

Donald Voet • Judith G. Voet • Charlotte W. Pratt

Fundamentals of Biochemistry

Second Edition

Chapter 5:

Proteins: Primary Structure

Unlimited variation in protein structure and function
 $20 \times 20 \times 20 \dots = 20^{100} = 1.27 \times 10^{130}$



Table 5-1 Compositions of Some Proteins

Protein	Amino Acid Residues	Subunits	Polypeptide Molecular Mass (D)
Proteinase inhibitor III (bitter gourd)	30	1	3,427
Cytochrome <i>c</i> (human)	104	1	11,617
Myoglobin (horse)	153	1	16,951
Interferon- γ (rabbit)	288	2	33,842
Chorismate mutase (<i>Bacillus subtilis</i>)	381	3	43,551
Triose phosphate isomerase (<i>E. coli</i>)	510	2	53,944
Hemoglobin (human)	574	4	61,986
RNA polymerase (bacteriophage T7)	883	1	98,885
Nucleoside diphosphate kinase (<i>Dictyostelium discoideum</i>)	930	6	100,764
Pyruvate decarboxylase (yeast)	2,252	4	245,456
Glutamine synthetase (<i>E. coli</i>)	5,616	12	621,264
Titin (human)	26,926	1	2,993,428

Table 5-1 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

The primary structure of insulin: synthesized as a polypeptide chain and cleaved

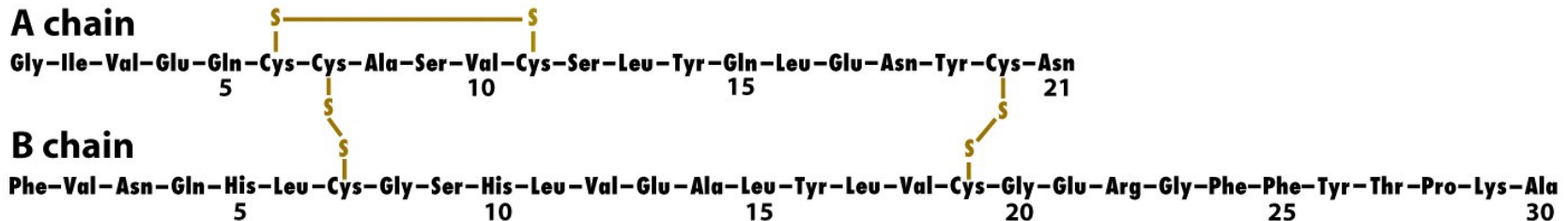


Figure 5-1 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Some restrictions in protein

Size: lower limit, around 40 residues; upper limit, many hundreds

Amino acid composition: most abundant (L, A, G, S, V, E) & least abundant (W, C, M, H)

Nonpolypeptide components

Protein purification and analysis

Purification is mandatory for studying macromolecules

The purification principle is universal to other molecules

A. General approach

Native proteins & recombinant proteins

Intracellular & extracellular

Soluble protein & membrane protein

Stabilizing protein

pH, temperature, degradative enzymes, adsorption to surface, storage

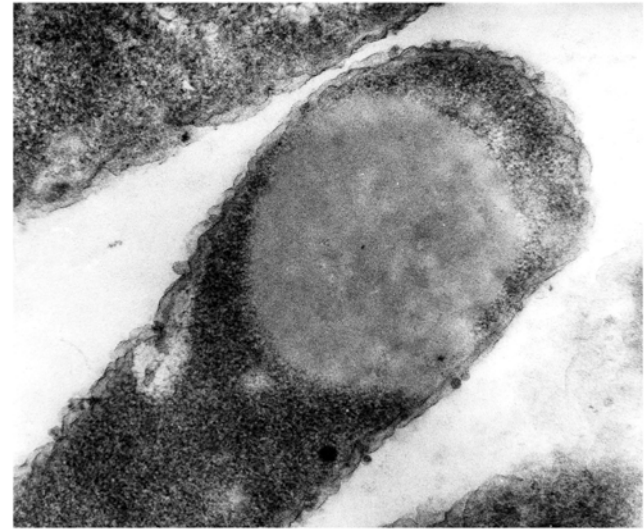


Figure 5-2 Fundamentals of Biochemistry, 2/e

Assaying proteins

Assay: quantitative detection method

catalysis reactions: measure substrate, product, cofactor

colorimetric,

coupled enzymatic reaction

Immunoassays: using antibody

RIA

ELISA

Absorption spectroscopy

Beer-Lambert law: $A = \log (I_0/I) = \epsilon cl$

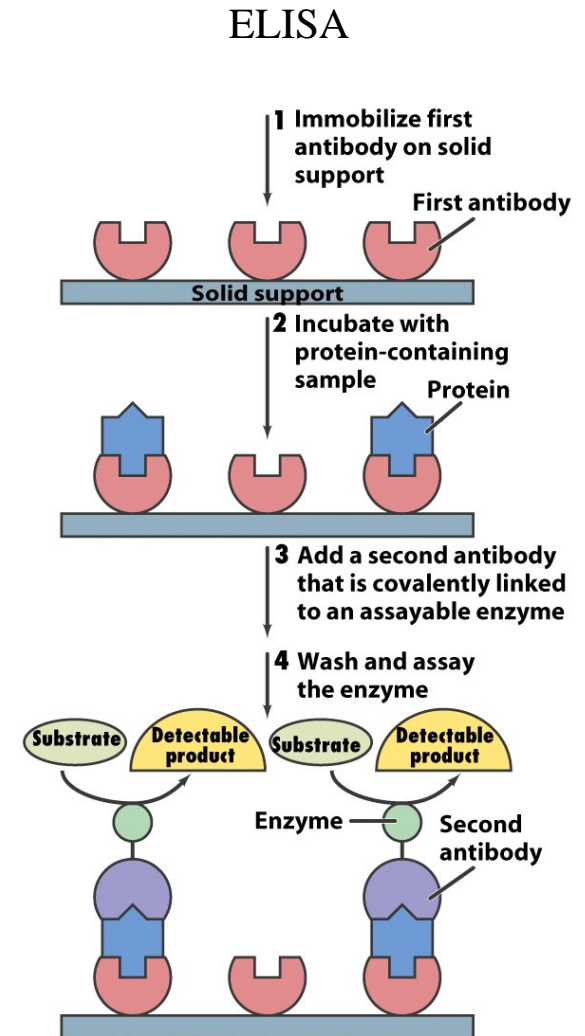


Figure 5-3 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Absorption spectrum
Chromophore

Protein quantitation
UV
Bradford assay

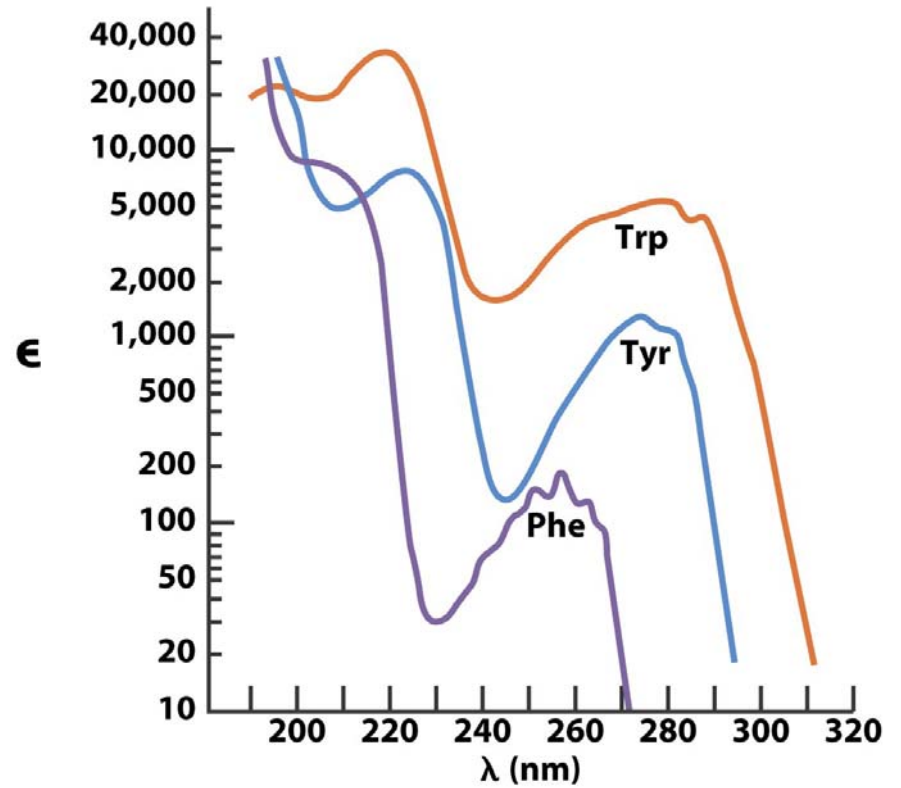


Figure 5-4 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

UV absorption spectrum

Separation techniques (fractionation procedures)

<u>Protein characteristic</u>	<u>purification procedure</u>
Solubility	salting out
Ionic charge	ion exchange chromatography electrophoresis isoelectric focusing
Polarity	hydrophobic interaction chromatography gel filtration chromatography SDS-PAGE ultracentrifugation ultrafiltration
Binding specificity	affinity chromatography

Protein solubility

Salting in

Salting out

Isoelectric point (pI)

