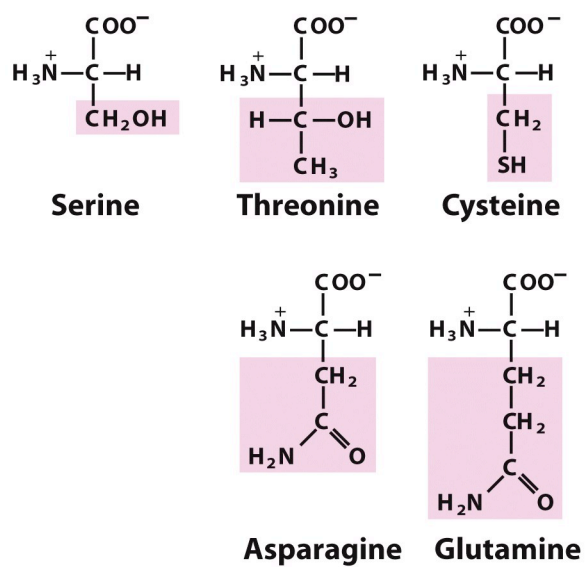**Figure 3-5 part 4**
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Negatively charged R groups

Aspartate Glutamate

Figure 3-5 part 5
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-7
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-8a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Phosphoserine

Phosphothreonine

Phosphotyrosine

σ-N-Methylarginine

6-N-Acetyllysine

Glutamate methyl ester

Adenylyltyrosine

Figure 3-8b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

$$H_3\overset{+}{N}-CH_2-CH_2-CH_2-CH-COO^-$$
$$\overset{|}{\underset{+NH_3}{}}$$

Ornithine

$$H_2N-C-N-CH_2-CH_2-CH_2-CH-COO^-$$

Citrulline

Figure 3-8c
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-6
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Box 3-1 figure 1
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Titration curve of amino acids

Neutral side chain

Acidic side chain

Basic side chain

$$R - \overset{\overset{\displaystyle H}{|}}{C} - C = O$$
$$\underset{H_2N \quad OH}{}$$

Nonionic form

$$R - \overset{\overset{\displaystyle H}{|}}{C} - C = O$$
$$\underset{H_3N^+ \quad O^-}{}$$

Zwitterionic form

$$R - \overset{\overset{\displaystyle H}{|}}{\underset{^+NH_3}{C}} - COO^- \; \rightleftharpoons \; R - \overset{\overset{\displaystyle H}{|}}{\underset{NH_2}{C}} - COO^- + H^+$$

**Zwitterion
as acid**

$$R - \overset{\overset{\displaystyle H}{|}}{\underset{^+NH_3}{C}} - COO^- + H^+ \; \rightleftharpoons \; R - \overset{\overset{\displaystyle H}{|}}{\underset{^+NH_3}{C}} - COOH$$

**Zwitterion
as base**

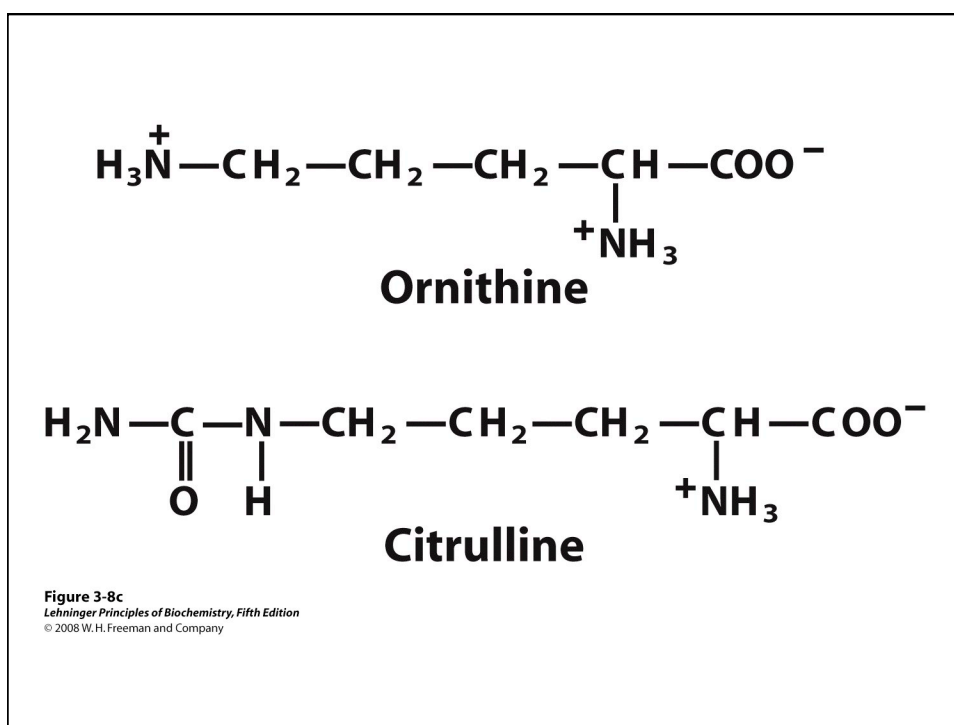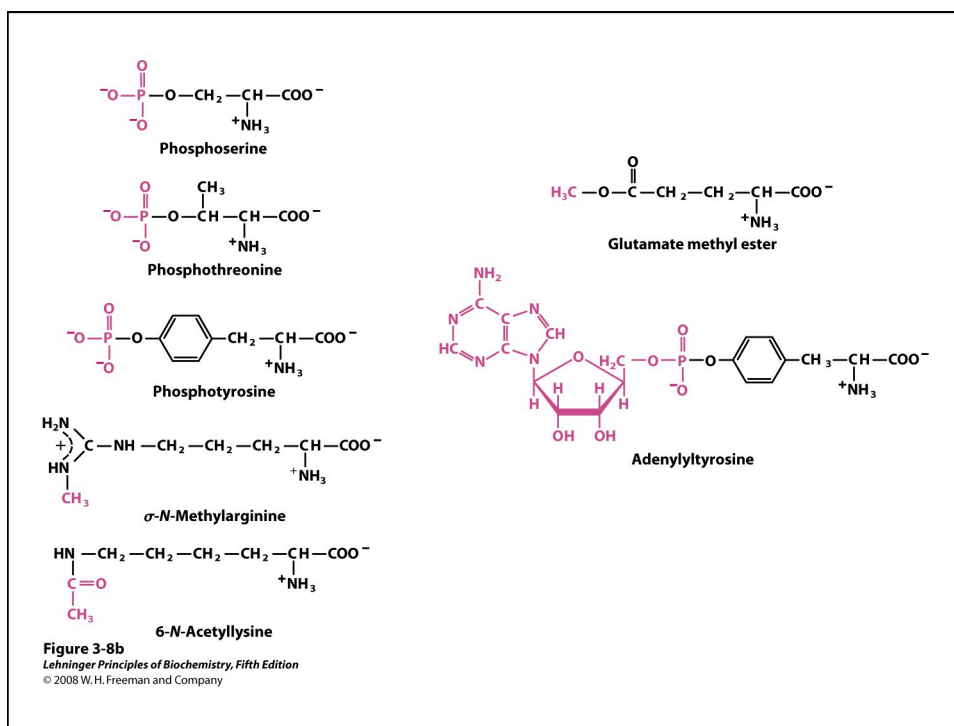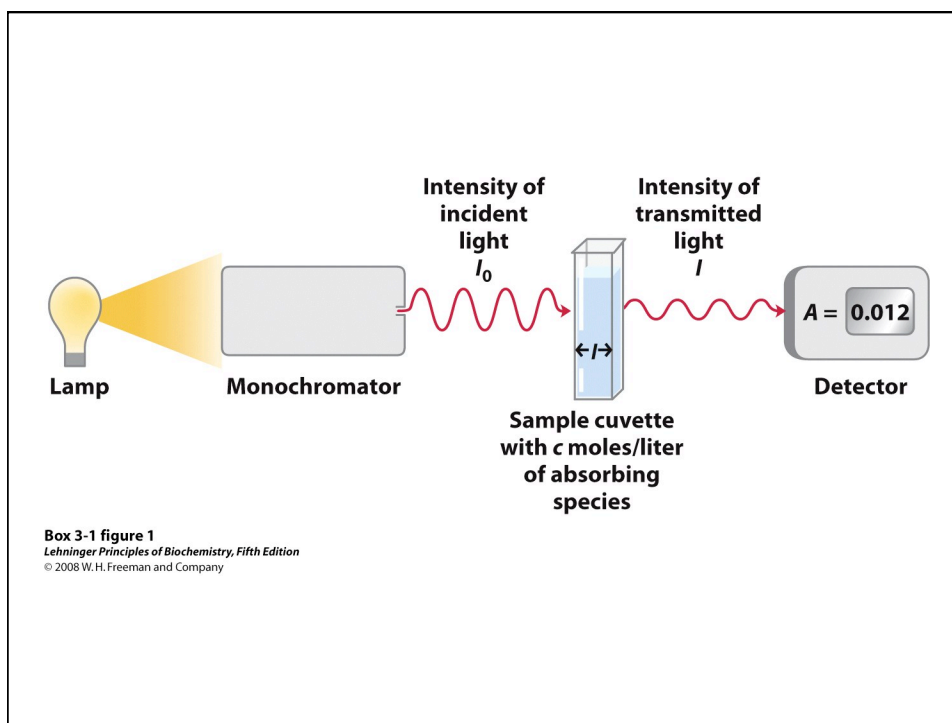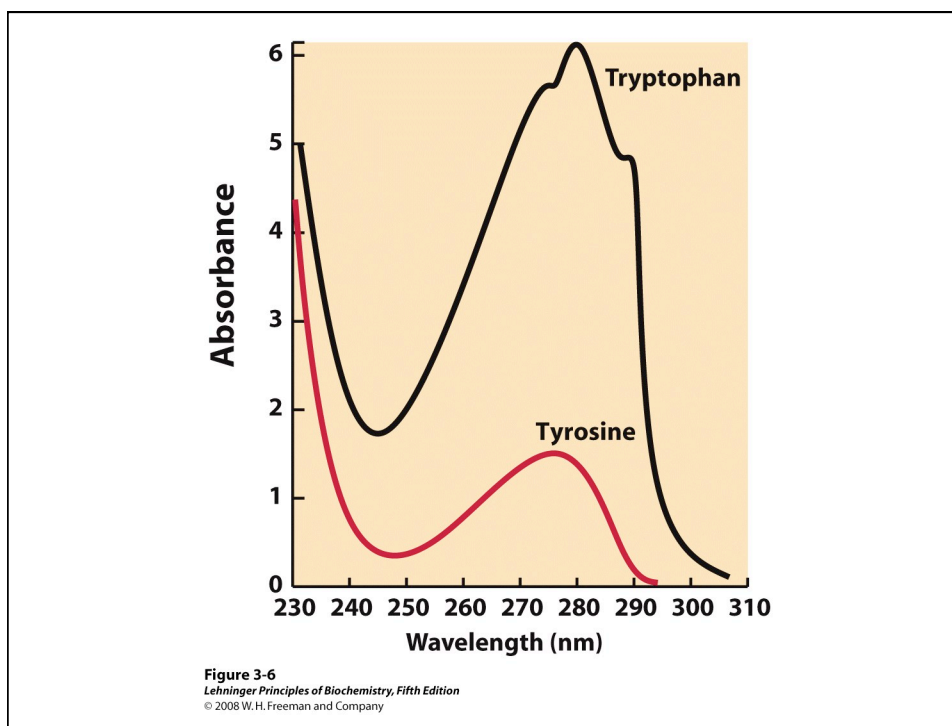**Figure 3-9**
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Net charge: $+1$ 0 -1

