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

Fundamentals of Biochemistry

Second Edition

Chapter 5:

Proteins: Primary Structure

The great variation in structure and function among proteins

Chapter 5 Opener Fundamentals of Biochemistry, 2/e

Protein structure

What is primary structure?

The amino acid sequence of its polypeptide chain

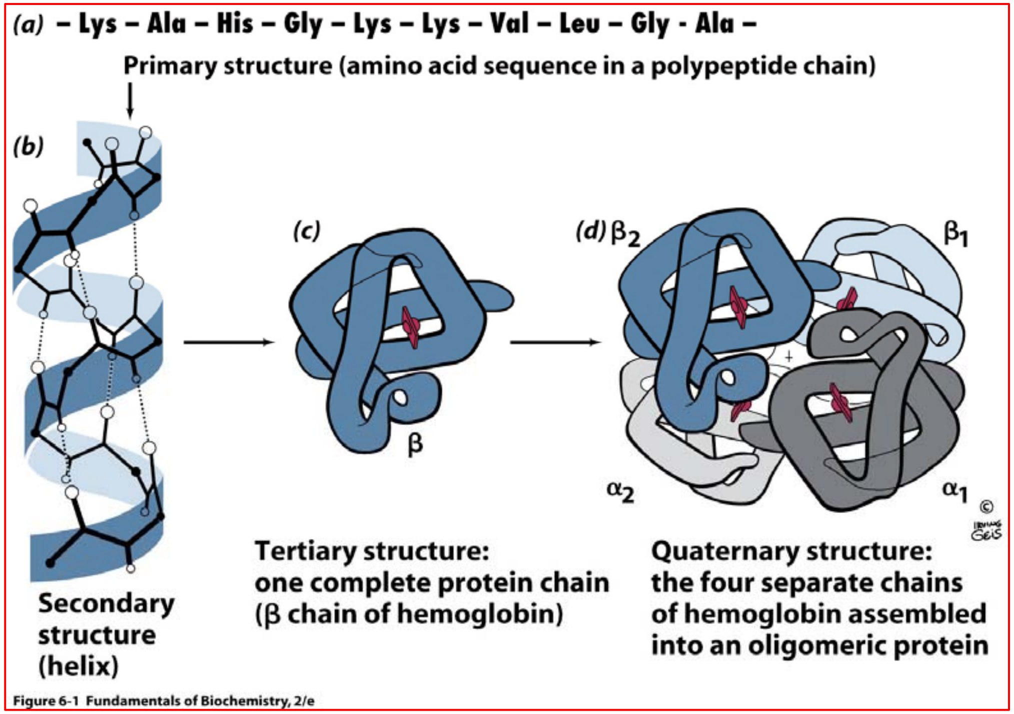


Figure 6-1 Fundamentals of Biochemistry, 2/e

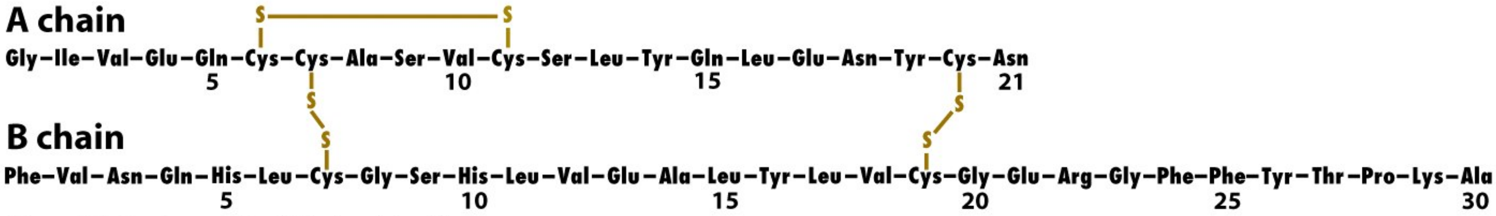


Figure 5-1 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Polypeptide diversity: theoretical possibilities are unlimited

Actual polypeptides are somewhat limited in size and composition

- Size: at least 40 residues
- The vast majority are between 100 and 1000 residues
- Monomeric & multimeric (multisubunits)
- Amino acid composition: average occurrence in proteins (Table 4-1)