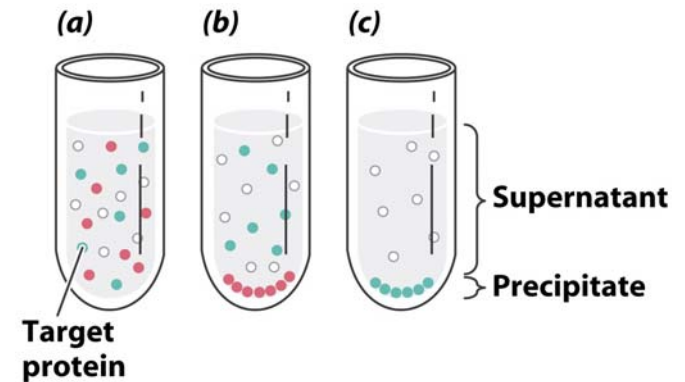
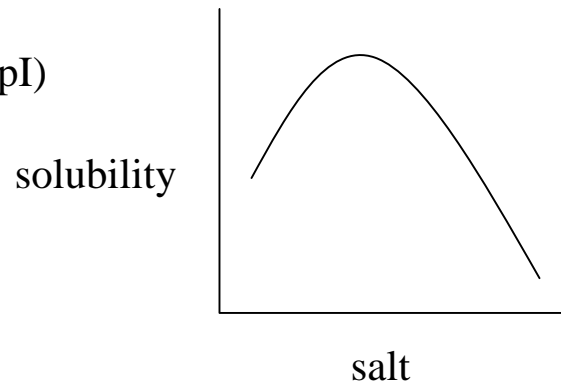


Figure 5-5 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Table 5-2 Isoelectric Points of Several Common Proteins

Protein	pI
Pepsin	<1.0
Ovalbumin (hen)	4.6
Serum albumin (human)	4.9
Tropomyosin	5.1
Insulin (bovine)	5.4
Fibrinogen (human)	5.8
γ -Globulin (human)	6.6
Collagen	6.6
Myoglobin (horse)	7.0
Hemoglobin (human)	7.1
Ribonuclease A (bovine)	9.4
Cytochrome <i>c</i> (horse)	10.6
Histone (bovine)	10.8
Lysozyme (hen)	11.0
Salmine (salmon)	12.1

Chromatography

Mobile phase & stationary phase (matrix)

Matrix type: paper, gel, HPLC

Ion exchange chromatography: electric charge

Anion exchangers: DEAE

Cation exchangers: CM

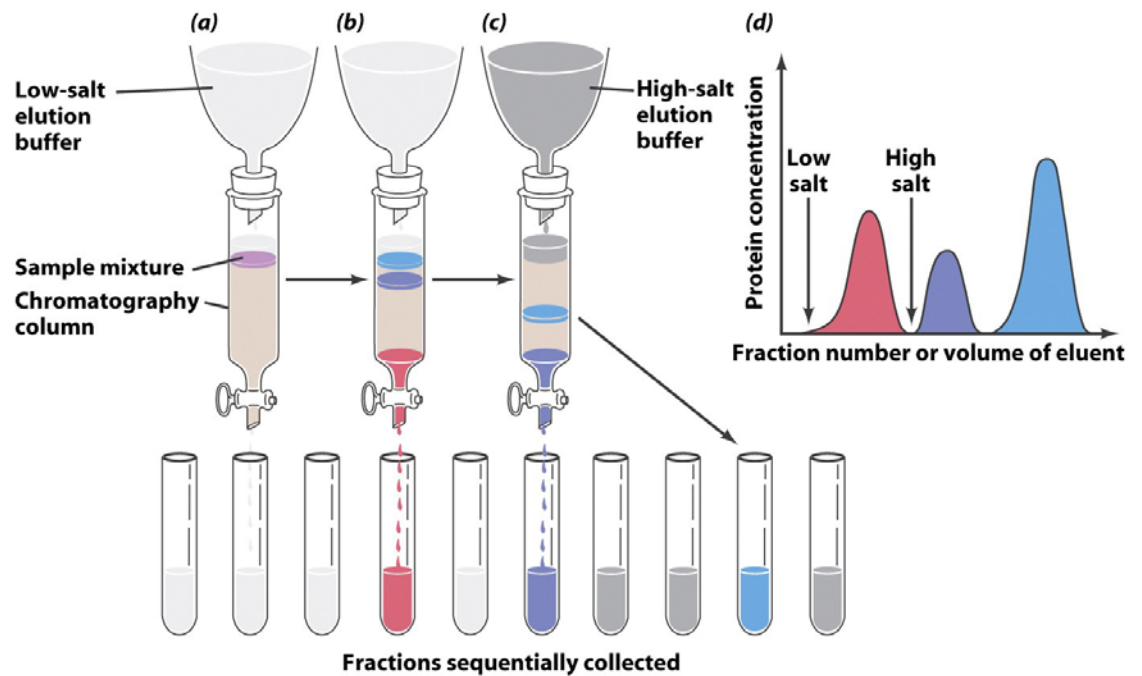


Figure 5-6 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Hydrophobic interaction chromatography: hydrophobicity
phenyl, octyl (C8), C18, etc

Gel filtration chromatography: molecular sieve: size & shape

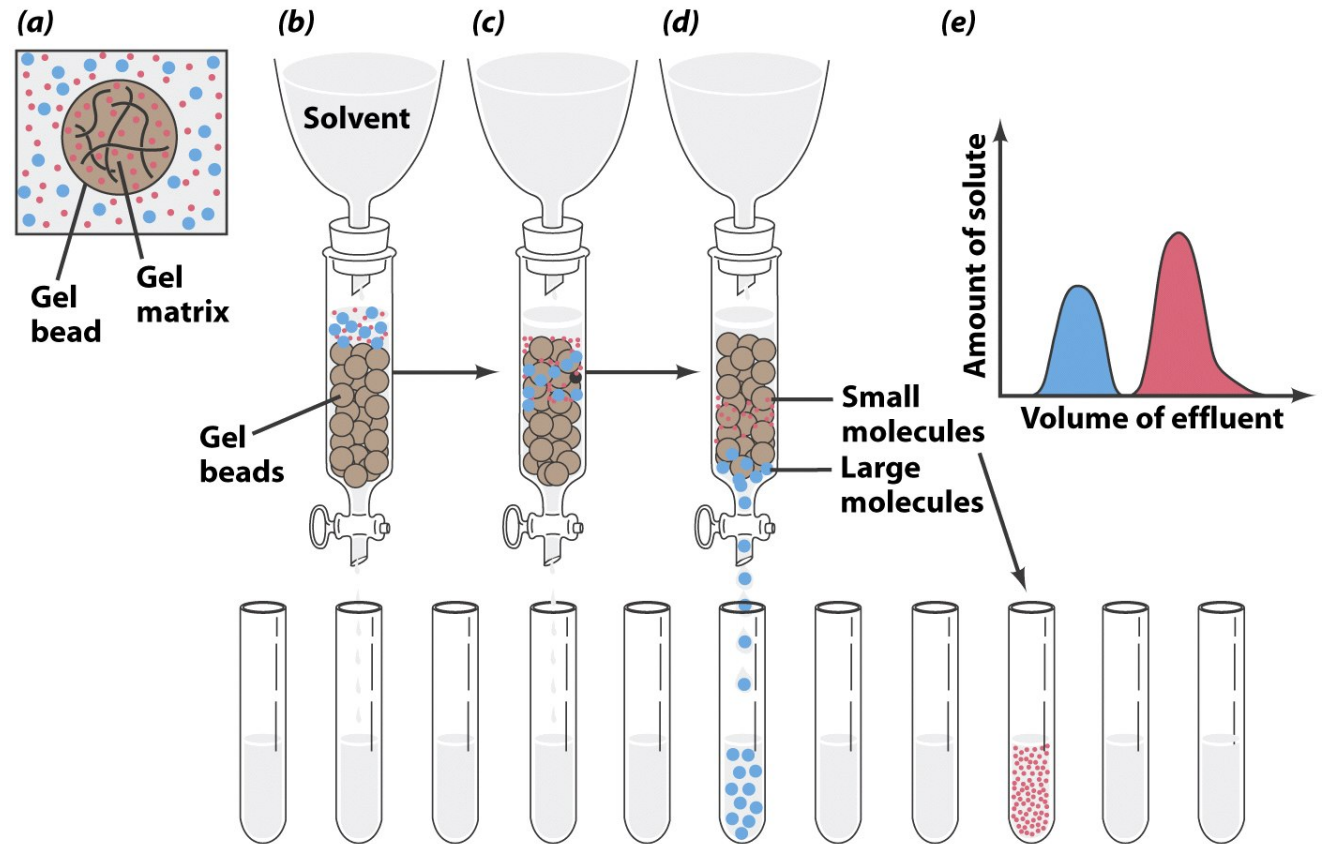


Figure 5-7 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Affinity chromatography