Unnumbered 3 p79
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Glycine

$pK_2 = 9.60$

$pI = 5.97$

$pK_1 = 2.34$

pH

OH$^-$ (equivalents)

Figure 3-10
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-11
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-12a
Lehninger Principles of Biochemistry, Fifth Edition
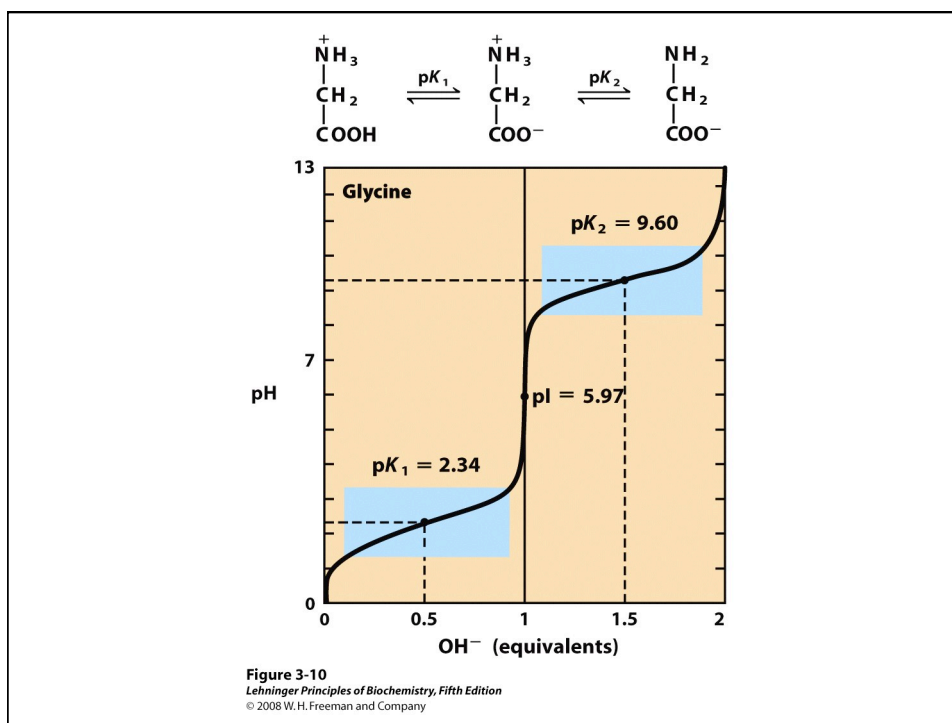© 2008 W. H. Freeman and Company

Figure 3-12b
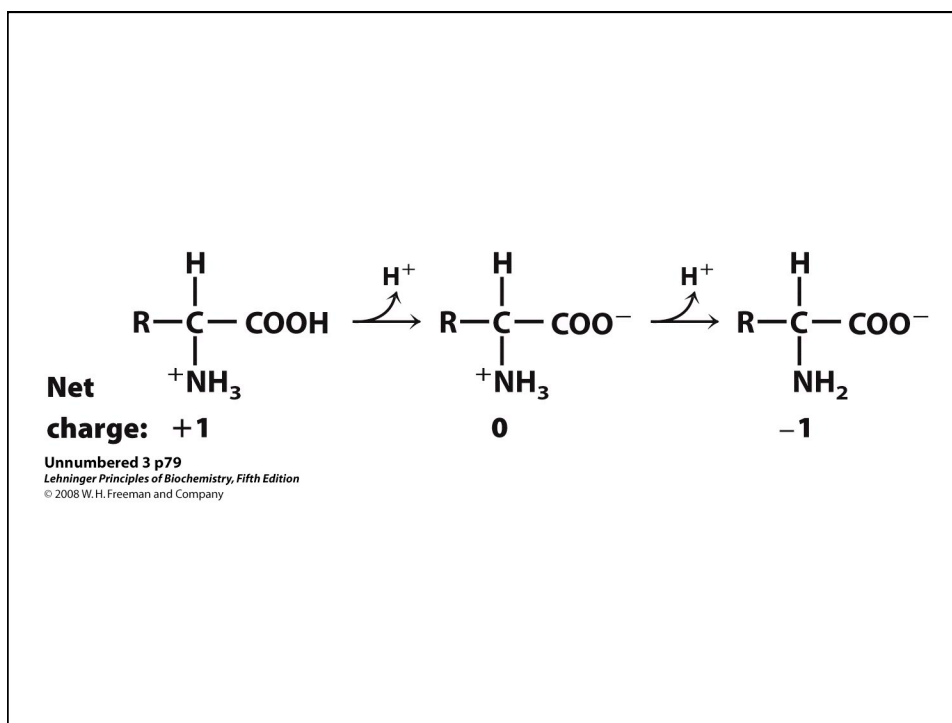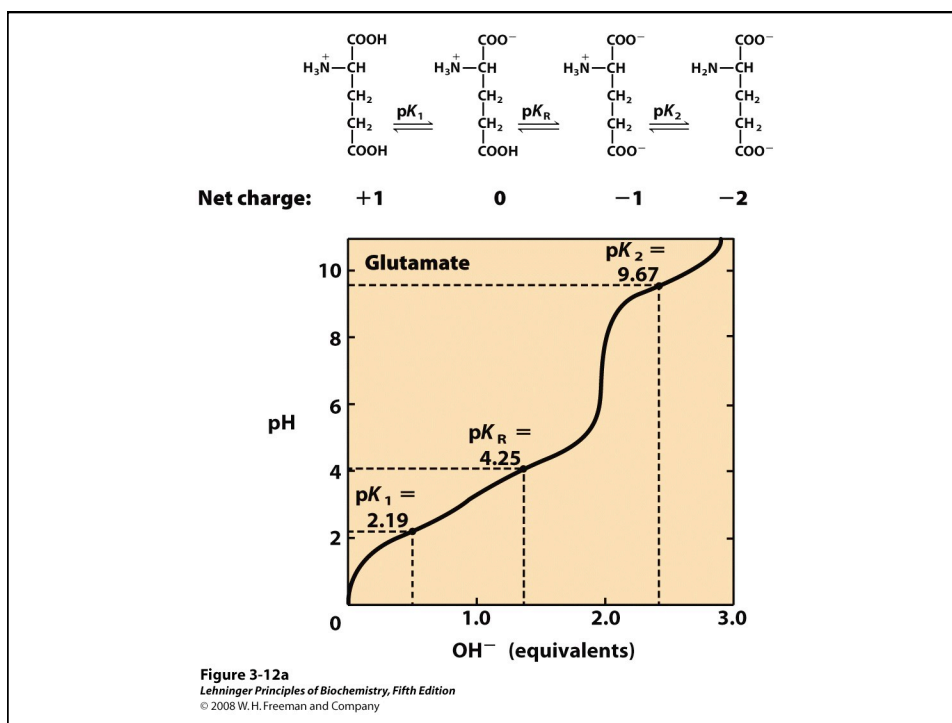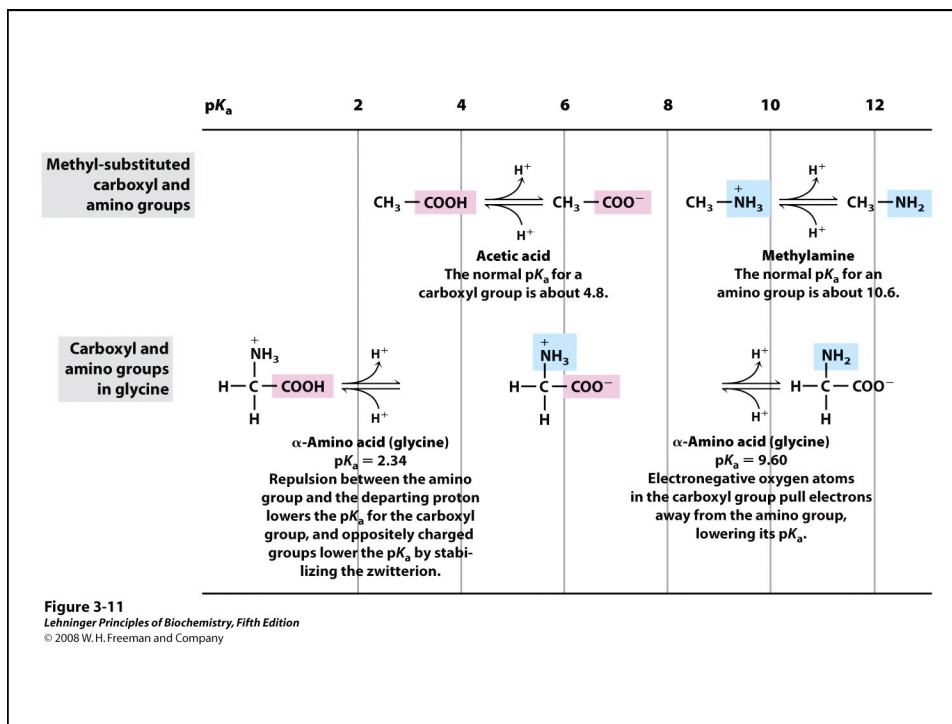*Lehninger Principles of Biochemistry, Fifth Edition*
© 2008 W. H. Freeman and Company

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

*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.

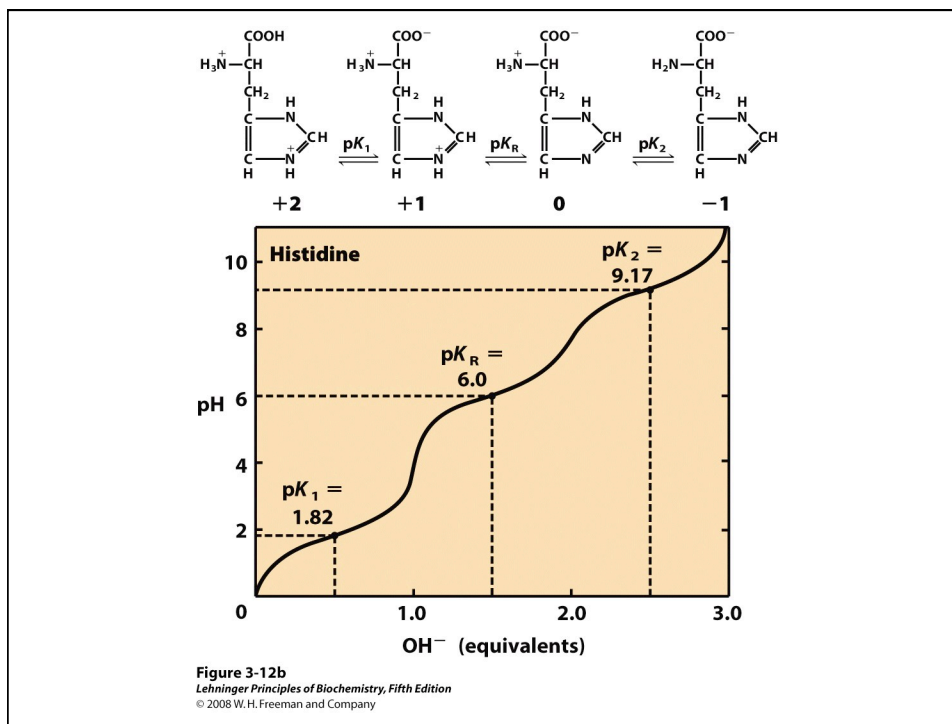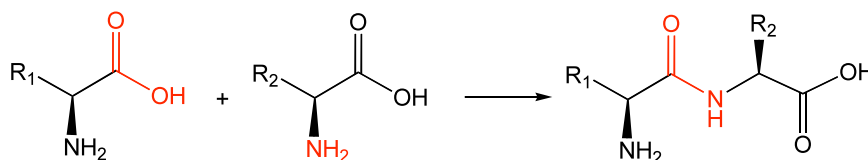**TABLE 3–1** Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r	pK_1 (—COOH)	pK_2 (—NH$_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)[†]
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.