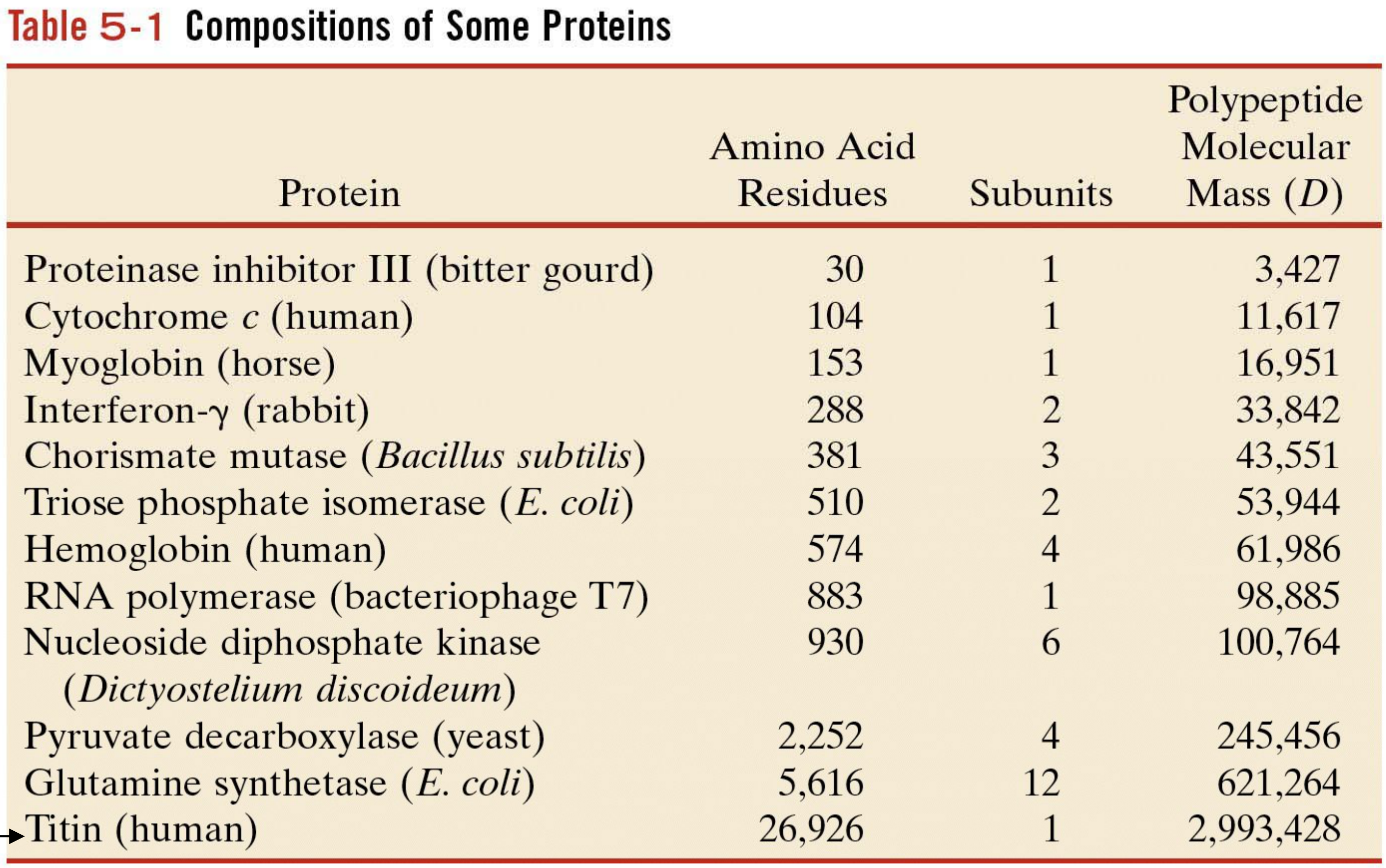


Table 5-1 Compositions of Some Proteins

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

The largest known → Titin (human)

Table 5-1 Fundamentals of Biochemistry, 2/e

Protein purification and analysis

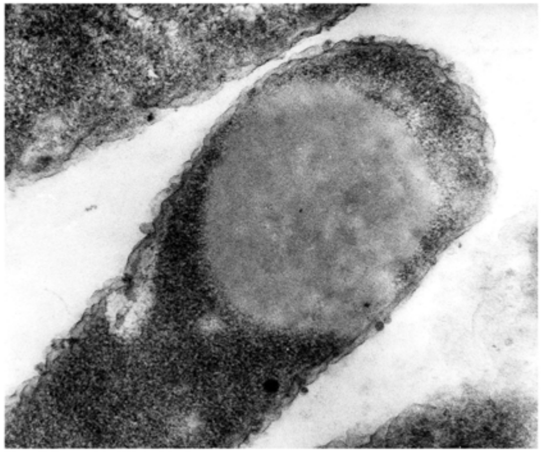


Figure 5-2 Fundamentals of Biochemistry, 2/e

Purification is a mandatory step for studying macromolecules
Starting from <0.1% to ~98% purity

A. General approach to purifying proteins
Genetic engineering: recombinant proteins
Protein source: organism, organs, cellular & subcellular locations

Protein stabilization
- pH
- Temperature
- Degradative proteins
- Adsorption to surfaces: air-water interface, glass or plastic surfaces
- Long-term storage

http://www.biochem.arizona.edu/classes/bioc462/462a/NOTES/Protein_Properties/protein_purification.htm
http://www.bio.ic.ac.uk/research/nield/expertise/protein_purification.html

Protein assay

Assay: quantitative detection

Specific, sensitive & convenient

Enzymatic reaction: substrate & product

Physiological and artificial substrates

Coupled enzymatic reaction

Immunoassays: antibody

Radioimmunoassay (RIA)

Enzyme-linked immunosorbent assay (ELISA)