Spacer

Ligand: substrate, inhibitor, etc

Immunoaffinity chromatography

Metal chelate affinity chromatography

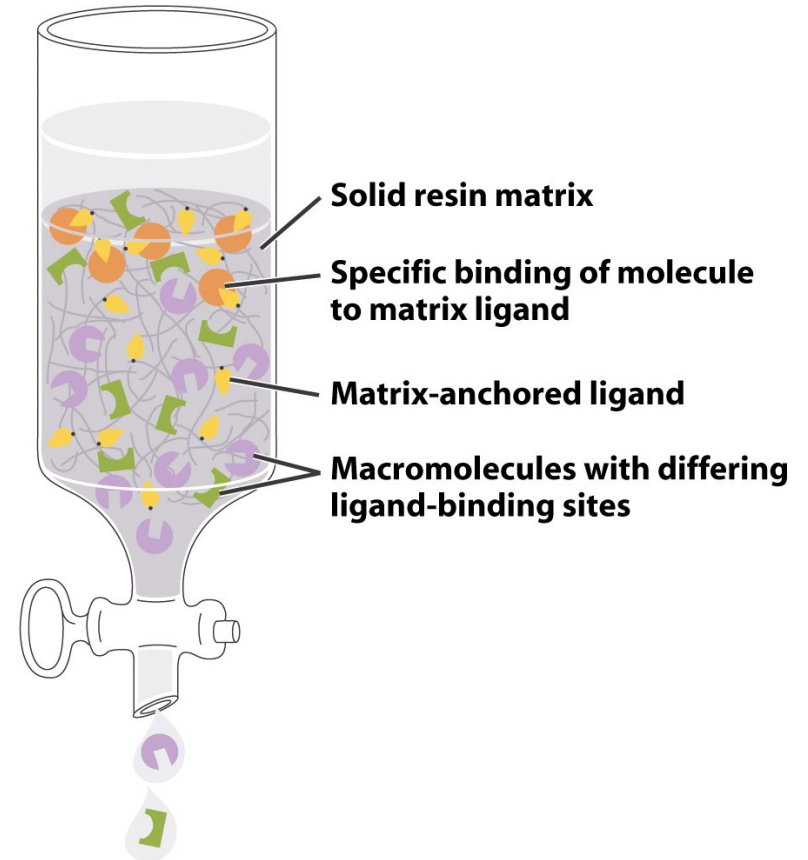


Figure 5-8 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Electrophoresis

Depends on size, shape, electric charge

Polyacrylamide

Agarose

paper

SDS-PAGE

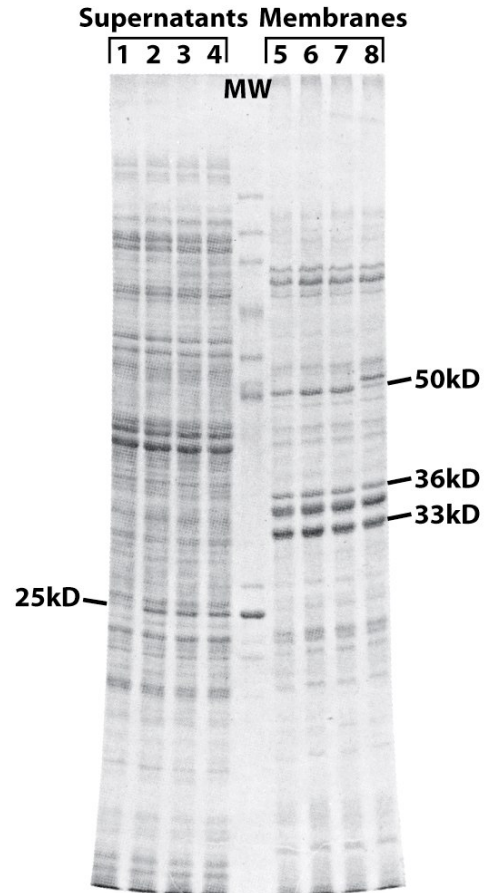


Figure 5-9 Fundamentals of Biochemistry, 2/e

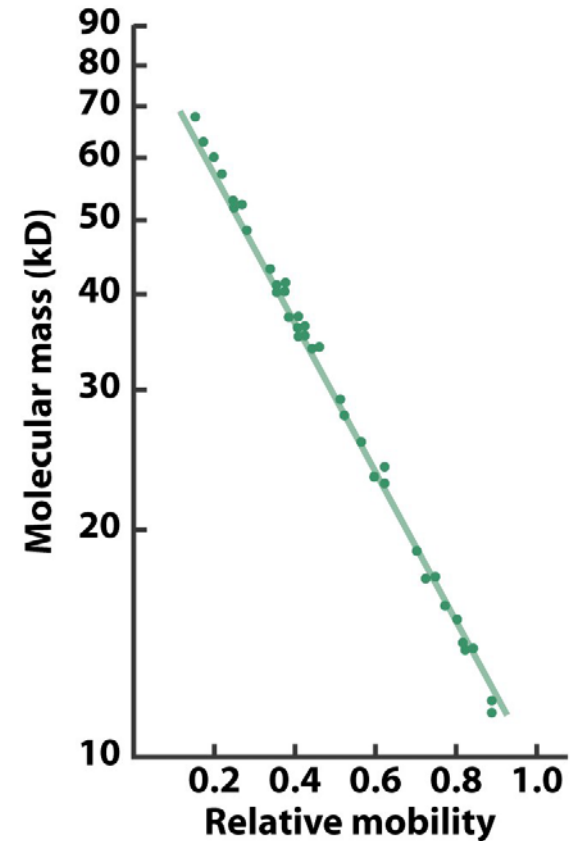


Figure 5-10 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

IEF & 2D-PAGE

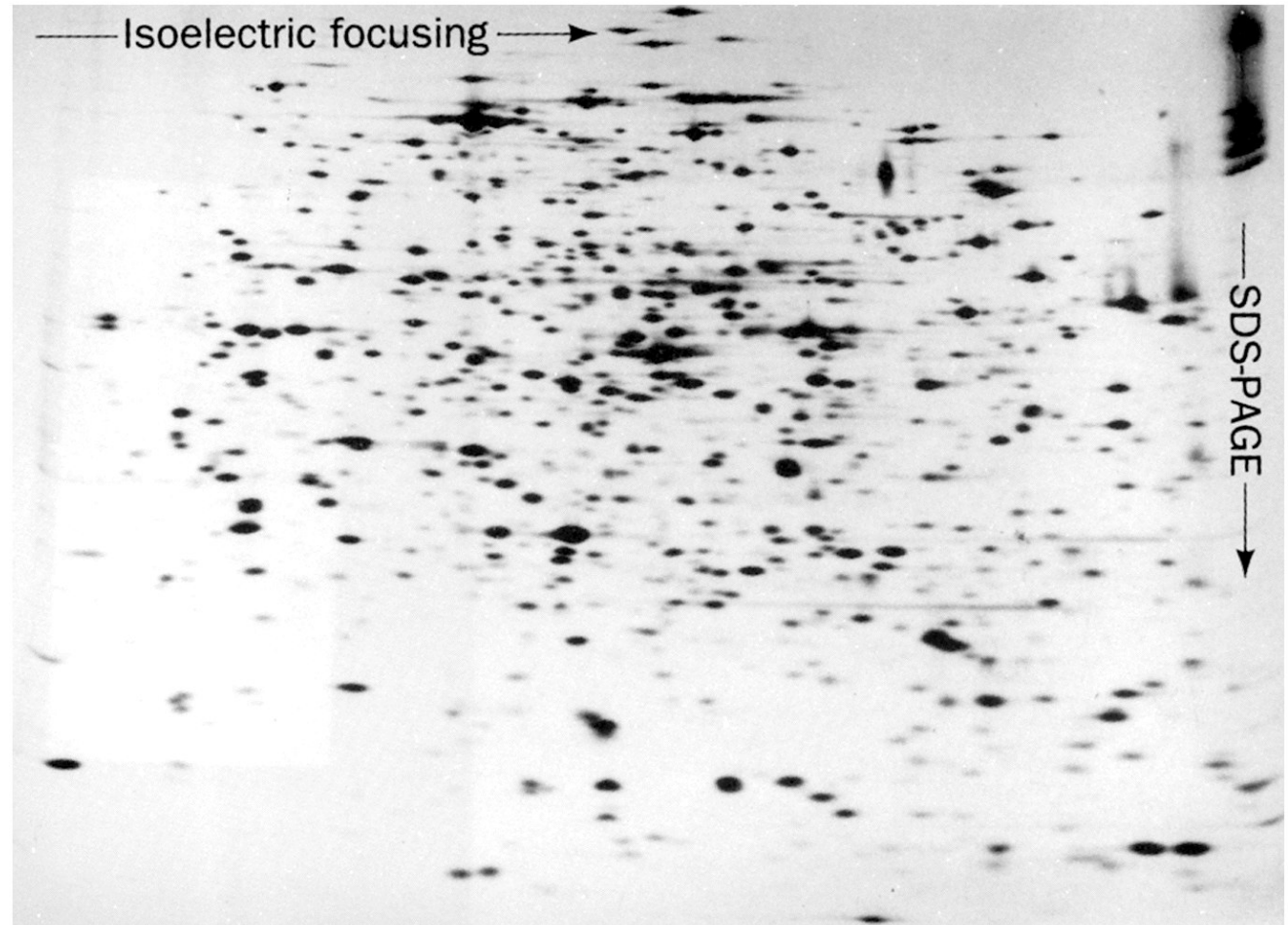


Figure 5-11 Fundamentals of Biochemistry, 2/e

Ultracentrifugation

Sedimentation rate: depends on mass, shape, density of the solution

Svedbergs (S) units: (10^{-13} s)

40S + 60S = 80S (not 100S)

Zonal ultracentrifugation: premade density gradient

Equilibrium density gradient centrifugation: CsCl

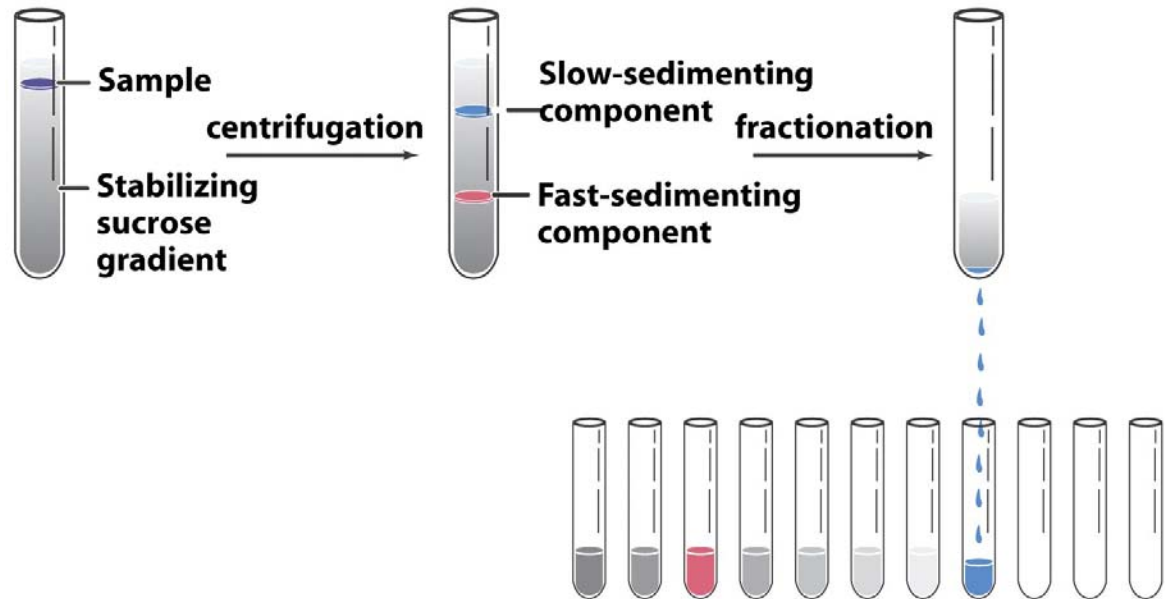


Figure 5-12 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Protein sequencing