Peptide Bond

Peptides

Dipeptide

Tripeptide

Tetrapeptide

Oligopeptide

Polypeptide

Protein

Peptide Bond

N-terminal
Amino Terminal

C-Terminal
Carboxy Terminal

**Replace -ine by -yl
but keep the last -ine !**

Ala-Gly-Arg   Alanylglycylarginine

Structure of Proteins

Primary structure

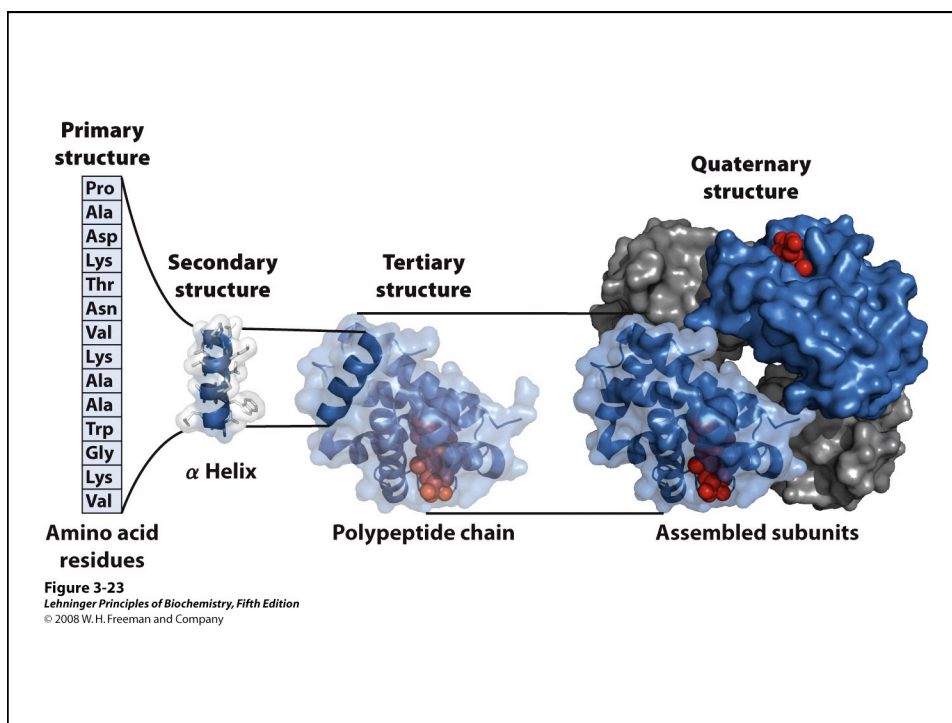**Secondary structure**

Tertiary structure

Quaternary structure

Figure 3-23
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

TABLE 3–2	Molecular Data on Some Proteins		
	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (chicken egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

Table 3-2
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

TABLE 3–3	Amino Acid Composition of Two Proteins	
	Number of residues per molecule of protein*	
Amino acid	Bovine cytochrome c	Bovine chymotrypsinogen
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	3	10
Glu	9	5
Gly	14	23
His	3	2
Ile	6	10
Leu	6	19
Lys	18	14
Met	2	2
Phe	4	6
Pro	4	9
Ser	1	28
Thr	8	23
Trp	1	8
Tyr	4	4
Val	3	23
Total	104	245

*In some common analyses, such as acid hydrolysis, Asp and Asn are not readily distinguished from each other and are together designated Asx (or B). Similarly, when Glu and Gln cannot be distinguished, they are together designated Glx (or Z). In addition, Trp is destroyed by acid hydrolysis. Additional procedures must be employed to obtain an accurate assessment of complete amino acid content.

Table 3-3
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

TABLE 3–4	Conjugated Proteins	
Class	Prosthetic group	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Table 3-4
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Peptide & Protein Charge

Arg—Ser—Gly—Asn—Gly—Phe—Pro—Lys—Met—Glu

pI = ?

Group	pKa	pH=1	pH=3	pH=7	pH=10	pH=11	pH=13
-COOH	2.2	0	-	-	-	-	-
-NH2	8.8	+	+	+	0	0	0
Glu	4.3	0	0	-	-	-	-
Lys	10.8	+	+	+	+	0	0
Arg	12.5	+	+	+	+	+	0
Net Charge		+3	+2	+1	0	-1	-2

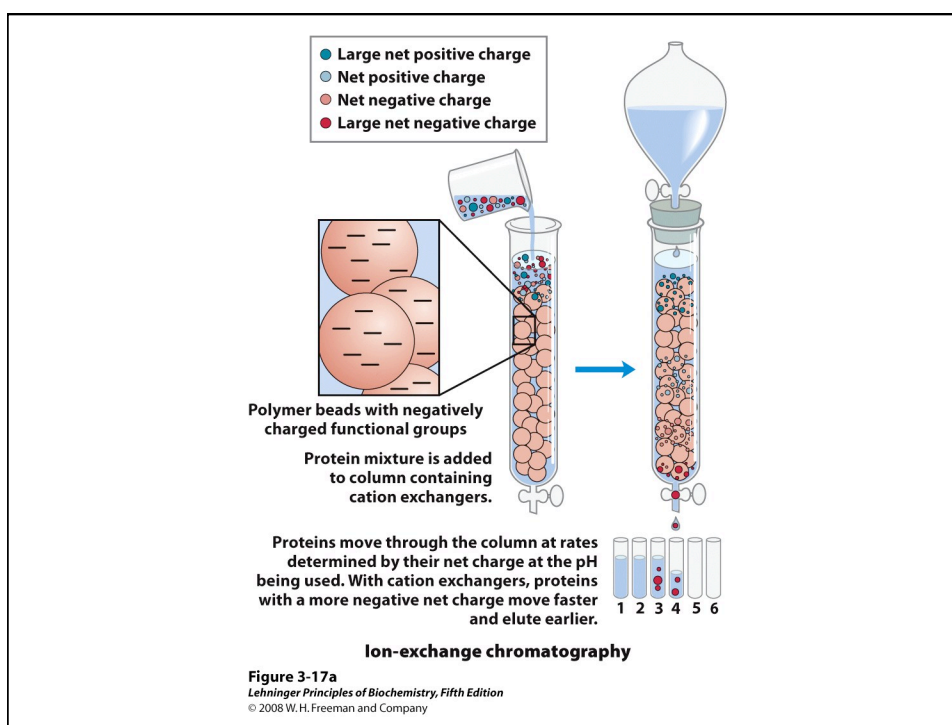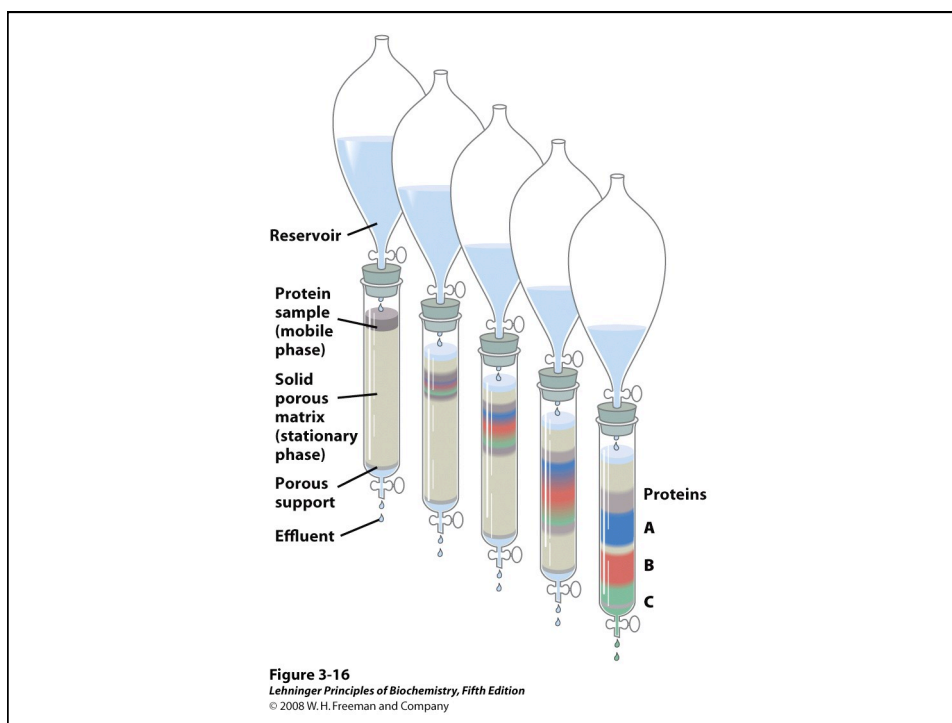pI = (8.8 + 10.8)/2 = 9.8

Protein Separation/Purification

- In general, proteins contain > 40 residues
  - Minimum needed to fold into tertiary structure
- Usually 100-1000 residues; percent of each AA varies
- Proteins separated based on differences in size and composition
- Proteins must be pure to analyze, determine structure/function
- Factors to control (to avoid denaturation or chemical degradation)
  - pH
  - Presence of enzymes
  - Temperature
  - Reactive thiol groups
  - Exposure to air, water