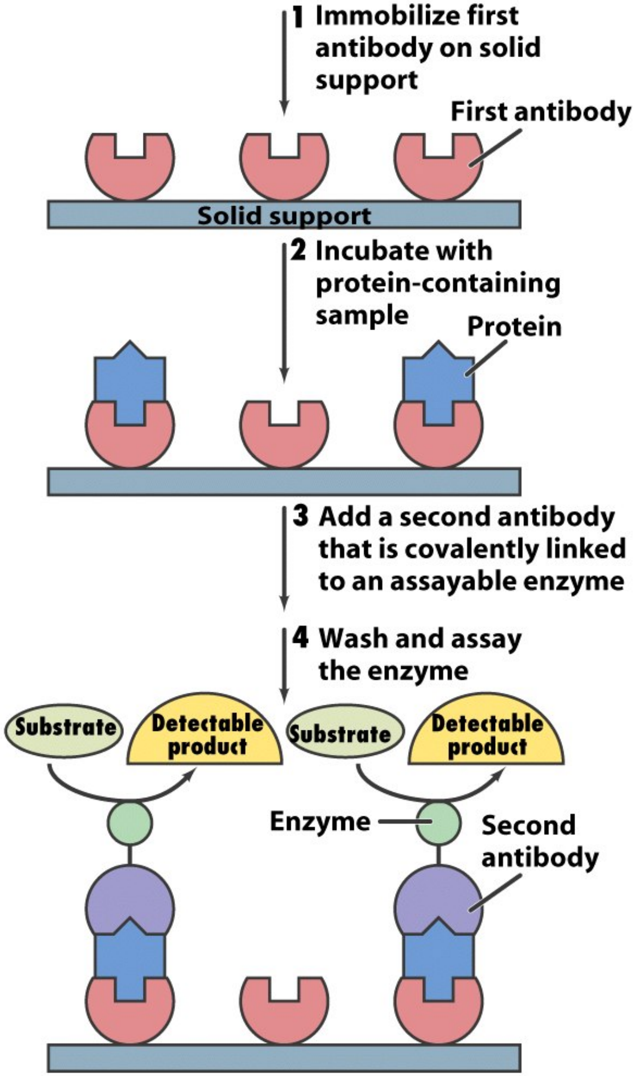


Figure 5-3 Fundamentals of Biochemistry, 2/e

© 2006 John Wiley & Sons

Absorbance spectroscopy

Beer-Lambert law
A= log (I_o/I) = εcl
A: absorbance (optical density)
I_o: incident light intensity
I: transmitted intensity
ε: molar absorptivity (molar extinction coefficient)
c: molar concentration
l: the length of light path

<u>Protein</u>
UV assay: absorption at 280 nm (~1 mg/ml)
    Chromophore: absorb light in the visible region

Dye-binding assay:
    Bradford assay (Coomassie brilliant blue)L at 595 nm

UV-absorption spectra

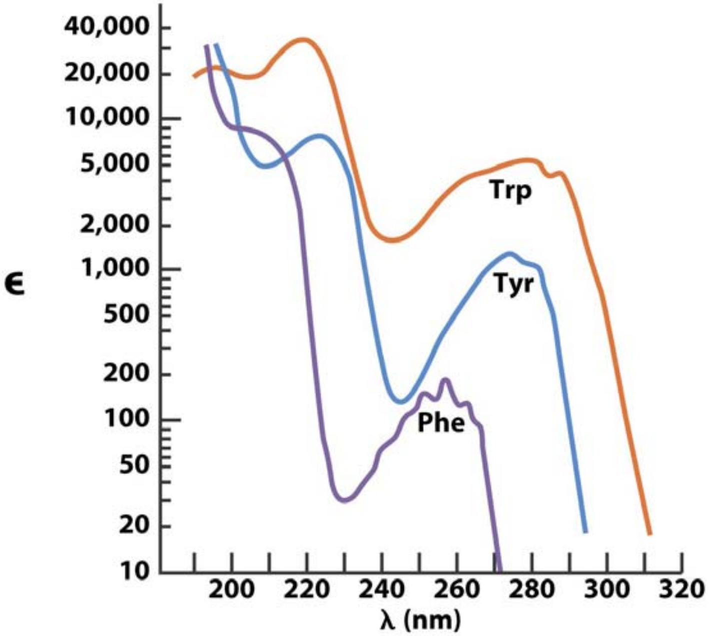
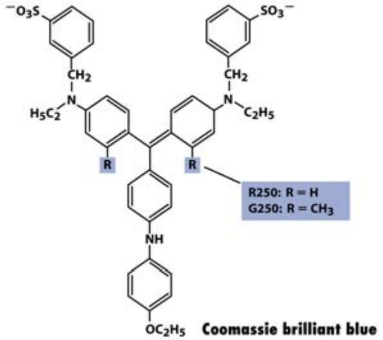
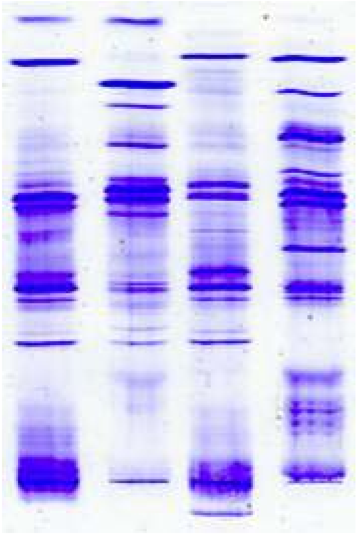


Figure 5-4 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons



Unnumbered figure pg 100 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons



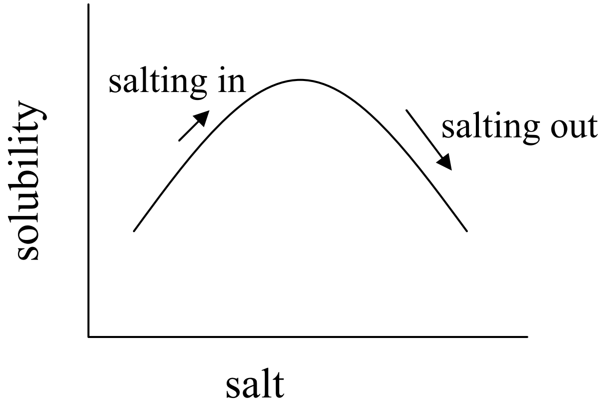
Separation techniques

Purification of target protein by fractionation procedures
Selective elimination of the other components
Using physicochemical properties of proteins

Protein characteristic	**purification procedure**
Solubility	salting out
Ionic charge	ion exchange chromatography
	electrophoresis
	ioselectric focusing
Polarity	hydrophobic interaction chromatography
Size	gel filtration chromatography
	SDS-PAGE
	ultracentrifugation
Binding specificity	affinity chromatography

Salting in

Proteins are polyelectrolytes (polyionic polymers)
Proteins are least soluble when its net charge is zero