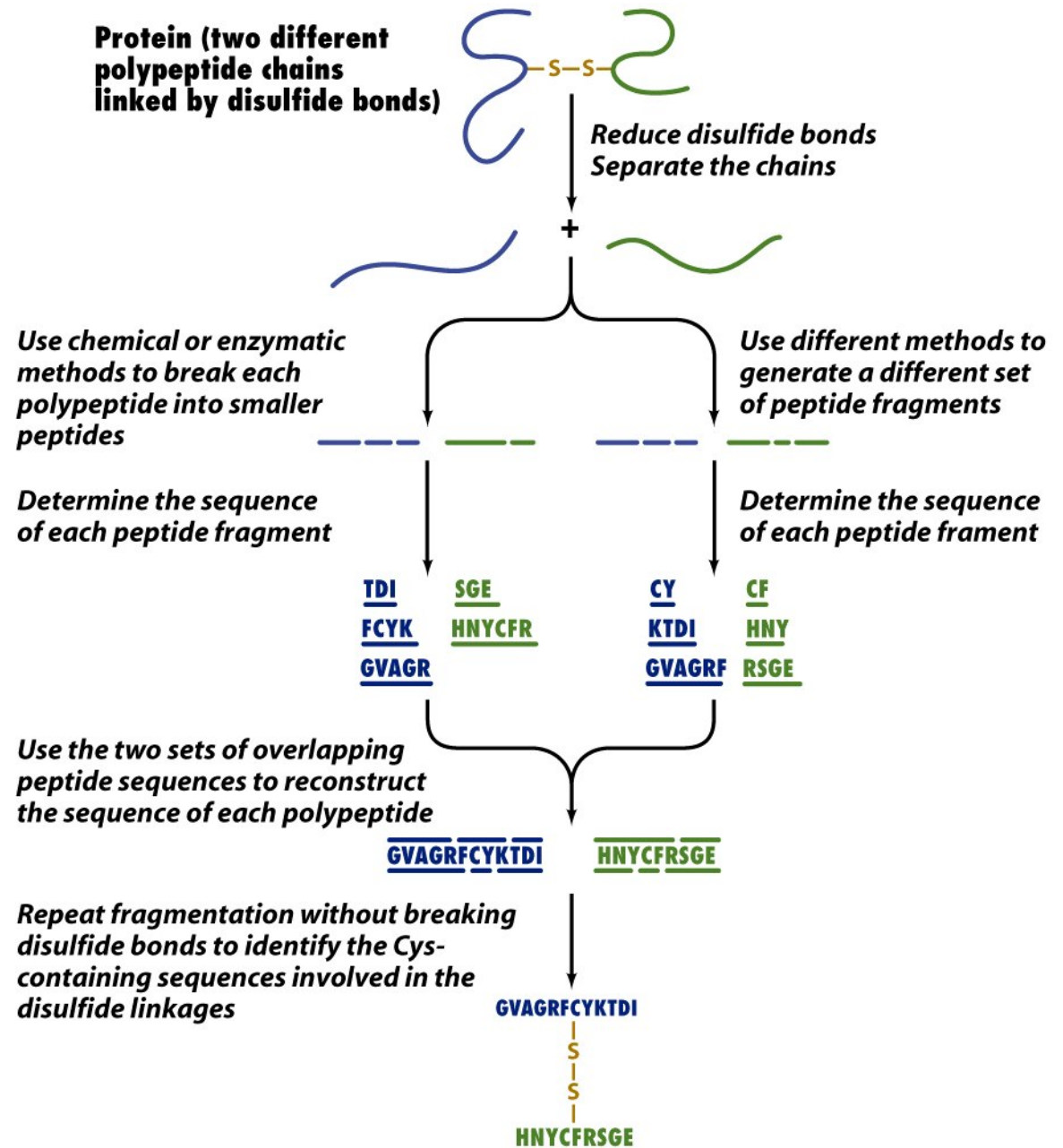


Figure 5-13 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

N-terminal sequencing

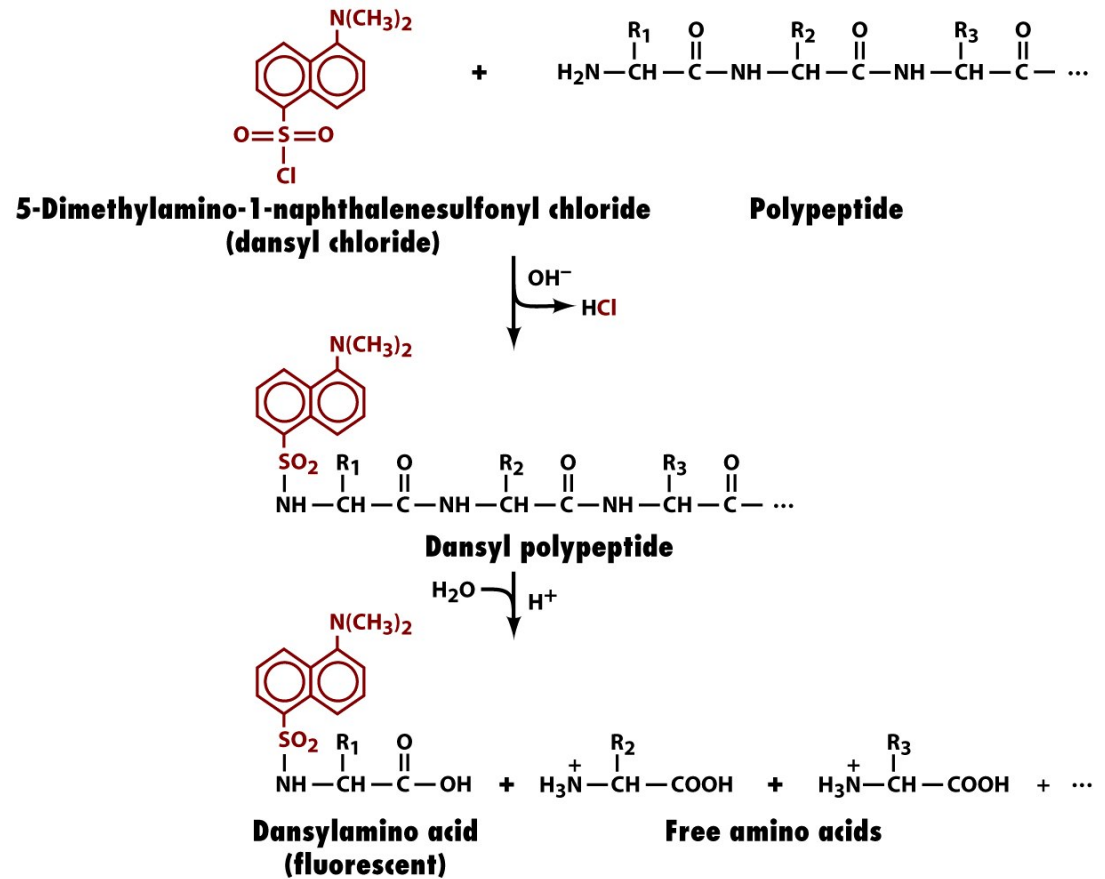
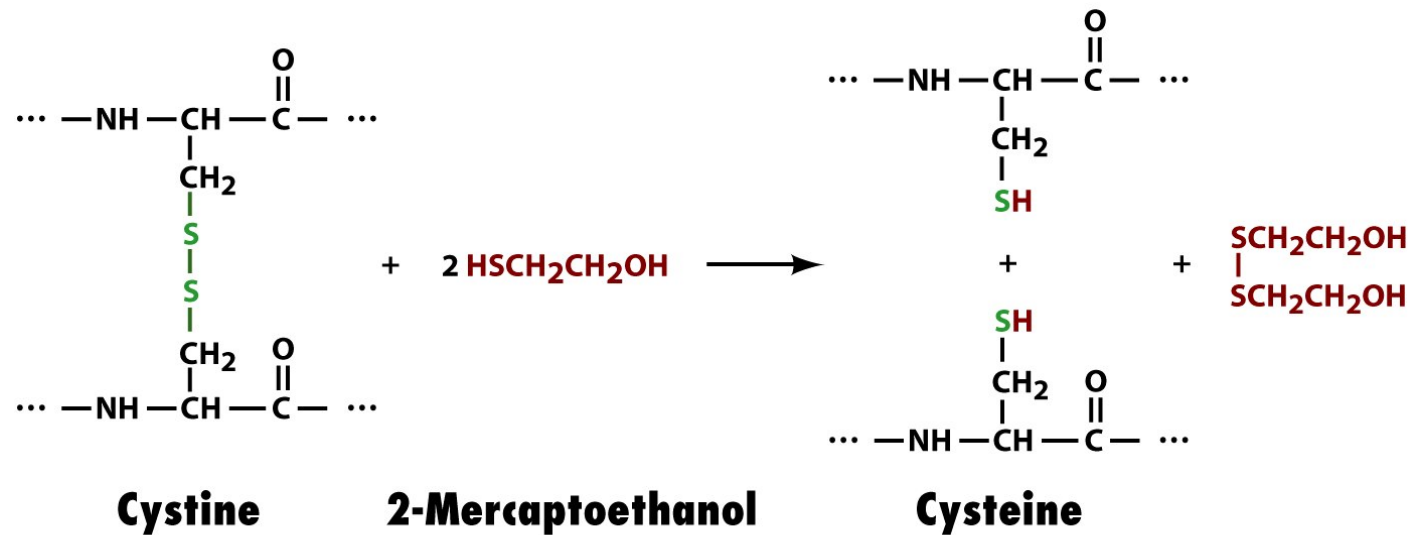
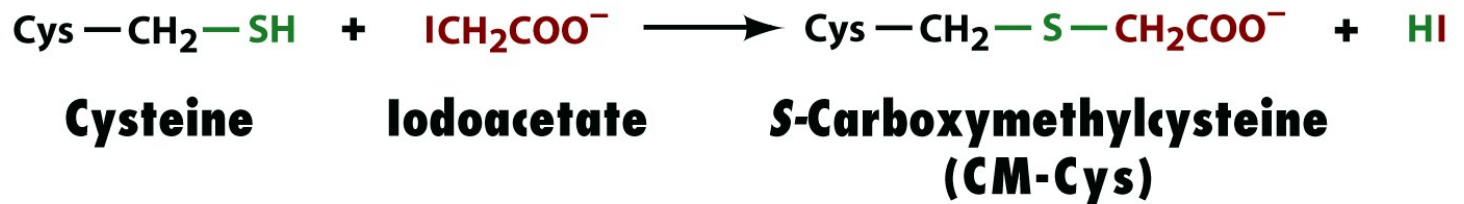