Methods of Separation/Purification

- Solubility (salts, solvents, pH, temperature)
- Chromatography
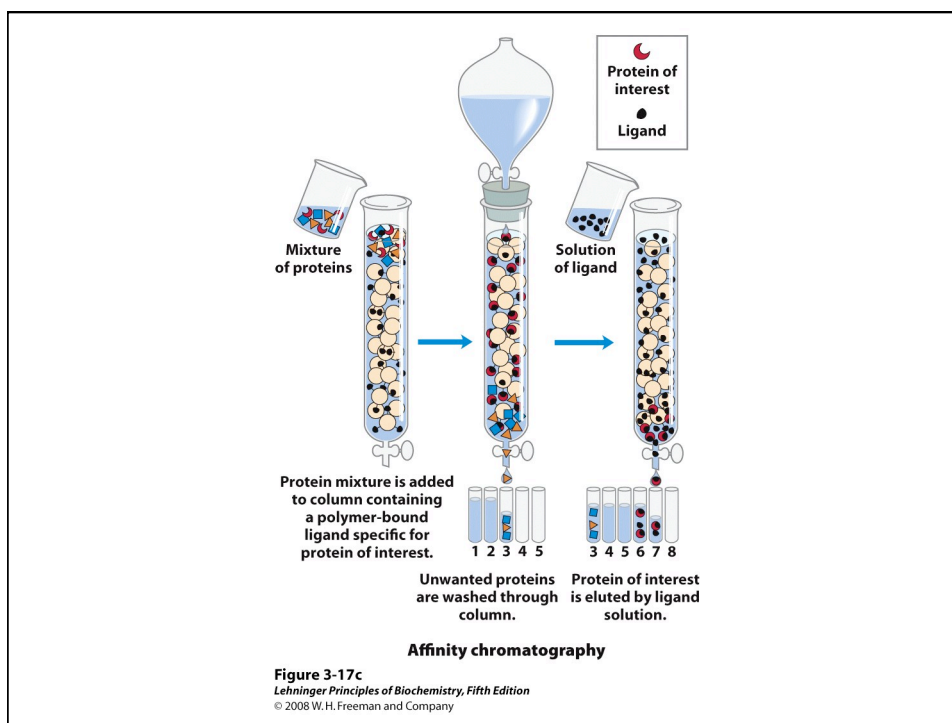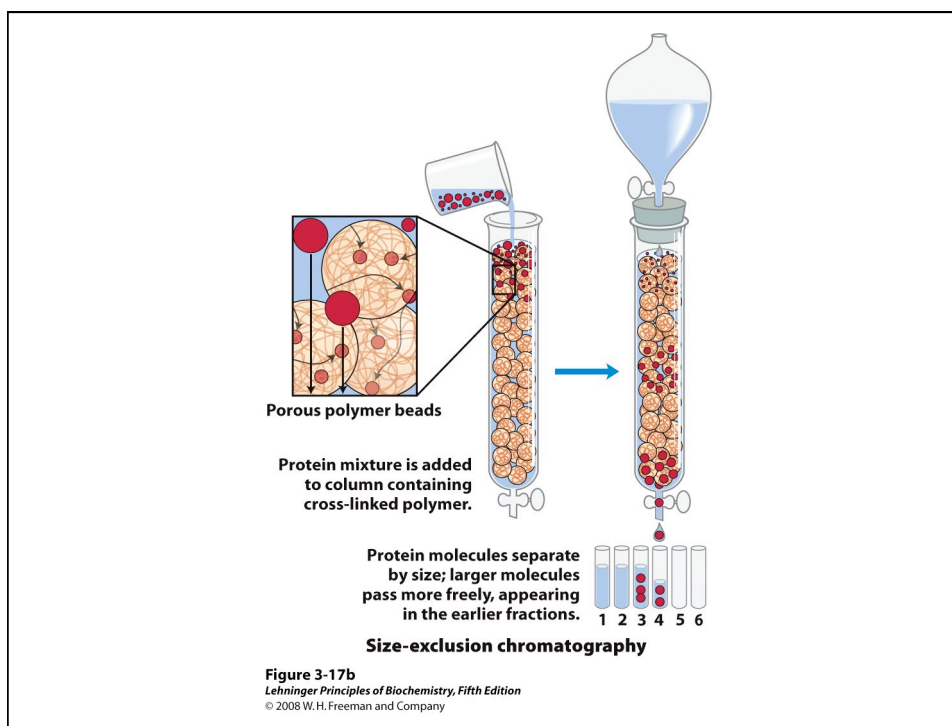 - Ion exchange
  - Gel filtration
  - Affinity
- Electrophoresis

Reservoir

Protein sample (mobile phase)

Solid porous matrix (stationary phase)

Porous support

Effluent

Proteins

A

B

C

Figure 3-16
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

● Large net positive charge
● Net positive charge
● Net negative charge
● Large net negative charge

Polymer beads with negatively charged functional groups

Protein mixture is added to column containing cation exchangers.

Proteins move through the column at rates determined by their net charge at the pH being used. With cation exchangers, proteins with a more negative net charge move faster and elute earlier.

1 2 3 4 5 6

Ion-exchange chromatography

Figure 3-17a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Porous polymer beads

**Protein mixture is added
to column containing
cross-linked polymer.**

**Protein molecules separate
by size; larger molecules
pass more freely, appearing
in the earlier fractions.**
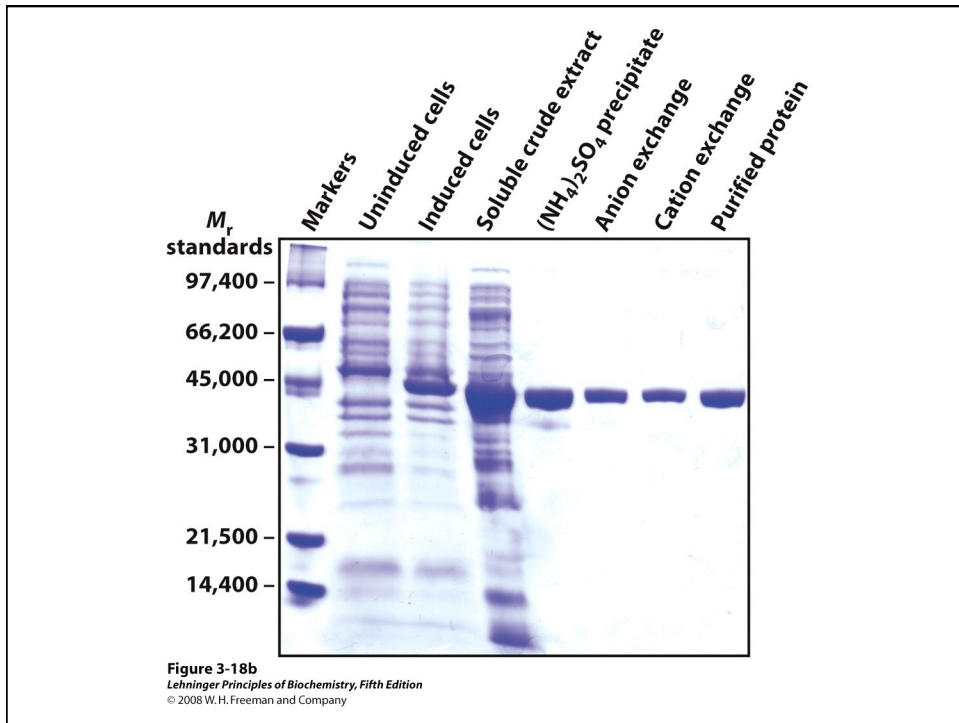
1 2 3 4 5 6

Size-exclusion chromatography

Figure 3-17b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

**Protein
of
interest**

Ligand

**Mixture
of proteins**

**Solution
of ligand**

**Protein mixture is added
to column containing
a polymer-bound
ligand specific for
protein of interest.**

1 2 3 4 5

3 4 5 6 7 8

**Unwanted proteins
are washed through
column.**

**Protein of interest
is eluted by ligand
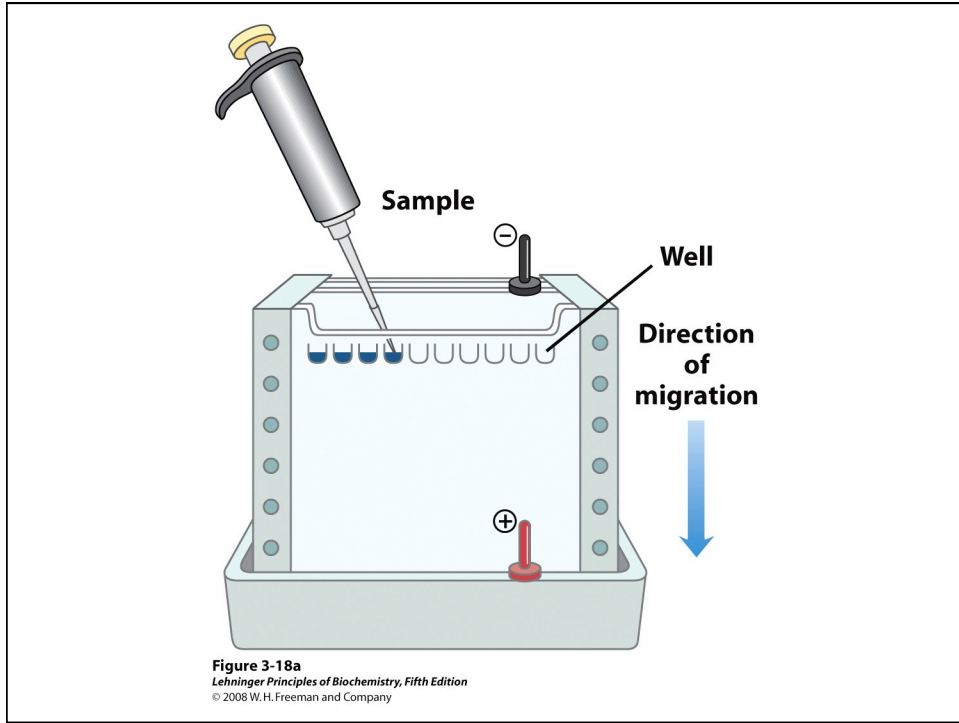solution.**

Affinity chromatography

Figure 3-17c
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

TABLE 3–5	A Purification Table for a Hypothetical Enzyme			
Procedure or step	Fraction volume (mL)	Total protein (mg)	Activity (units)	Specific activity (units/mg)
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation with ammonium sulfate	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

Note: All data represent the status of the sample *after* the designated procedure has been carried out. Activity and specific activity are defined on page 91.
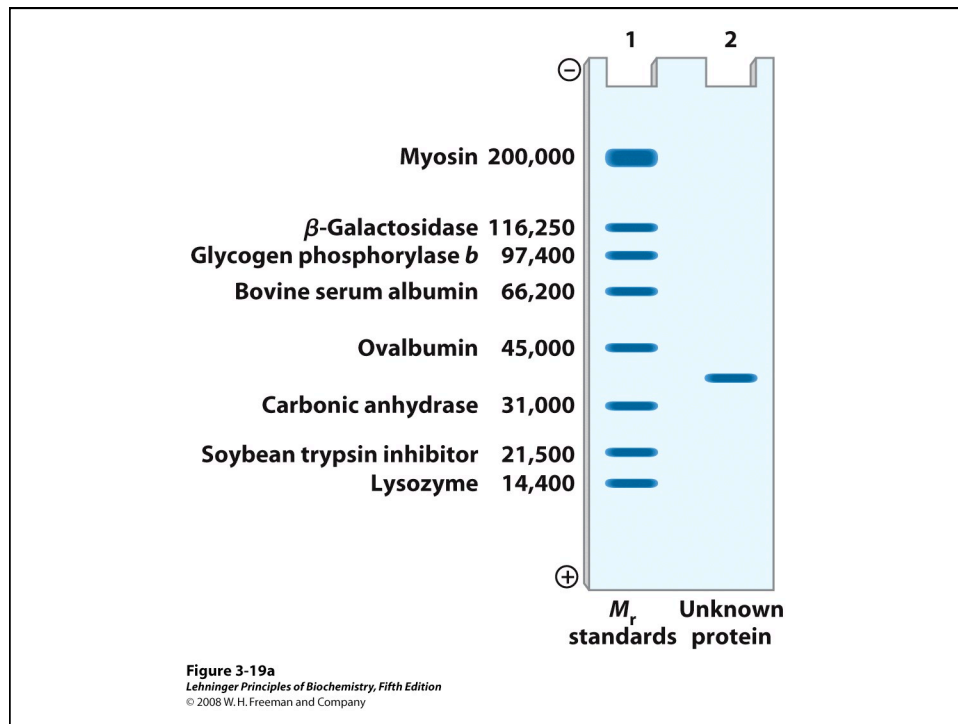
Table 3-5
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