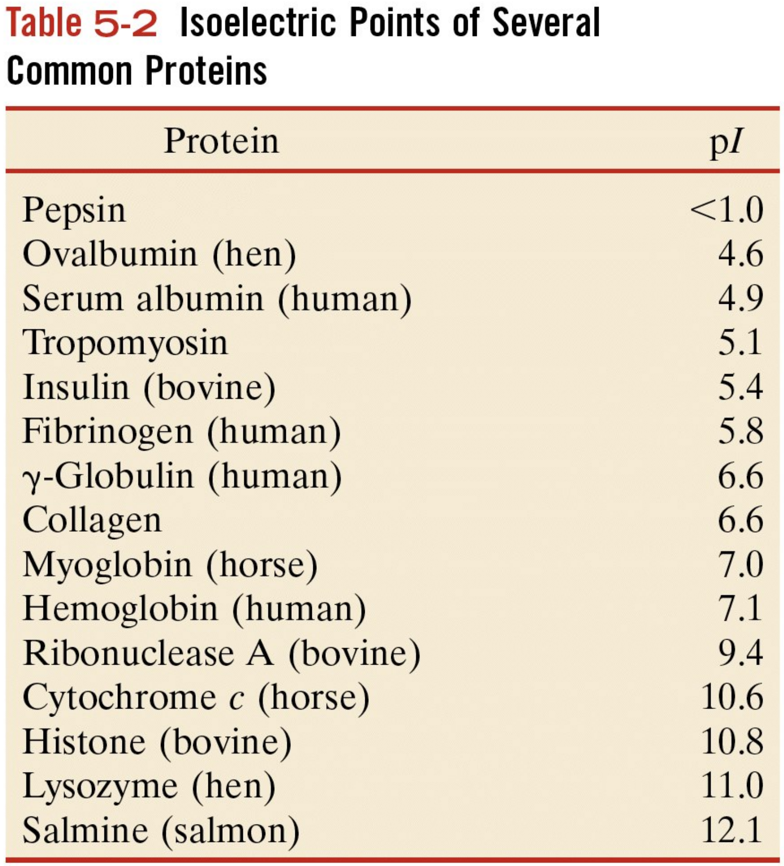


Table 5-2 Isoelectric Points of Several Common Proteins

Protein	p*I*
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 *c* (horse)	10.6
Histone (bovine)	10.8
Lysozyme (hen)	11.0
Salmine (salmon)	12.1

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

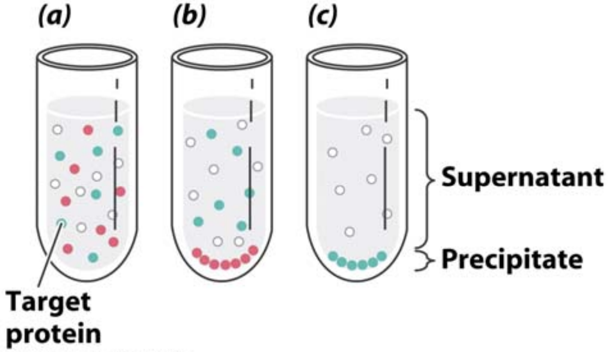
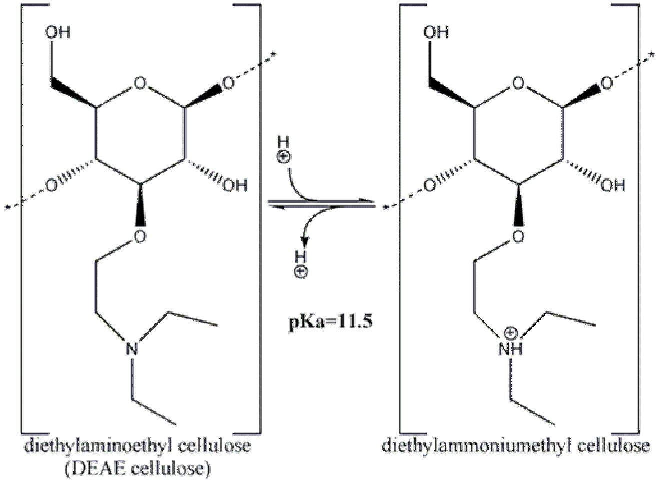


Figure 5-5 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Chromatography

Mobile phase: liquid
Stationary phase: porous solid matrix
From paper chromatography to HPLC



Ion-exchange chromatography
Anion exchanger: diethylaminoethyl (DEAE)
Cation exchanger: carboxy-methyl (CM) $–CH_2-COO^-$