Figure 5-14 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Disulfide bond cleavage



Unnumbered figure pg 112a Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons



Unnumbered figure pg 112b Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Amino acid composition

Hydrolysis
Derivatization
HPLC

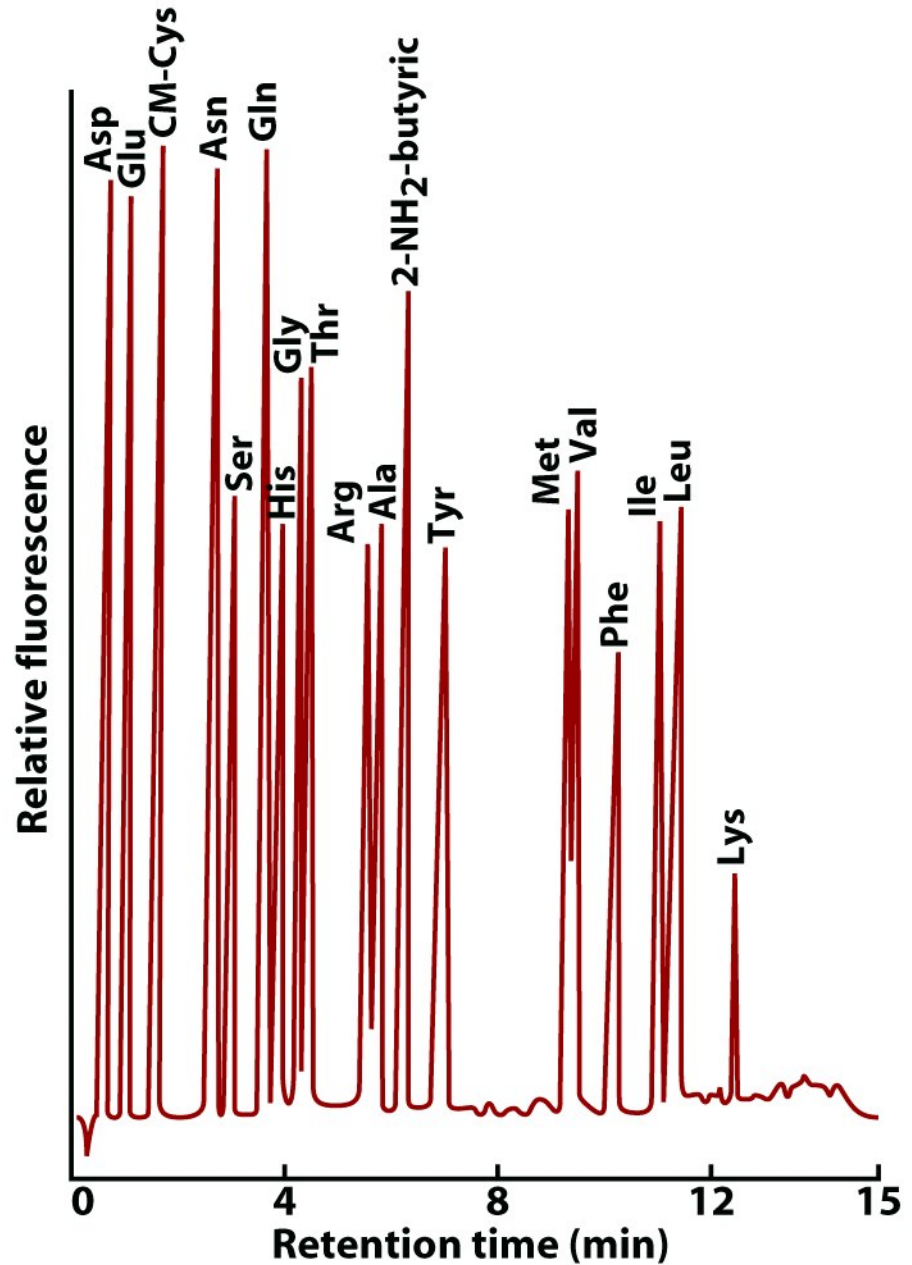


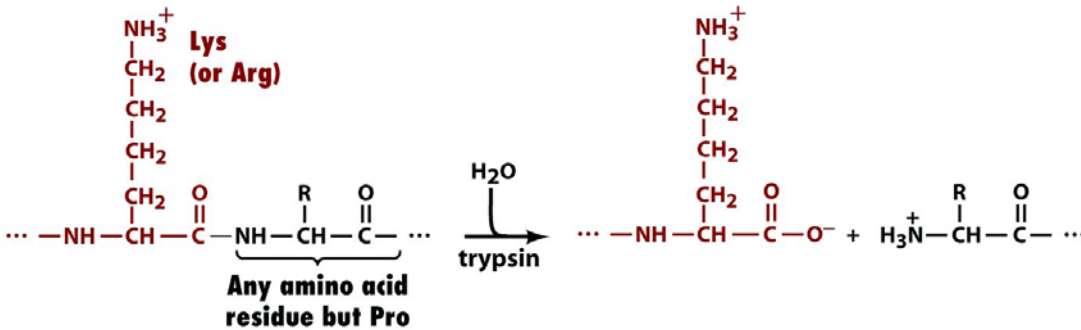
Figure 5-15 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Table 5-3 Specificities of Various Endopeptidases

Enzyme	Source	Specificity	Comments
Trypsin	Bovine pancreas	R_{n-1} = positively charged residues: Arg, Lys; $R_n \neq$ Pro	Highly specific
Chymotrypsin	Bovine pancreas	R_{n-1} = bulky hydrophobic residues: Phe, Trp, Tyr; $R_n \neq$ Pro	Cleaves more slowly for R_{n-1} = Asn, His, Met, Leu
Elastase	Bovine pancreas	R_{n-1} = small neutral residues: Ala, Gly, Ser, Val; $R_n \neq$ Pro	
Thermolysin	<i>Bacillus thermoproteolyticus</i>	R_n = Ile, Met, Phe, Trp, Tyr, Val; $R_{n-1} \neq$ Pro	Occasionally cleaves at R_n = Ala, Asp, His, Thr; heat stable
Pepsin	Bovine gastric mucosa	R_n = Leu, Phe, Trp, Tyr; $R_{n-1} \neq$ Pro	Also others; quite nonspecific; pH optimum = 2
Endopeptidase V8	<i>Staphylococcus aureus</i>	R_{n-1} = Glu	

Table 5-3 Fundamentals of Biochemistry, 2/e

© 2006 John Wiley & Sons



Unnumbered figure pg 113 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Polypeptide cleavage

Enzymatic

Chemical

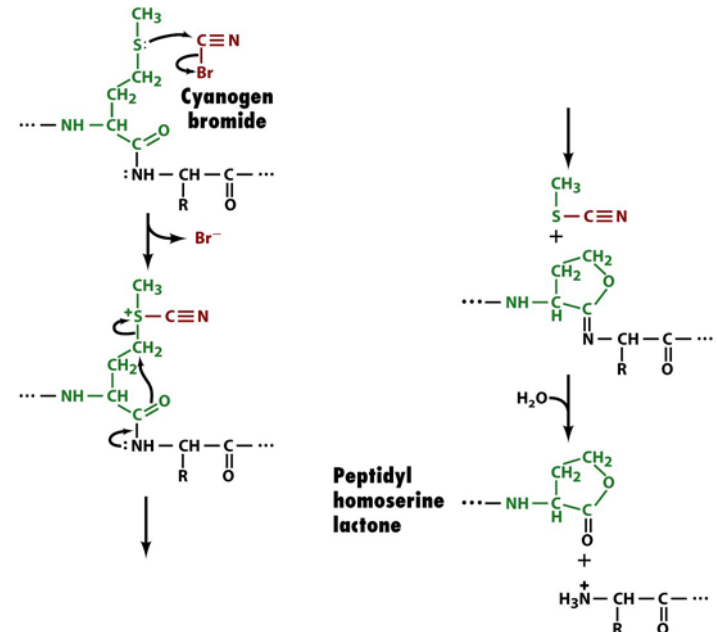


Figure 5-16 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Edman degradation