$$Na^+ \ {}^-O-\overset{\overset{\displaystyle O}{\|}}{\underset{\underset{\displaystyle O}{\|}}{S}}-O-(CH_2)_{11}CH_3$$
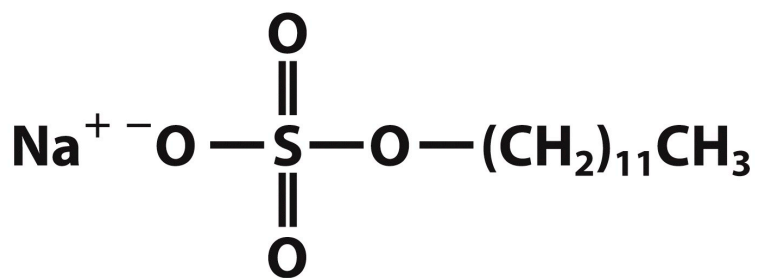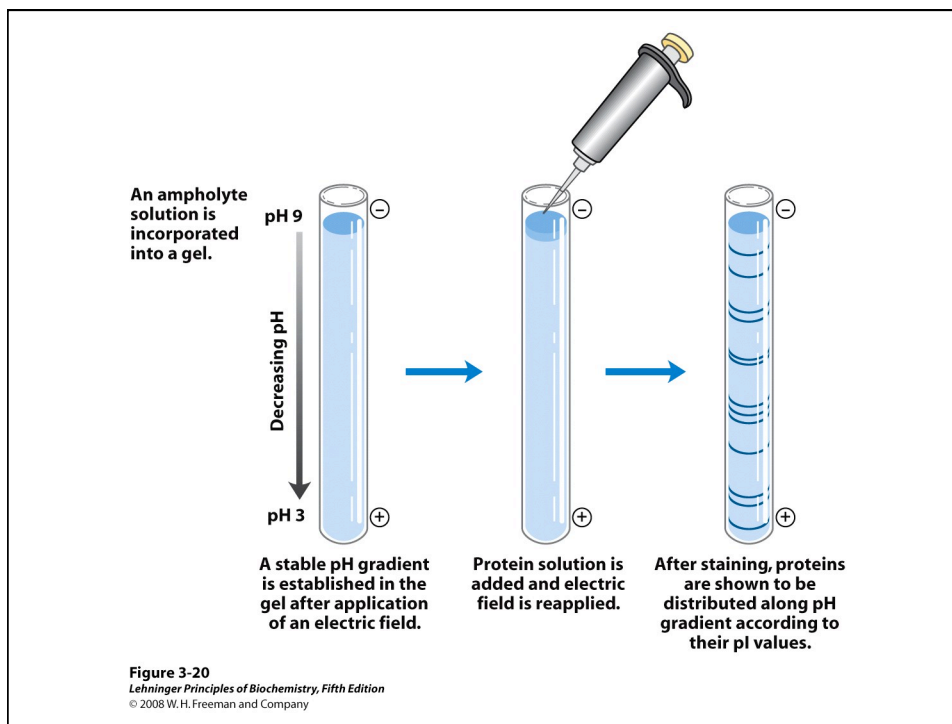
Sodium dodecyl sulfate (SDS)

Figure 3-18a
Lehninger Principles of Biochemistry, Fifth Edition
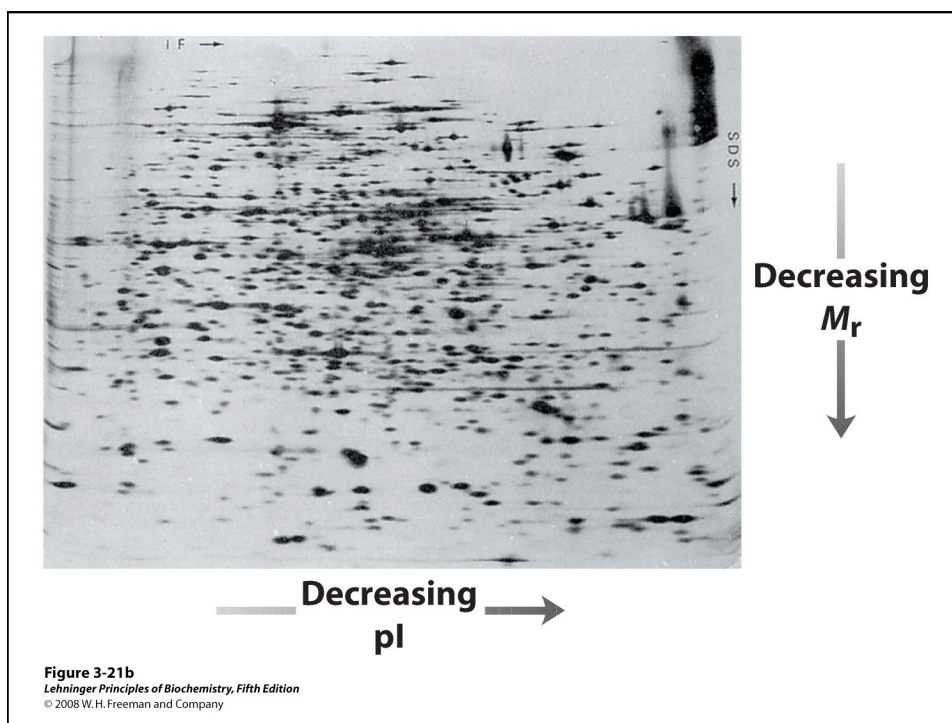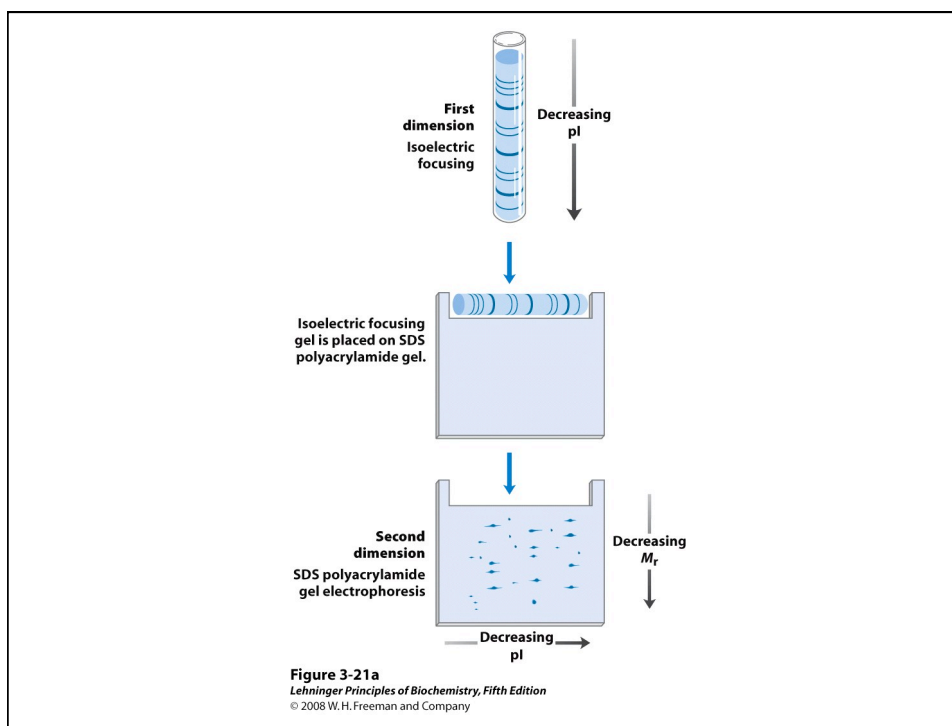© 2008 W. H. Freeman and Company

Figure 3-18b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Myosin 200,000

β-Galactosidase 116,250
Glycogen phosphorylase *b* 97,400
Bovine serum albumin 66,200

Ovalbumin 45,000

Carbonic anhydrase 31,000

Soybean trypsin inhibitor 21,500
Lysozyme 14,400

M_r standards   Unknown protein

Figure 3-19a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Unknown protein

$\log M_r$

Relative migration

Figure 3-19b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-20
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

TABLE 3–6	The Isoelectric Points of Some Proteins
Protein	**pI**
Pepsin	<1.0
Egg albumin	4.6
Serum albumin	4.9
Urease	5.0
β-Lactoglobulin	5.2
Hemoglobin	6.8
Myoglobin	7.0
Chymotrypsinogen	9.5
Cytochrome *c*	10.7
Lysozyme	11.0

Table 3-6
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-21a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-21b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Protein Sequencing

Traditional Chemical Method (Sanger Method)

Genetic Method

Proteomics
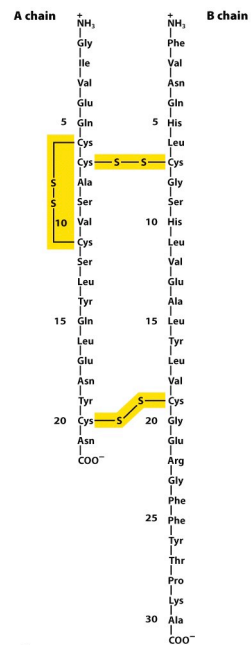
Figure 3-24
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

A Common Strategy for Protein Sequencing

Protein chemists follow a basic strategy when they attempt to determine the sequence of most proteins. This strategy is outlined below. Keep in mind that this strategy is only a guide, and should not inhibit your own ingenuity in solving the sequence of a protein.

Determine the Amino Acid Composition

In order to know which amino acids and how many of each amino acid there are in a polypeptide, we must break the peptide bonds. This can be accomplished with strong acids (i.e. 6N HCl) or strong bases or by exhaustive enzymatic digestion. By performing an acid hydrolysis or base hydrolysis experiment you obtain a minimum length for the polypeptide. The hydrolyzed amino acids are then grouped into different groups according to their pI values, and each amino acid are isolated by HPLC with pH gradient buffer, and determined by reacting with nihydrin. The concentration of individual amino acid are normalized to integral number, and by comparing with the molecular weight, the actual number of individual amino acids within that particular protein can be determined.

Hydrolysis

- A polypeptide can be hydrolyzed by refluxing with 6M hydrochloric acid for 24h
- The individual amino acids can be separated from each other using a cation-exchange resin
  - An acidic solution of the amino acids is passed through the cation-exchange column; the strength of adsorption varies with the basicity of each amino acid (the most basic are held most strongly)
  - Washing the column with a sequence of buffered solutions causes the amino acids to move through it at different rates
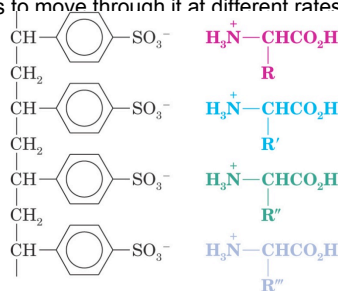
- In the original method, the column eluant is treated with ninhydrin, a dye used for detecting and quantifying each amino acid as it comes off the column


Indane-1,2,3-trione   Ninhydrin

Purple anion
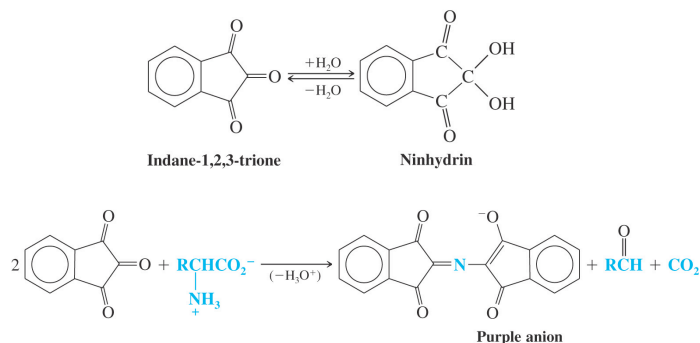
- In modern practice, analysis of amino acid mixtures is routinely accomplished using high performance liquid chromatography (HPLC)

64

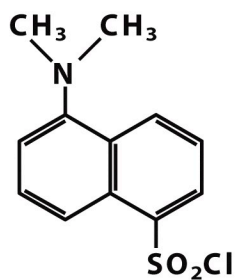– Sanger *N*-Terminal Analysis

- The *N*-terminal end of the polypeptide is labeled with 2,4-dinitrofluorobenzene and the polypeptide is hydrolyzed
 - The labeled *N*-terminal amino acid is separated from the mixture and identified