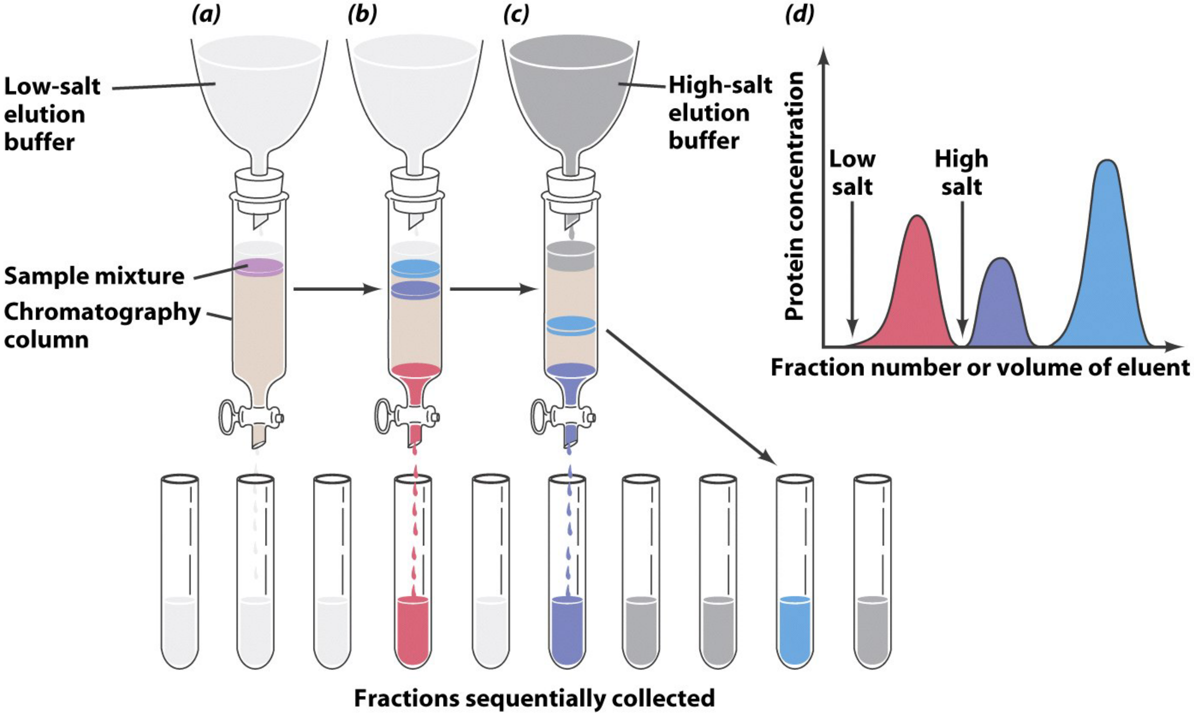
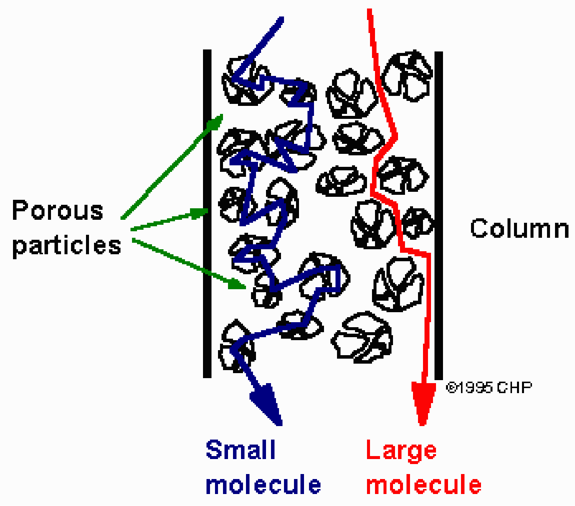


Figure 5-6 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

HIC (hydrophobic interaction chromatography)

Gel filtration chromatography

(size exclusion or molecular sieve chromatography)



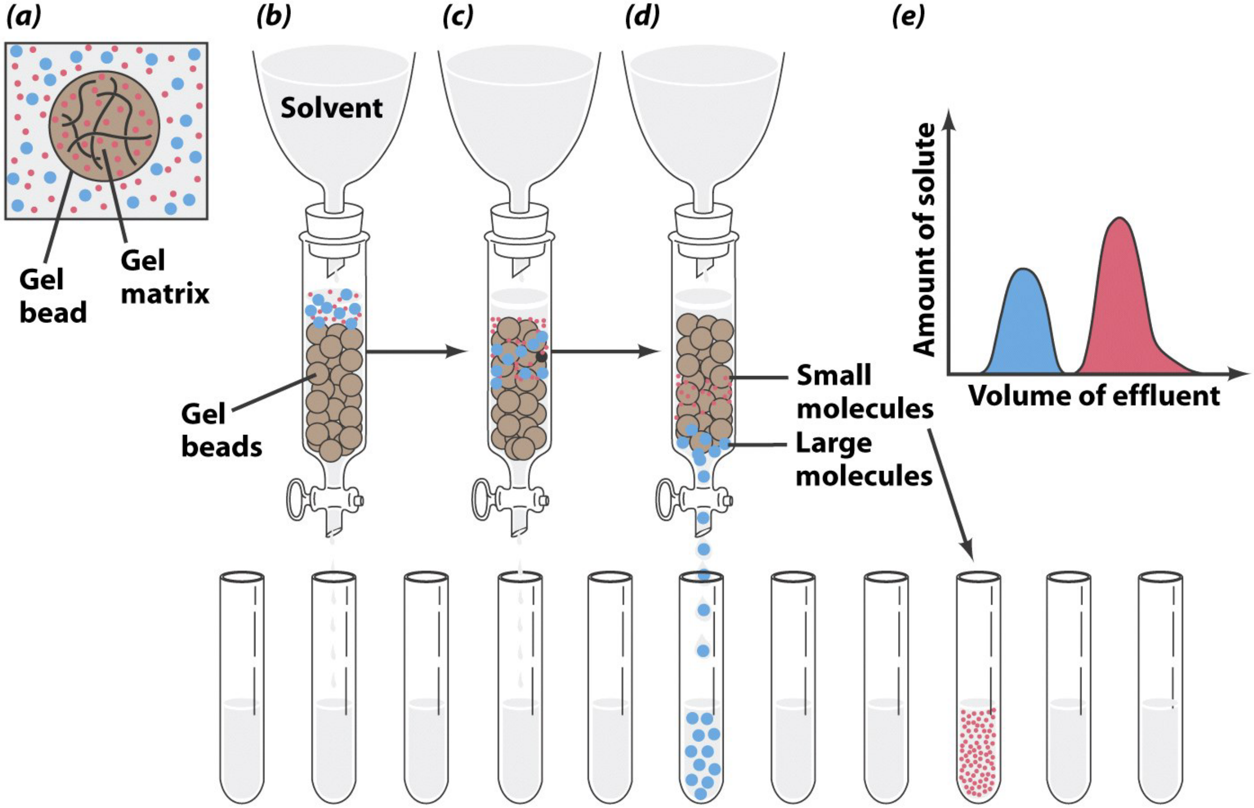


Figure 5-7 Fundamentals of Biochemistry, 2/e

Affinity chromatography

Ligand bound: cAMP, NADH, etc
Antibody bound: immunoaffinity chromatography
Metal bound: metal chelate affinity chromatography
    Zn2+, Ni2+
    his-tag column