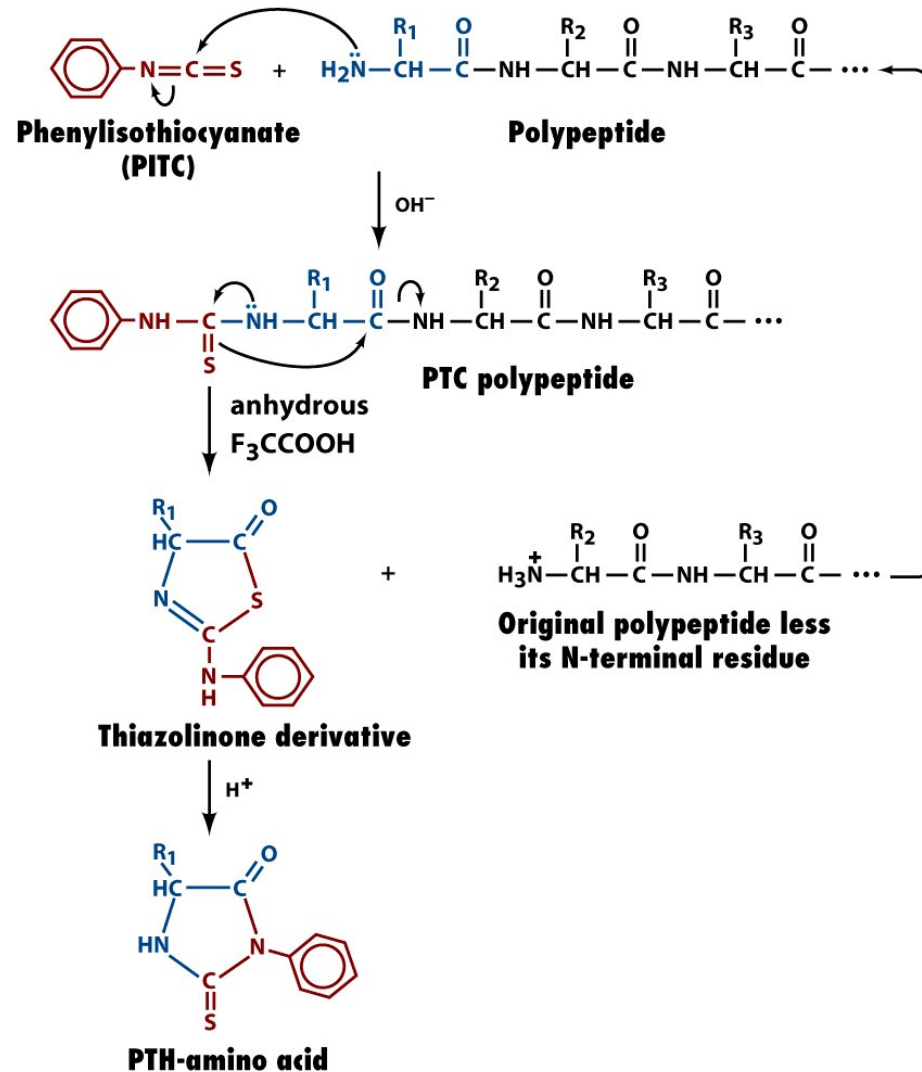


Figure 5-17 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Sequencing by mass spectrometry

Mass-to-charge ratio (m/z) for ions in the gas phase

Electron spray ionization (ESI)

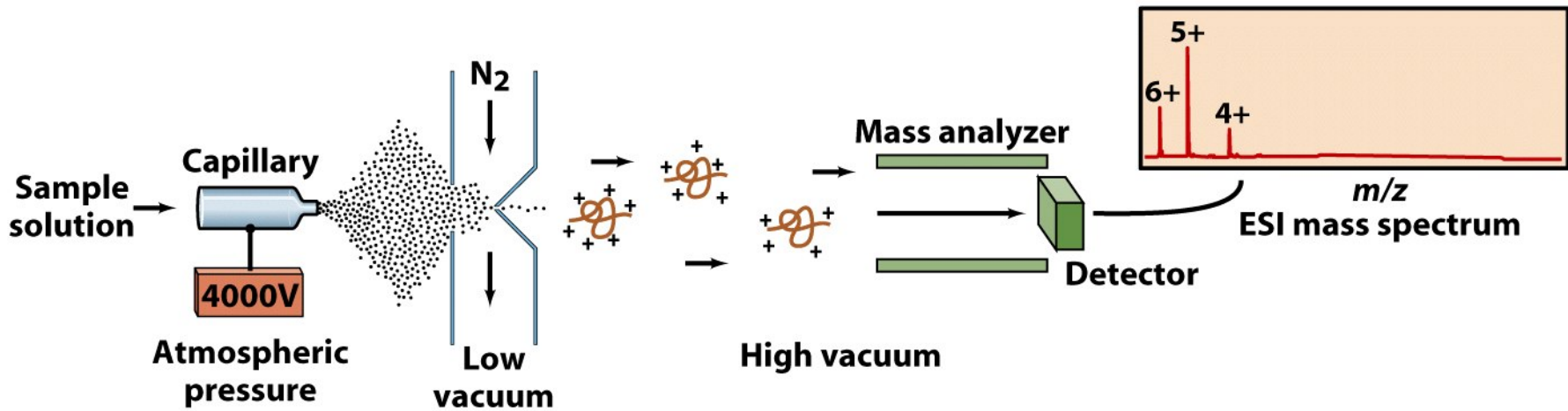


Figure 5-18a Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

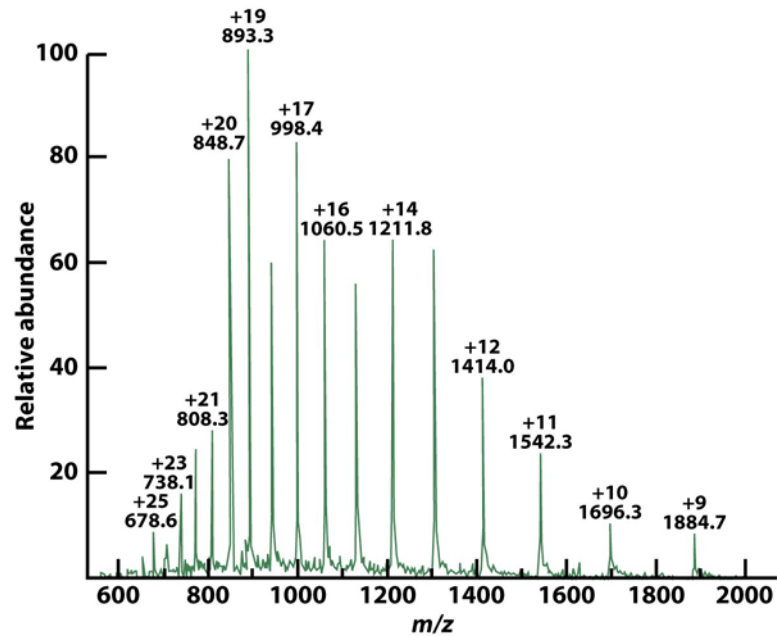


Figure 5-18b Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

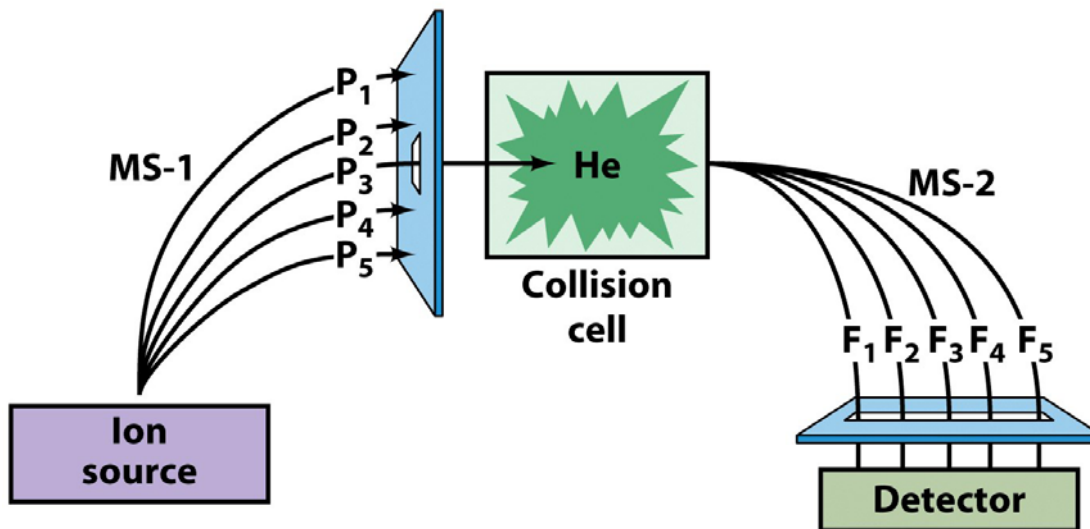


Figure 5-19 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

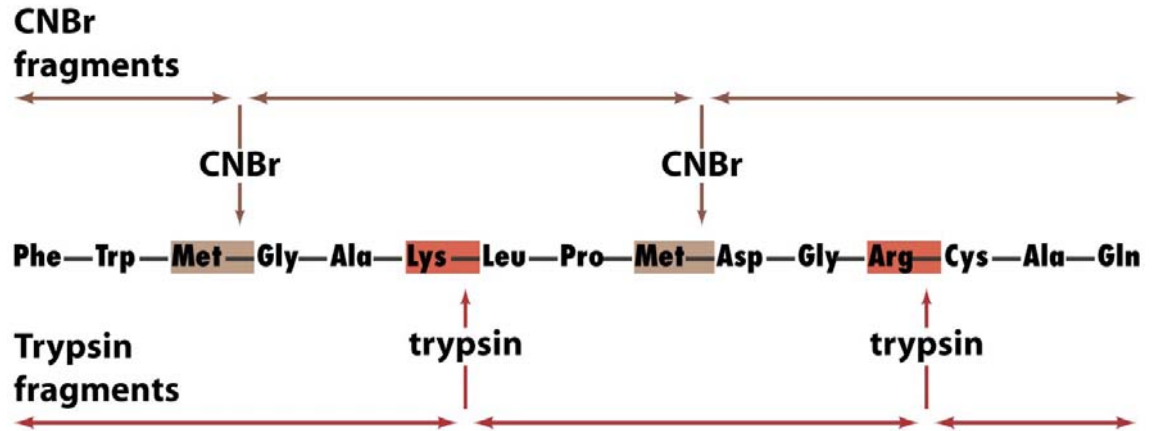


Figure 5-20 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Final step in protein sequencing