2,4-Dinitrofluorobenzene (DNFB)   Polypeptide

Labeled polypeptide

Labeled N-terminal amino acid   Mixture of amino acids

Separate and identify
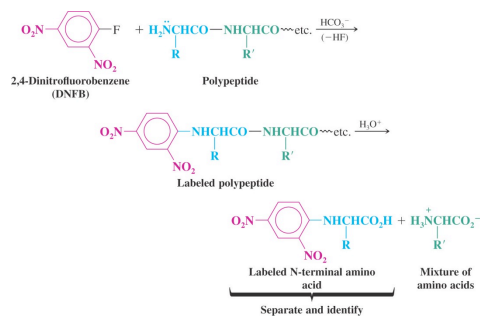
- The Sanger method is not as widely used as the Edman method

33

CH₃   **CH₃**

Dansyl chloride   **Dabsyl chloride**

Unnumbered 3 p95
Lehninger Principles of Biochemistry, Fifth Edition
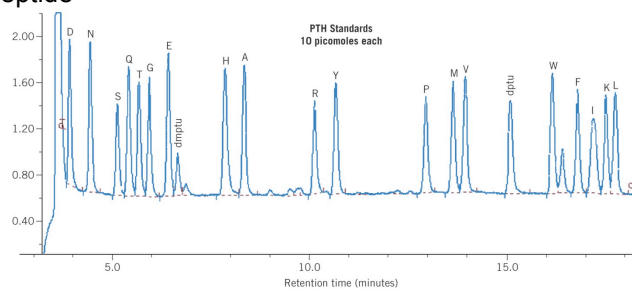© 2008 W. H. Freeman and Company
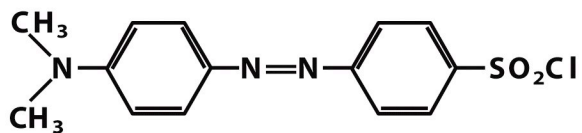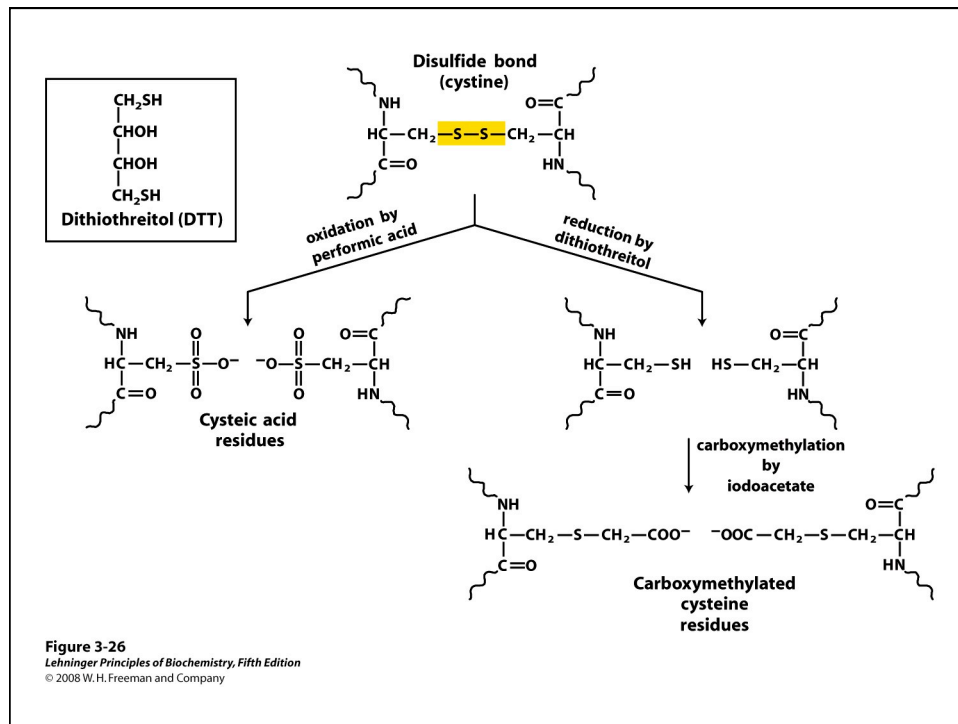

– *C*-Terminal Analysis
  • Enzymes called carboxypeptidases hydrolyze *C*-terminal amino acids selectively
    – The enzyme continues to release each newly exposed *C*-terminal amino acid as the peptide is hydrolyzed; it is necessary to monitor the release of *C*-terminal amino acids as a function of time to identify them

- Primary Structure of Polypeptides and Proteins
  - The sequence of amino acids in a polypeptide is called its primary structure
    - Several methods exist to elucidate the primary structure of peptides
  - Edman Degradation
    - Edman degradation involve sequential cleavage and identification of *N*-terminal amino acids
    - Edman degradation works well for polypeptide sequence analyses up to approximately 60 amino acid residues
      - The N-terminal residue of the polypeptide reacts with phenyl isothiocyanate
      - The resulting phenylthiocarbamyl derivative is cleaved from the peptide chain

68

- The unstable product rearranges to a stable phenylthiohydantoin (PTH) which is purified by HPLC and identified by comparison with PTH standards

- Automated amino acid sequencing machines use the Edman degradation and high performance liquid chromatography (HPLC)
  - One Edman degradation cycle beginning with a picomolar amount of polypeptide can be completed in approximately 30 minutes
  - Each cycle results in identification of the next amino acid residue in the peptide

Figure 3-26
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

– Complete Sequence Analysis
- The Sanger and Edman methods of analysis apply to short polypeptide sequences (up to about 60 amino acid residues by Edman degradation)
- For large proteins and polypeptides, the sample is subjected to partial hydrolysis with dilute acid to give a random assortment of shorter polypeptides which are then analyzed
  – The smaller polypeptides are sequenced, and regions of overlap among them allow the entire polypeptide to be sequenced

- Larger polypeptides can also be cleaved into smaller sequences using *site-specific* reagents and enzymes
  - The use of these agents gives more predictable fragments which can again be overlapped to obtain the sequence of the entire polypeptide
  - Cyanogen bromide (CNBr) cleaves peptide bonds only on the C-terminal side of methionine residues
    Mass spectrometry can be used to determine polypeptide and protein sequences

TABLE 3–7	The Specificity of Some Common Methods for Fragmenting Polypeptide Chains
Reagent (biological source)*	**Cleavage points†**
Trypsin (bovine pancreas)	**Lys, Arg (C)**
***Submaxillarus* protease (mouse submaxillary gland)**	**Arg (C)**
Chymotrypsin (bovine pancreas)	**Phe, Trp, Tyr (C)**
***Staphylococcus aureus* V8 protease (bacterium *S. aureus*)**	**Asp, Glu (C)**
Asp-*N*-protease (bacterium *Pseudomonas fragi*)	**Asp, Glu (N)**
Pepsin (porcine stomach)	**Leu, Phe, Trp, Tyr (N)**
Endoproteinase Lys C (bacterium *Lysobacter enzymogenes*)	**Lys (C)**
Cyanogen bromide	**Met (C)**

*All reagents except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

Table 3-7
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

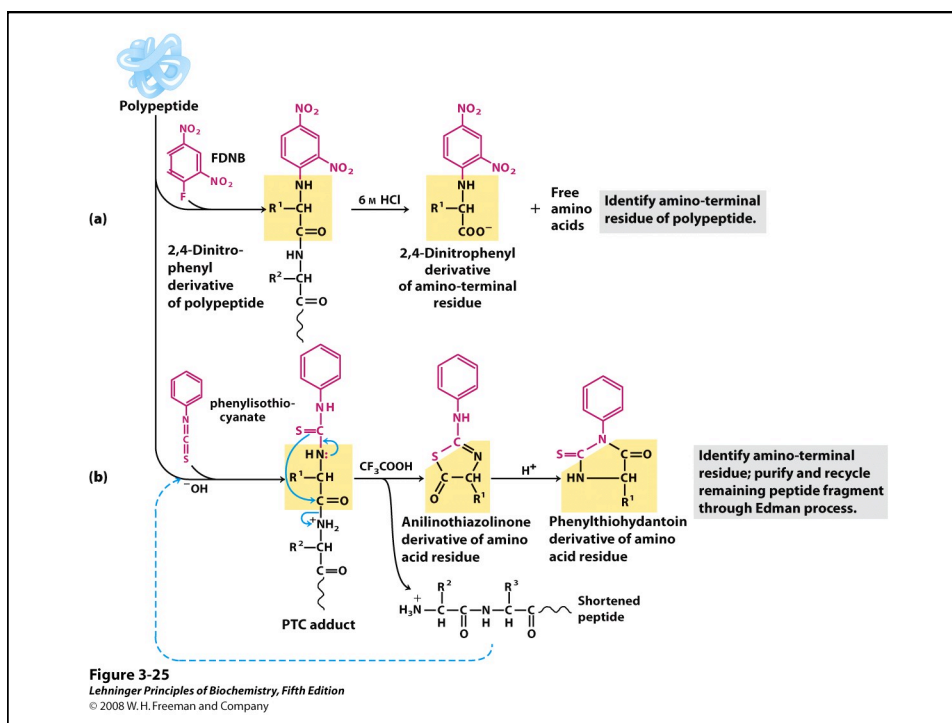**Repeat Steps 3 and 4 to Determine Sub-sequences and Create ÒOverlappingsÓ**

The initial cleavage is generally made as specific as possible in order to generate large peptide fragments. It i s easy to arra nge fewer fragments. These fragments can be positioned relative to one another after treatment of the original polypeptide by a seco nd cleavage procedure that generates fragments whose sequences extend across the initial cleavage points (referred to as overlapping peptides).
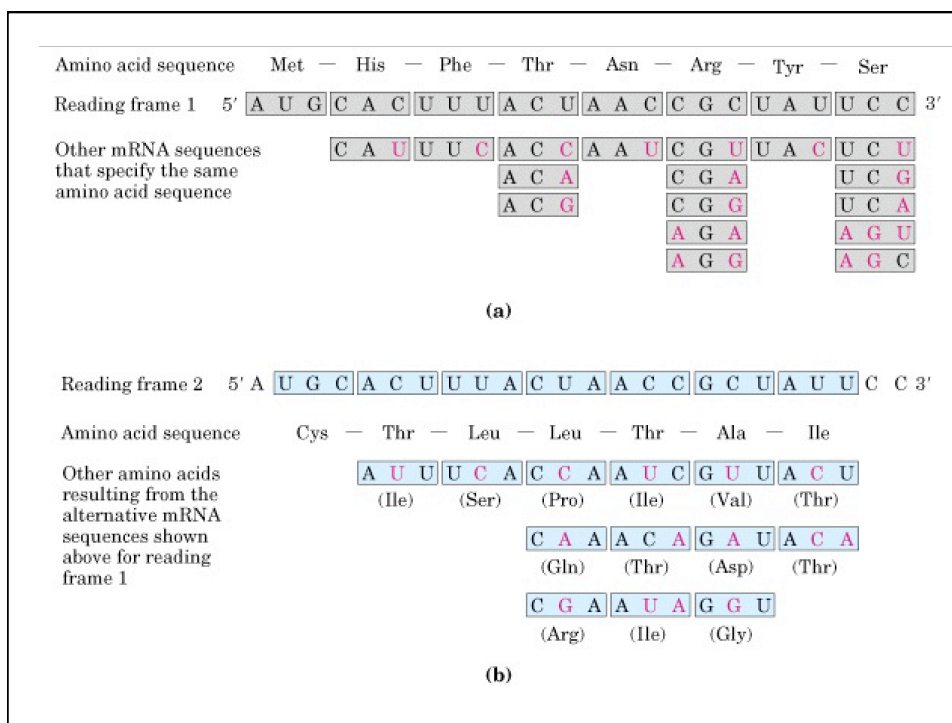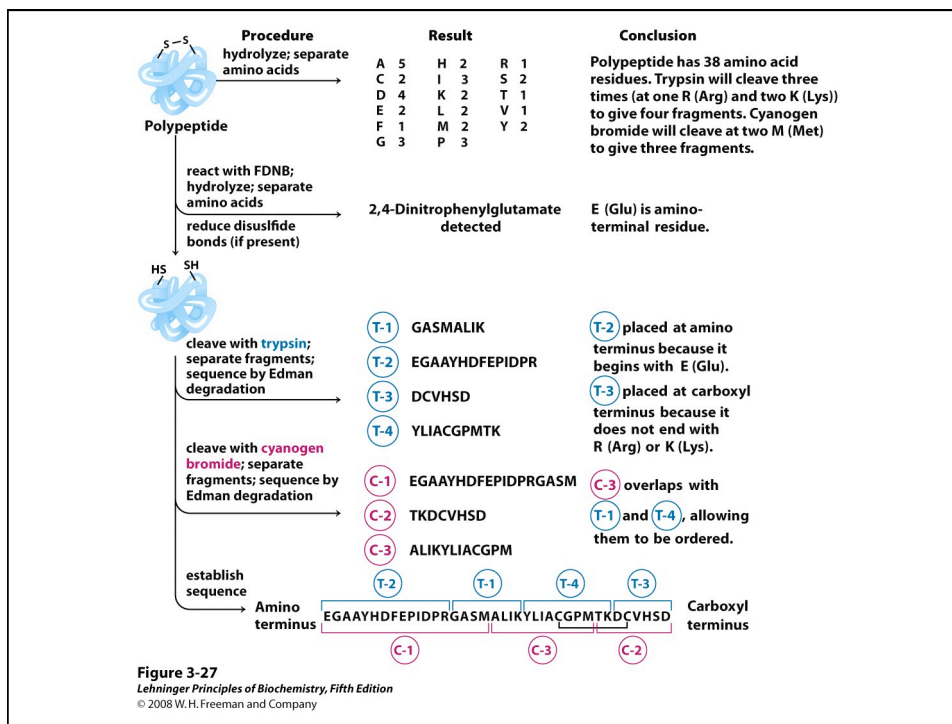
Locate the Disulfide Bonds