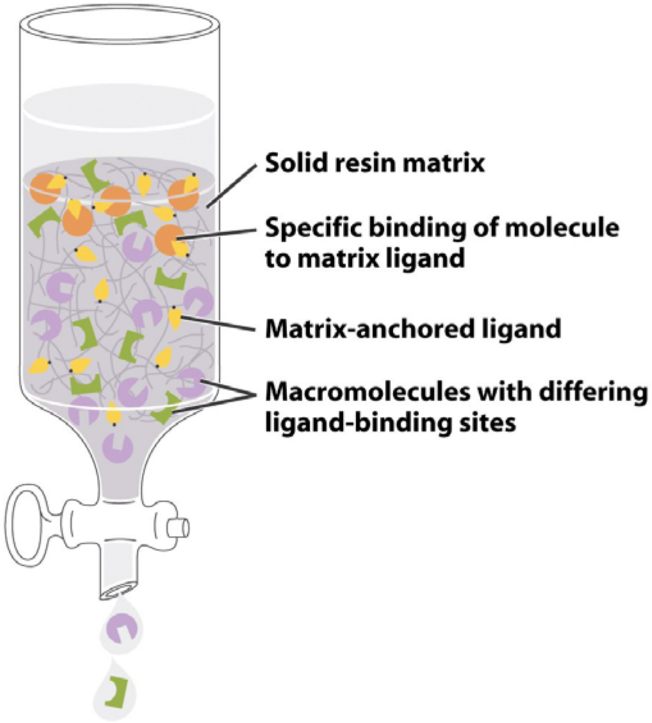
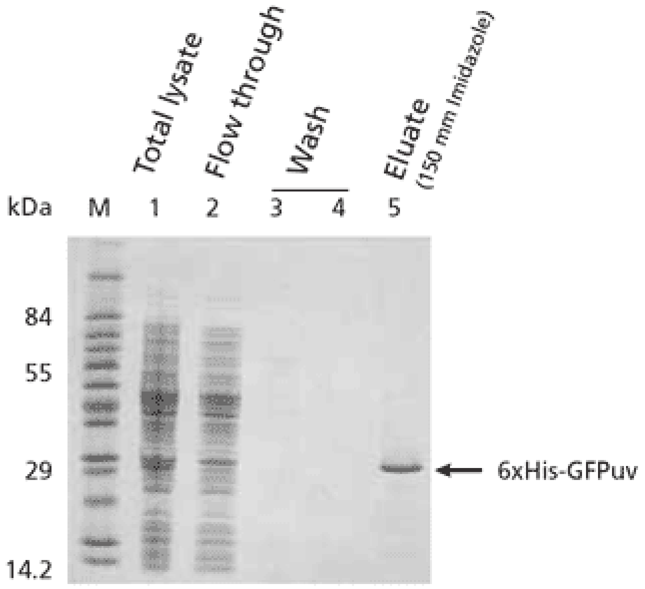


Figure 5-8 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Nickel Affinity Gel