Generating overlapping fragments
Determining the positions of disulfide bonds

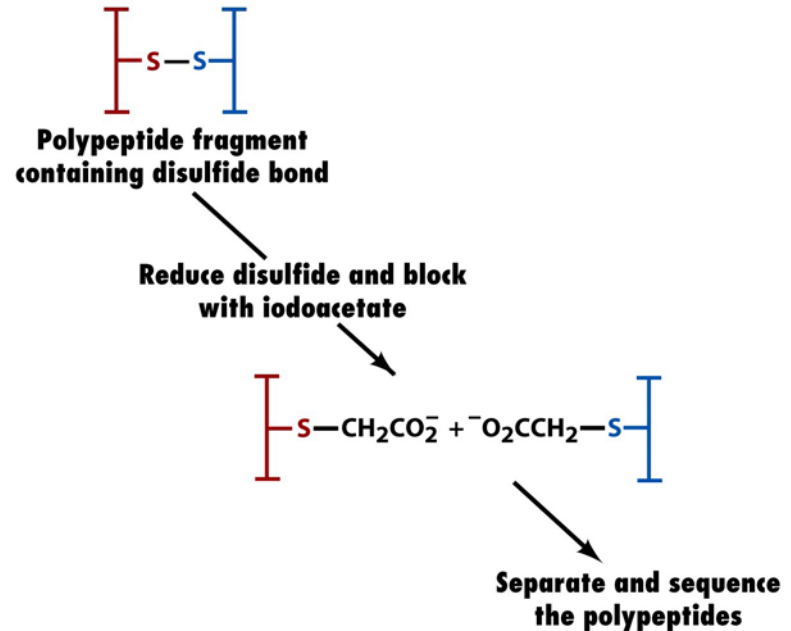


Figure 5-21 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Homology search

Table 5-4 Internet Addresses for the Major Protein and DNA Sequence Data Banks

Data Banks Containing Protein Sequences

ExPASy Molecular Biology Server (Swiss-Prot): <http://au.expasy.org>

Protein Information Resource (PIR): <http://pir.georgetown.edu/>

Protein Research Foundation (PRF): <http://www4.prf.or.jp/>

UniProt: <http://www.ebi.uniprot.org/>

Data Banks Containing Gene Sequences

GenBank: <http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html>

European Bioinformatics Institute (EBI): <http://srs.ebi.ac.uk/>

DBGET/Integrated Database Retrieval System: <http://www.genome.ad.jp/dbget>

General information about the UniProt/Swiss-Prot entry		
Entry name	RSN_HUMAN	
Primary accession number	Q9HD89	
Entered in Swiss-Prot	Release 40, 16-OCT-2001	
Sequence was last modified	Release 40, 16-OCT-2001	
Annotations were last modified	Release 44, 05-JUL-2004	
Protein description		
Protein name	Resistin precursor	
Synonyms	Cysteine-rich secreted protein FIZZ3 Adipose tissue-specific secretory factor ADSF C/EBP-epsilon regulated myeloid-specific secreted cysteine-rich protein Cysteine-rich secreted protein A12-alpha-like 2 UNQ407/PRO1199	
Origin of the protein		
Gene	Gene name	RETN
	Synonyms	RSTN, FIZZ3, HXCP1
From	Homo sapiens (Human)[TaxID:9606]	
Taxonomy	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

Figure 5-22 Fundamentals of Biochemistry, 2/e

Protein evolution

Protein sequence evolution

Sequence comparison

Invariant residue

Conservatively substituted

Hypervariable

Neutral drift

TABLE 5-5 Amino Acid Sequences of Cytochrome c from 38 species^a

	-9	-5	-1	1	5	10	15	20	25	30	35	40
Mammals												
Human, chimpanzee												
Rhesus monkey												
Horse												
Donkey												
Cow, pig, sheep												
Dog												
Rabbit												
California gray whale												
Great gray kangaroo												
Other vertebrates												
Chicken, turkey												
Pigeon												
Pekin duck												
Snapping turtle												
Rattlesnake												
Bullfrog												
Tuna												
Dogfish												
Insects												
<i>Samia cynthia</i> (a moth)												
Tobacco hornworm moth												
Screwworm fly												
<i>Drosophila</i> (fruit fly)												
Fungi												
Baker's yeast												
<i>Candida krusei</i> (a yeast)												
<i>Neurospora crassa</i> (a mold)												
Higher plants												
Wheat germ												
Buckwheat seed												
Sunflower seed												
Mung bean												
Cauliflower												
Pumpkin												
Sesame seed												
Castor bean												
Cottonseed												
<i>Abutilon</i> seed												
Number of different amino acids	1	3	5	5	1	3	3	4	1	3	2	1

^aThe amino acid side chains have been shaded according to their polarity characteristics so that an invariant or conservatively substituted residue is identified by a vertical band of a single color. The letter a at the beginning of the chain indicates that the N-terminal amino group is acetylated; an h indicates that the acetyl group is absent.

Source: After Dickerson, R.E., *Sci. Am.* 226(4): 58-72 (1972), with corrections from Dickerson, R.E., and Timkovich, R., in Boyer, P.D. (Ed.), *The Enzymes* (3rd ed.), Vol. 11, pp. 421-422, Academic Press (1975). Table copyrighted © by Irving Geis.

Table 5-5 part 1 Fundamentals of Biochemistry, 2/e

	45	50	55	60	65	70	75	80	85	90	95	100	104
PGY SYT AANKNKG I WGEDT LMEY LEPKKY IPG TKMIFVGI KKKKEERADLIAYLKKATNE													
PGY SYT AANKNKG I WGEDT LMEY LEPKKY IPG TKMIFVGI KKKKEERADLIAYLKKAAE													
PGF TYT DANKNKG I TWKEET LMEY LEPKKY IPG TKMIFAGI KKKTERADLIAYLKKATNE													
PGF SYT DANKNKG I TWGEET LMEY LEPKKY IPG TKMIFAGI KKKGERADLIAYLKKATNE													
PGF SYT DANKNKG I TWGEET LMEY LEPKKY IPG TKMIFAGI KKKGERADLIAYLKKATNE													
PGF SYT DANKNKG I TWGEET LMEY LEPKKY IPG TKMIFAGI KKKGERADLIAYLKKATNE													
VGF SYT DANKNKG I TWGEDT LMEY LEPKKY IPG TKMIFAGI KKKDERADLIAYLKKATNE													
VGF SYT DANKNKG I TWGEET LMEY LEPKKY IPG TKMIFAGI KKKGERADLIAYLKKATNE													
PGF TYT DANKNKG I WGEDT LMEY LEPKKY IPG TKMIFAGI KKKGERADLIAYLKKATNE													
EGF SYT DANKNKG I TWGEDT LMEY LEPKKY IPG TKMIFAGI KKKSERVDLIAYLKDATS K													
EGF SYT DANKNKG I TWGEDT LMEY LEPKKY IPG TKMIFAGI KKKKAERADLIAYLKKQATA K													
EGF SYT DANKNKG I TWGEDT LMEY LEPKKY IPG TKMIFAGI KKKSERADLIAYLKKDATA K													
EGF SYT EANKNKG I TWGEET LMEY LEPKKY IPG TKMIFAGI KKKKAERADLIAYLKKDATA K													
VGF SYT AANKNKG I WGEDT LMEY LEPKKY IPG TKMVFGLSKKKERNLIAYLKKKATAA													
AGF SYT DANKNKG I TWGEDT LMEY LEPKKY IPG TKMIFAGI KKKGERQDLIAYLKSACSK													
EGF SYT DANKSKGI VWNNDT LMEY LEPKKY IPG TKMIFAGI KKKGERDLVLIYLSATS S													
QGF SYT DANKSKGI WQDET LRIY LEPKKY IPG TKMIFAGI KKKSERQDLIAYLKKTAAS													
PGF SYSNANKAKGI TWGDDT LFEY LEPKKY IPG TKMVFAGLKKANERADLIAYLKESTK													
PGF SYSNANKAKGI TWQDDT LFEY LEPKKY IPG TKMVFAGLKKANERADLIAYLKKQATK													
AGFAYTNANKAKGI TWQDDT LFEY LEPKKY IPG TKMIFAGLKKPNERADLIAYLKSATK													
AGFAYTNANKAKGI TWQDDT LFEY LEPKKY IPG TKMIFAGLKKPNERADLIAYLKSATK													
QGY SYT DANIKNVLDWENNMEYLTPPKY IPG TKMAFGGLKKEKDRNDLIY LKKACE													
DGY SYT DANKRAGVEWAPTMSDY LEPKKY IPG TKMAFGGLKKAERADLIY LKKLEASK													
DGY SYT DANKKQGI TWGENT LFEY LEPKKY IPG TKMVFAGLKKKDRNDLIY LKKKATAA													
AGY SYT AANKNKA VEWEENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATS S													
AGY SYT AANKNKA VTWGEDT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKDSTE													
AGY SYT AANKNKA VIWEENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKSTAS													
AGY SYT AANKNKA VIWEENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKASTA													
AGY SYT AANKNKA VEWEENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATA													
PGY SYT AANKNKA VIWEENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATA													
PGY SYT AANKNKA VIWEENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATA													
AGY SYT AANKNKA VQWENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATA													
AGY SYT AANKNKA VQWENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATA													
PGY SYT AANKNKA VQWENT LYDY LEPKKY IPG TKMVFAGLXKPPQDRADLIAYLKKATA													

Hydrophilic, acidic: **D** Asp **E** Glu

Hydrophilic, basic: **H** His **K** Lys **R** Arg **X** TrimethylLys

Polar, uncharged: **B** Asn or Asp **G** Gly **N** Asn **Q** Gln
S Ser **T** Thr **W** Trp **Y** Tyr **Z** Gln or Glu

Hydrophobic: **A** Ala **C** Cys **F** Phe **I** Ile **L** Leu
M Met **P** Pro **V** Val