No primary structure analysis of a c ysteine-containing protein can be regarded as complete before the presence and location of disulfide bonds has been established.

Reconstruct the Original Protein.

From the overlapping peptides and information gained from the original protein, a uni que sequence for the protein or polypeptide of interest can be worked out.

Figure 3-25
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-27
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

```
                          133
Gene C                    AUG AGA AAA UUC GAC CUA U --- 3'
                          Met – Arg – Lys – Phe – Asp – Leu    ---

Gene A 5' --- GGC GGA AAA UGA
         --- Gly – Gly – Lys – Stop

Gene K 5' --- G GCG GAA AAU GAG AAA AUU CGA CCU AU --- 3'
         --- Ala – Glu – Asn – Glu – Lys – Ile – Arg – Pro –   ---

                          (b)
```
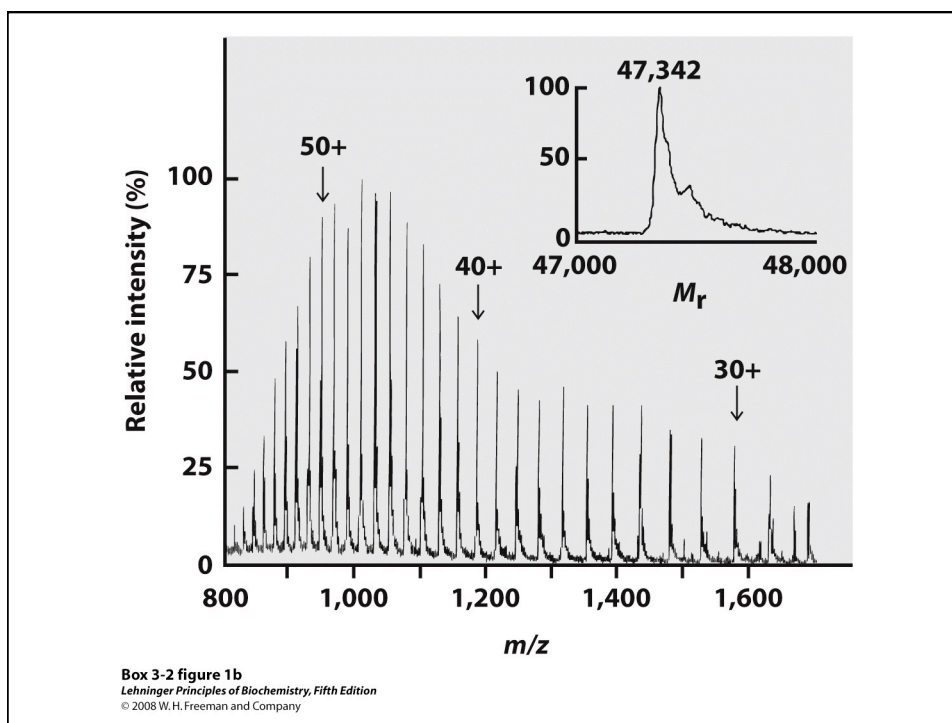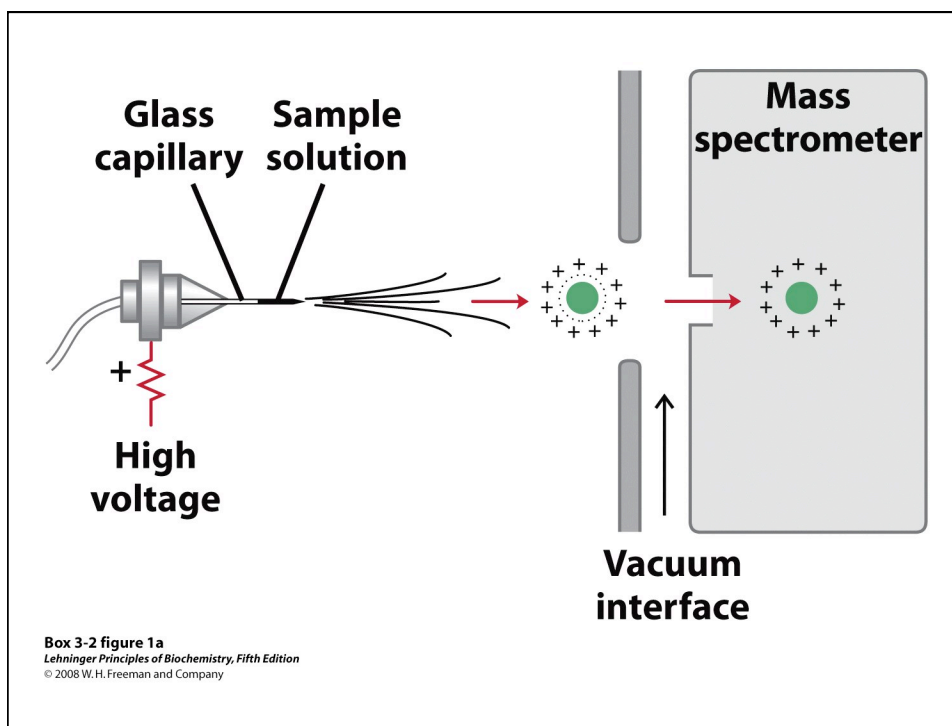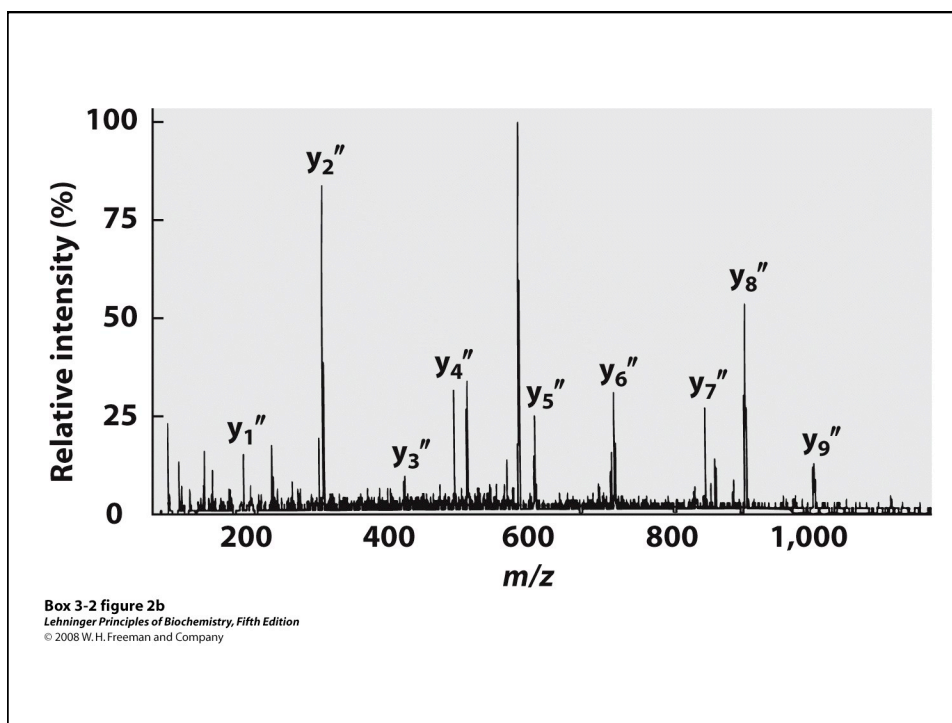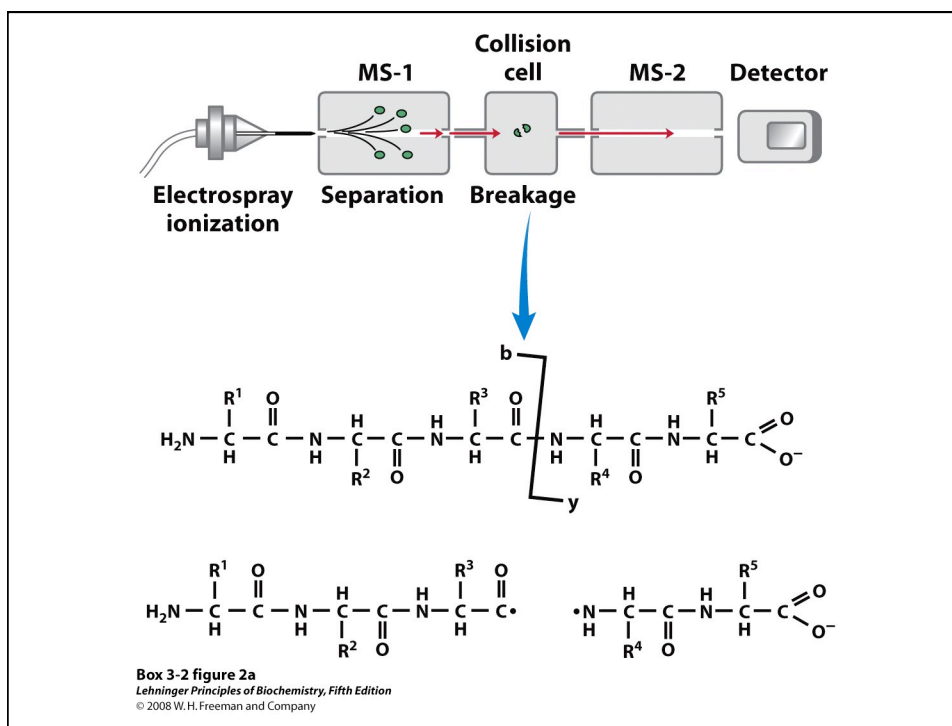
"Ladder sequencing" involves analyzing a polypeptide digest by mass spectrometry, wherein each polypeptide in the digest differs by one amino acid in length; the difference in mass between each adjacent peak indicates the amino acid that occupies that position in the sequence

Mass spectra of polypeptide fragments from a protein can be compared with databases of known polypeptide sequences, thus leading to an identification of the protein or a part of its sequence by matching

Box 3-2 figure 1a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Box 3-2 figure 1b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Box 3-2 figure 2a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Box 3-2 figure 2b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

The Blind Men and the Elephant

by John Godfrey Saxe

American poet John Godfrey Saxe (1816–1887) based the following poem on a fable which was told in India many years ago.