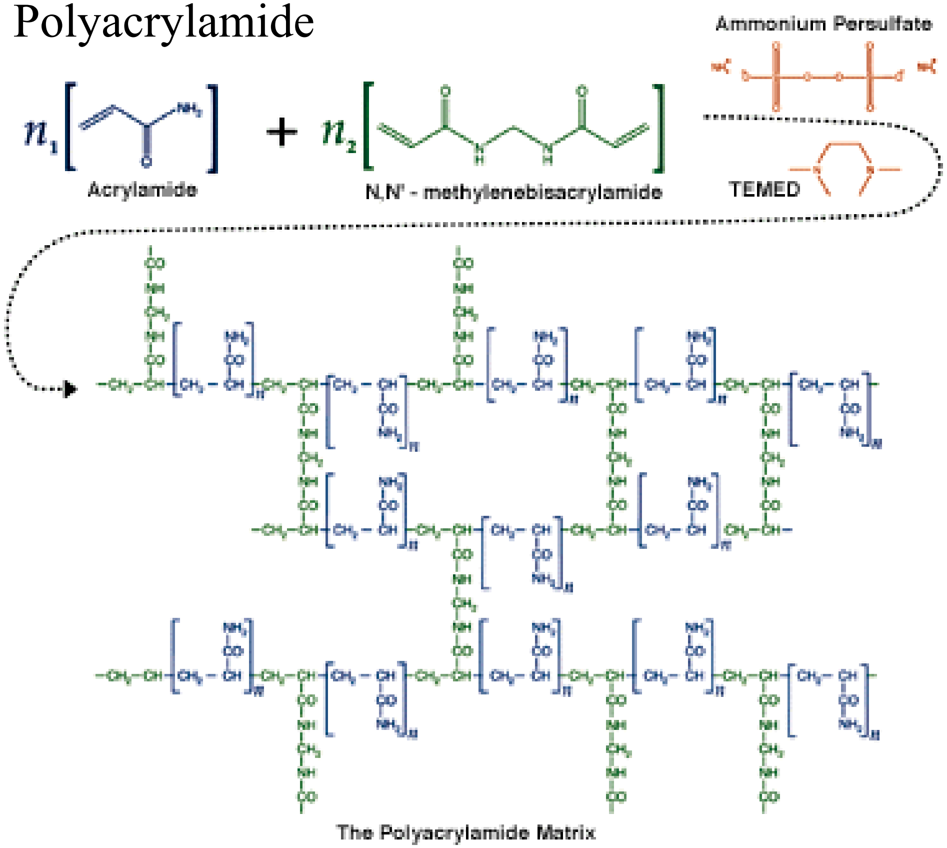


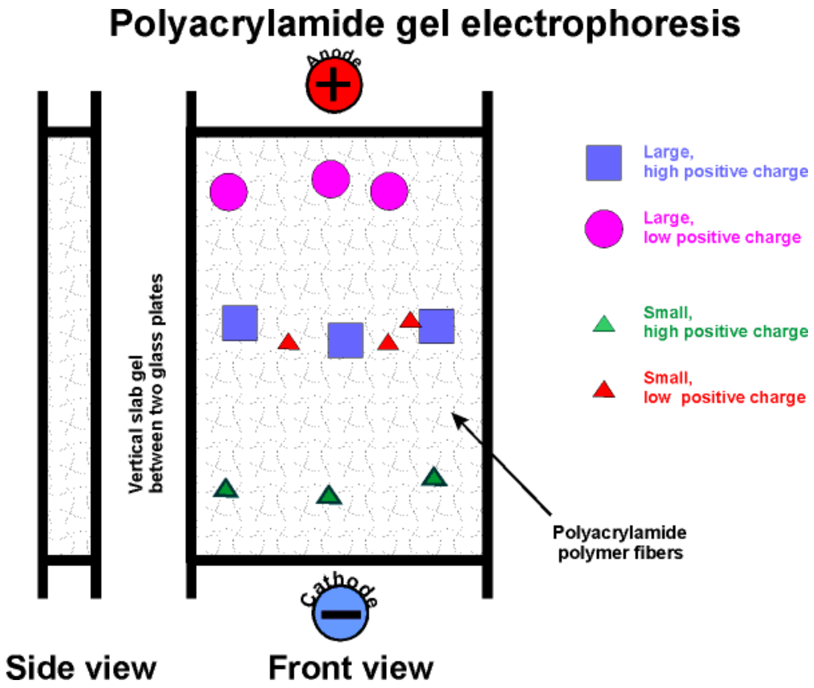
Electrophoresis

Polyacrylamide gel electrophoresis

Mobility difference depending on size, shape, electric charge

Polyacrylamide





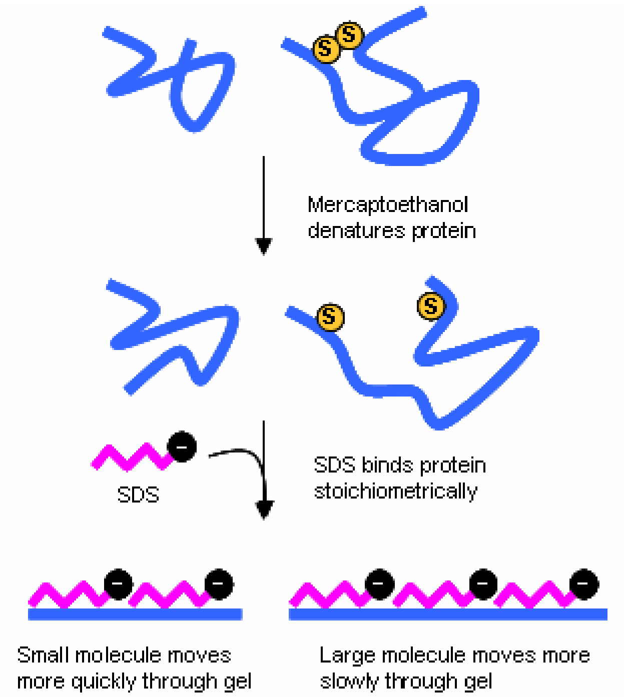
SDS-PAGE

SDS: sodium dodecylsulfate

Uniform binding of SDS to protein and denaturation

Net charge is equal in every protein

Separation depends on size



Mercaptoethanol denatures protein

SDS binds protein stoichiometrically

Small molecule moves more quickly through gel