Table 5-5 part 2 Fundamentals of Biochemistry, 2/e

Phylogenetic trees

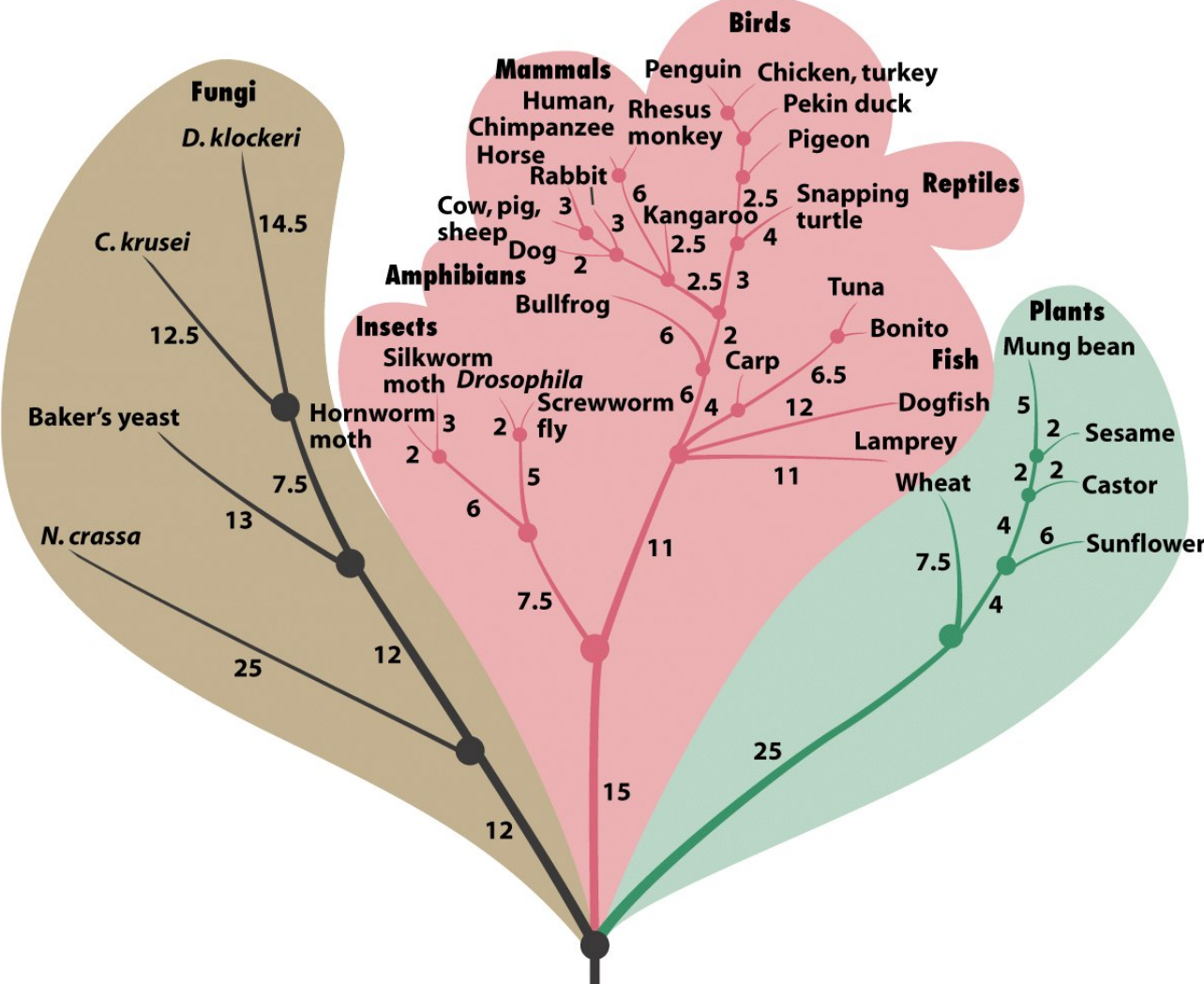


Figure 5-23 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Protein evolve at characteristic rates

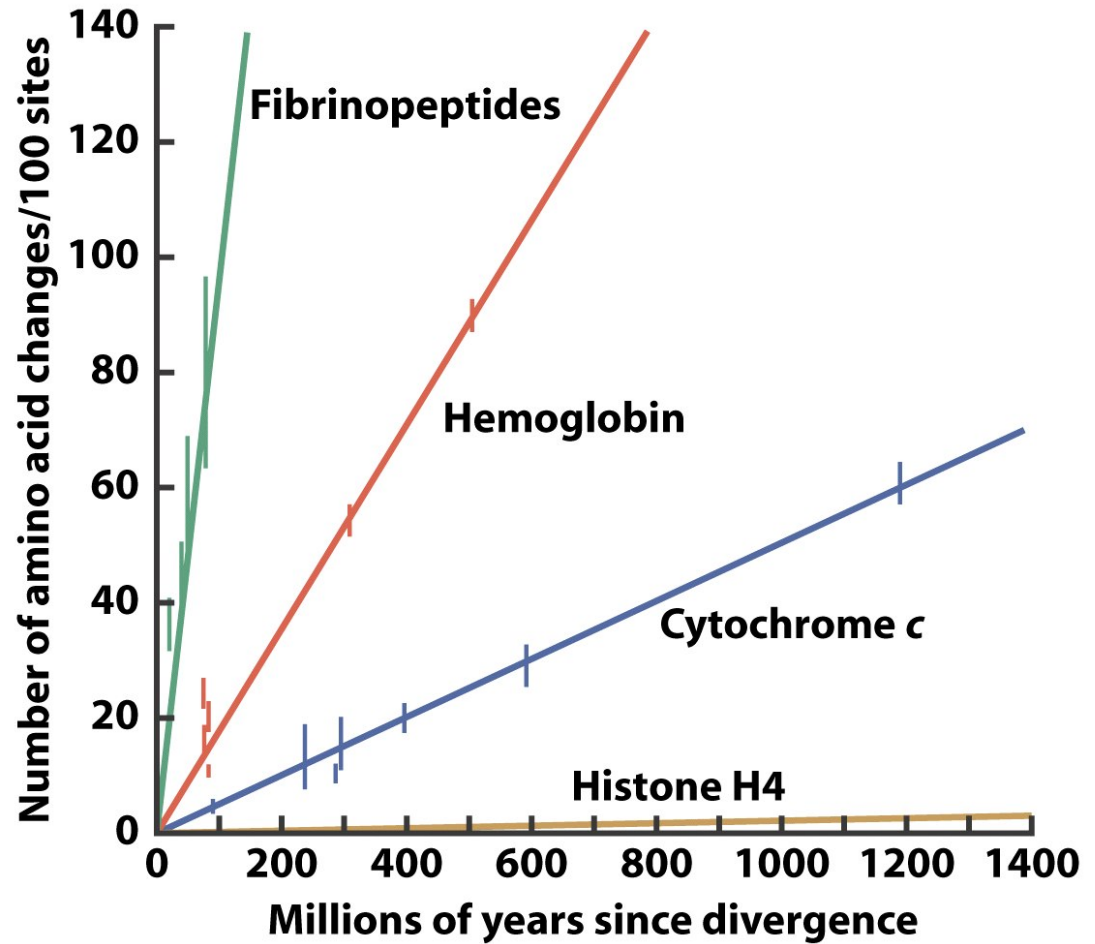


Figure 5-24 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Gene duplication and protein families

Homologous
Orthologous
Paralogous
Pseudogenes

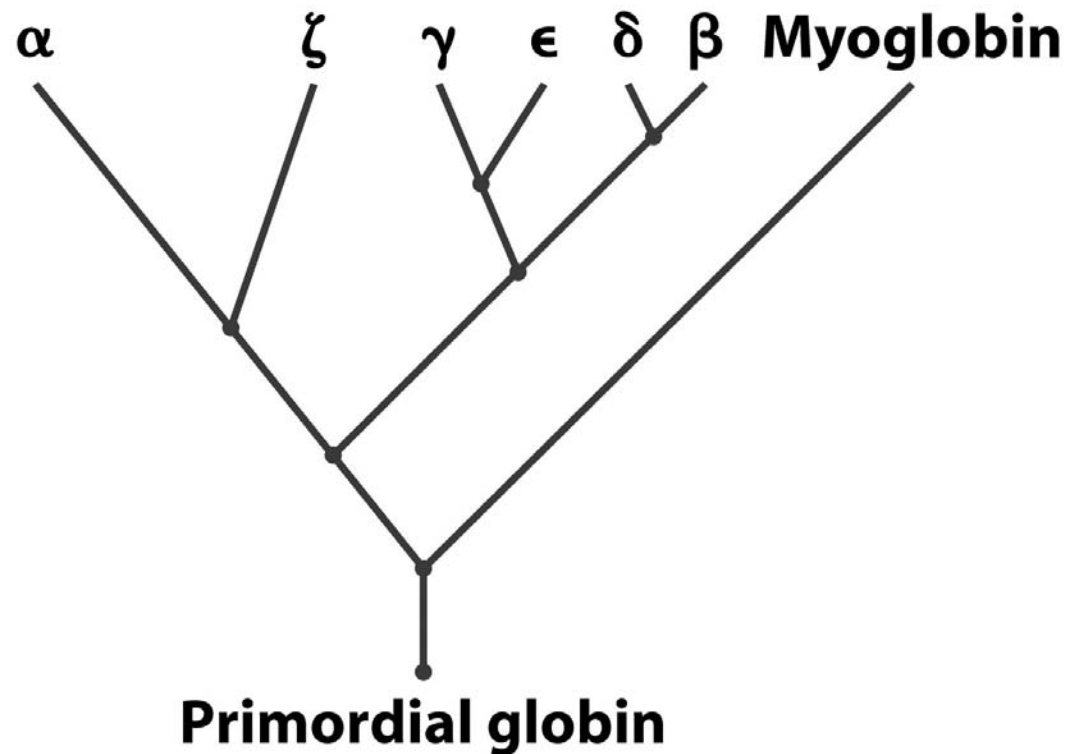


Figure 5-25 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Protein modules

Module

Understanding of modular structure

(a) Fibronectin



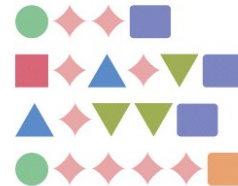
(b) Blood clotting proteins

Factors VII, IX, X, and protein C

Factor XII

Tissue-type plasminogen activator

Protein S



Key

▲ Fibronectin domain 1

■ Fibronectin domain 2

● Fibronectin domain 3

● γ -Carboxyglutamate domain

◆ Epidermal growth factor domain

■ Serine protease domain

▼ Kringle domain

■ Unique domain