The Blind Men and the Elephant

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind

The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant
Is very like a wall!"

The Second, feeling of the tusk,
Cried, "Ho! what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"

The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"

The Fourth reached out an eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
" 'Tis clear enough the Elephant
Is very like a tree!"

The Fifth, who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can
This marvel of an Elephant
Is very like a fan!?

The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a rope!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!
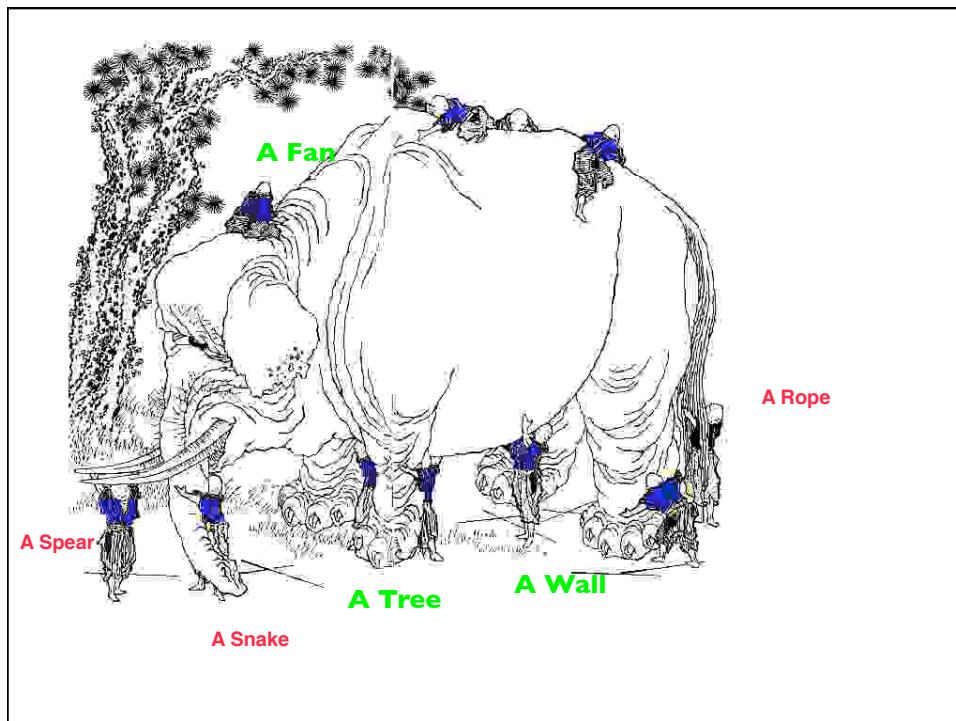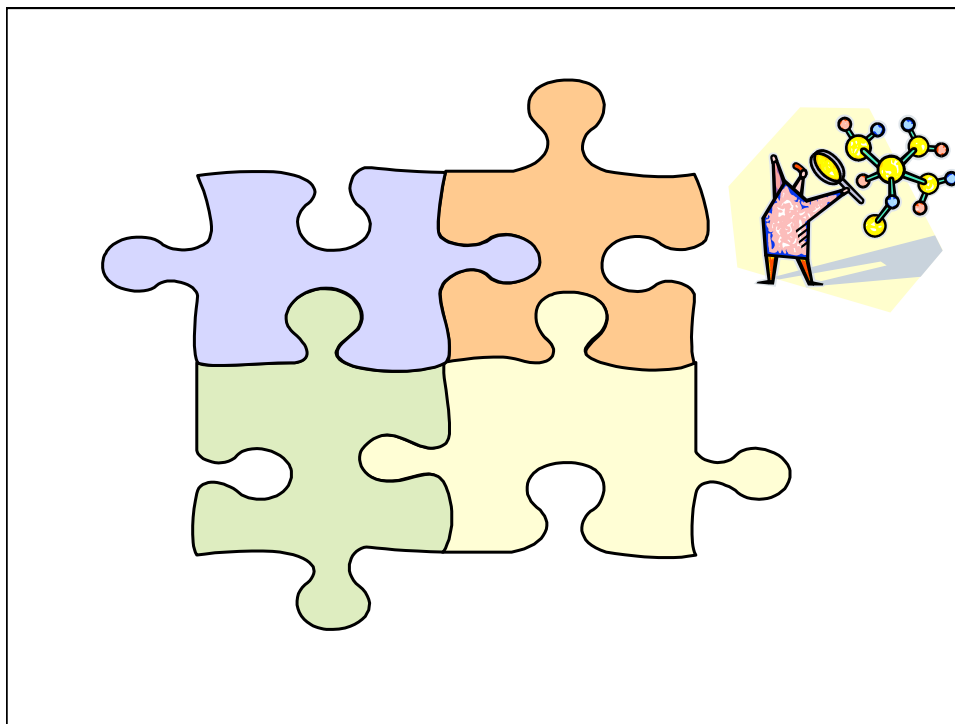
Determine the sequence of protein from the following experimental details.

a. A peptide had the following amino acid composition obtained by acid hydrolysis and amino acid analysis: Arg, Phe, Glu, Asp, Lys, Met, Pro, Ser, His, Trp. N-terminal and C-terminal analyses of the peptide yielded no amino acids. The peptide strongly absorbed UV light at 280 nm.

b. Treatment of the original peptide with trypsin yield 2 peptides. T-1 had an amino acid composition after acid hydrolysis of Ser, Asp, Lys and yielded in Edman degradation Asp 1st and Ser 2nd. T-2 had amino acid composition after acid hydrolysis of His, Phe, Arg, Pro, Glu, Met, Trp and yielded in Edman degradation Met 1st, Glu 2nd, Trp 3rd and His 4th.

c. Treatment of the original peptide with chymotrypsin gave one large peptide with amino acid composition obtained by acid hydrolysis and amino acid analysis the same as the original peptide. Edman degradation yielded His 1st, Phe 2nd, Pro 3rd and Arg 4th.

d. Treatment of the original peptide with cyanogen bromide (CNBr) gave one large peptide with amino acid composition obtained by acid hydrolysis and amino acid analysis the same as the original peptide, except there was no Met and it contained an unknown amino acid. Edman degradation yielded Glu 1st, Trp 2nd, His 3rd and Phe 4th.
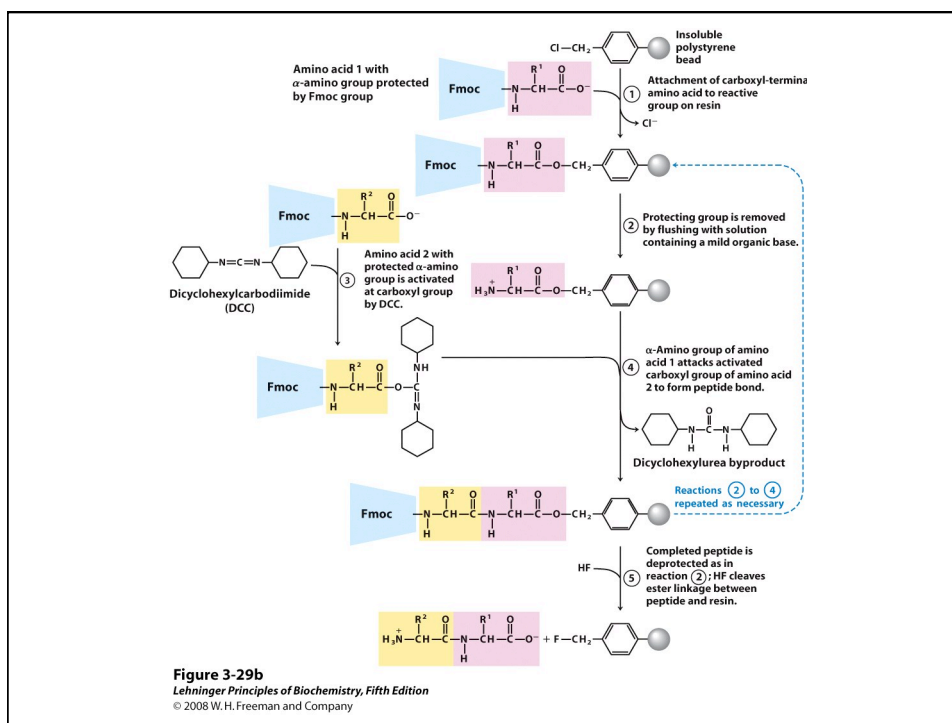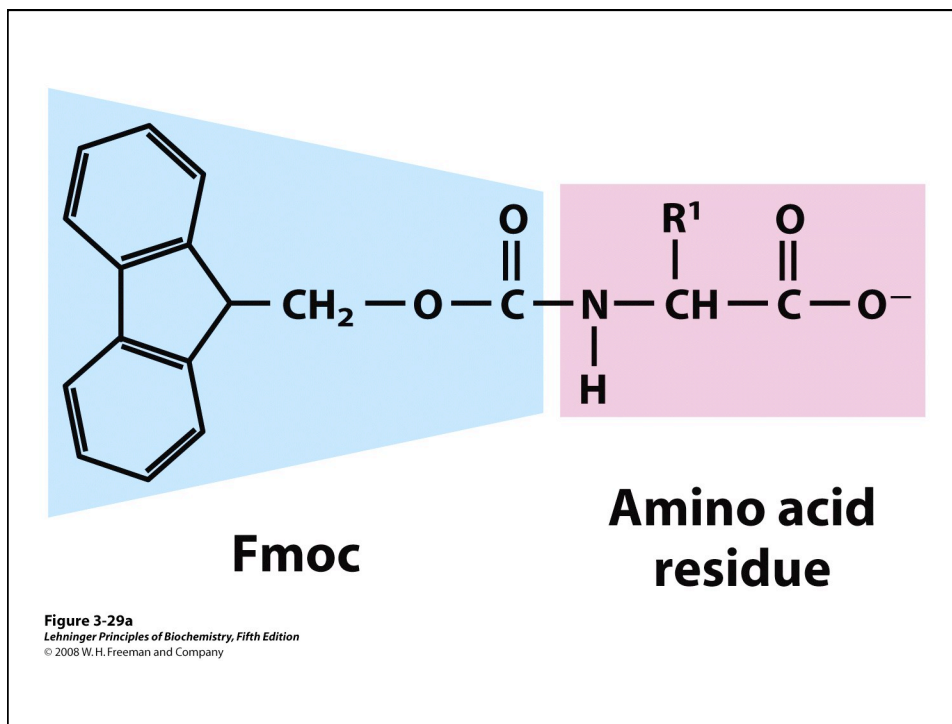
Figure 3-29a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

Figure 3-29b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company

TABLE 3–8	Effect of Stepwise Yield on Overall Yield in Peptide Synthesis	
Number of residues in the final polypeptide	**Overall yield of final peptide (%) when the yield of each step is:**	
	96.0%	99.8%
11	64	98
21	42	96
31	28	94
51	12	90
100	1.7	82

Table 3-8
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W. H. Freeman and Company