Large molecule moves more slowly through gel

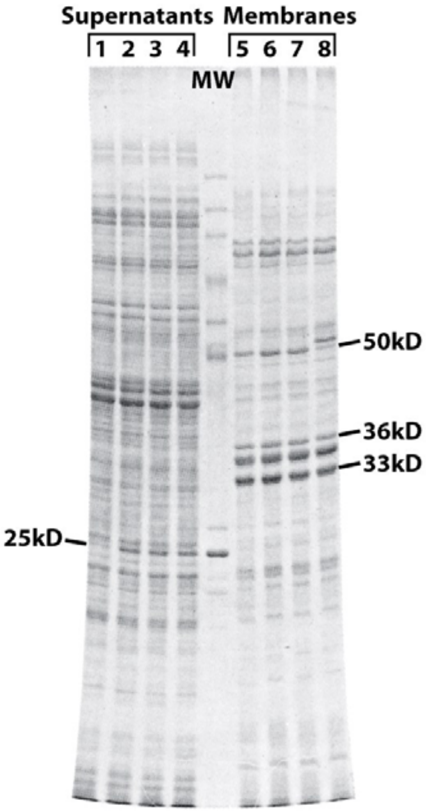


Figure 5-9 Fundamentals of Biochemistry, 2/e

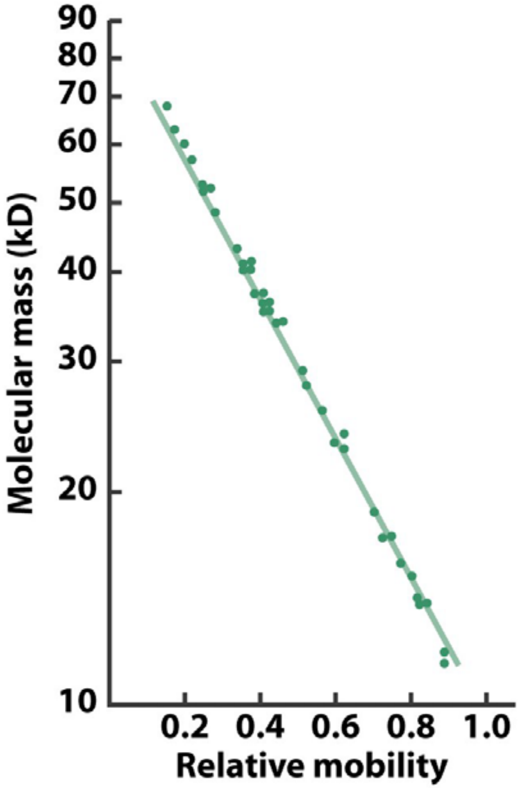


Figure 5-10 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Capillary electrophoresis

Electrophoresis in very thin capillary tubes (20-100 um diameter)

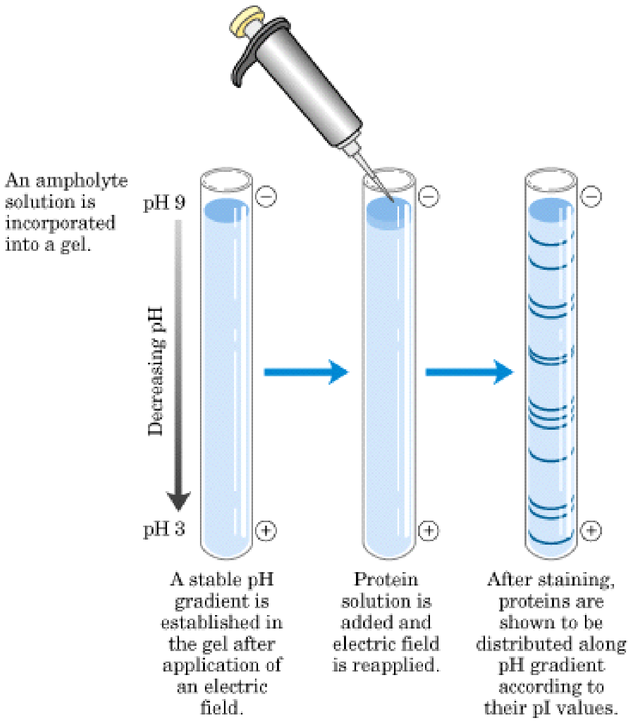
Dissipate heat and permit high electric fields

Fast separation

2D-PAGE

Isoelectric focusing (IEF) & SDS-PAGE

Application to proteomics



+ (pH 3) - (pH 10)

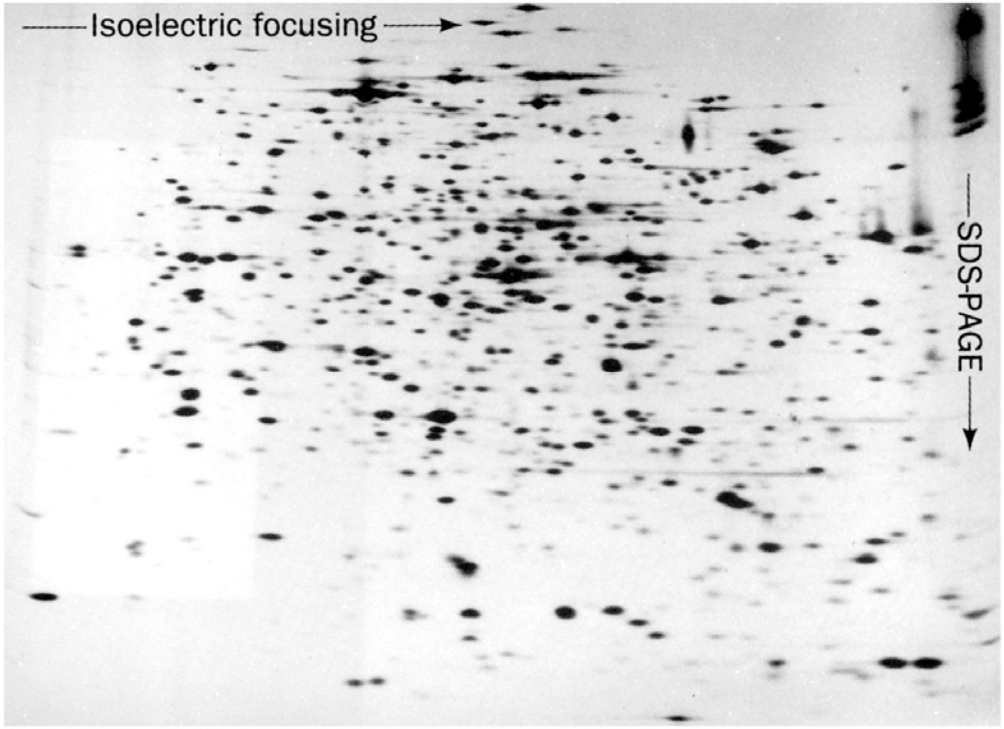
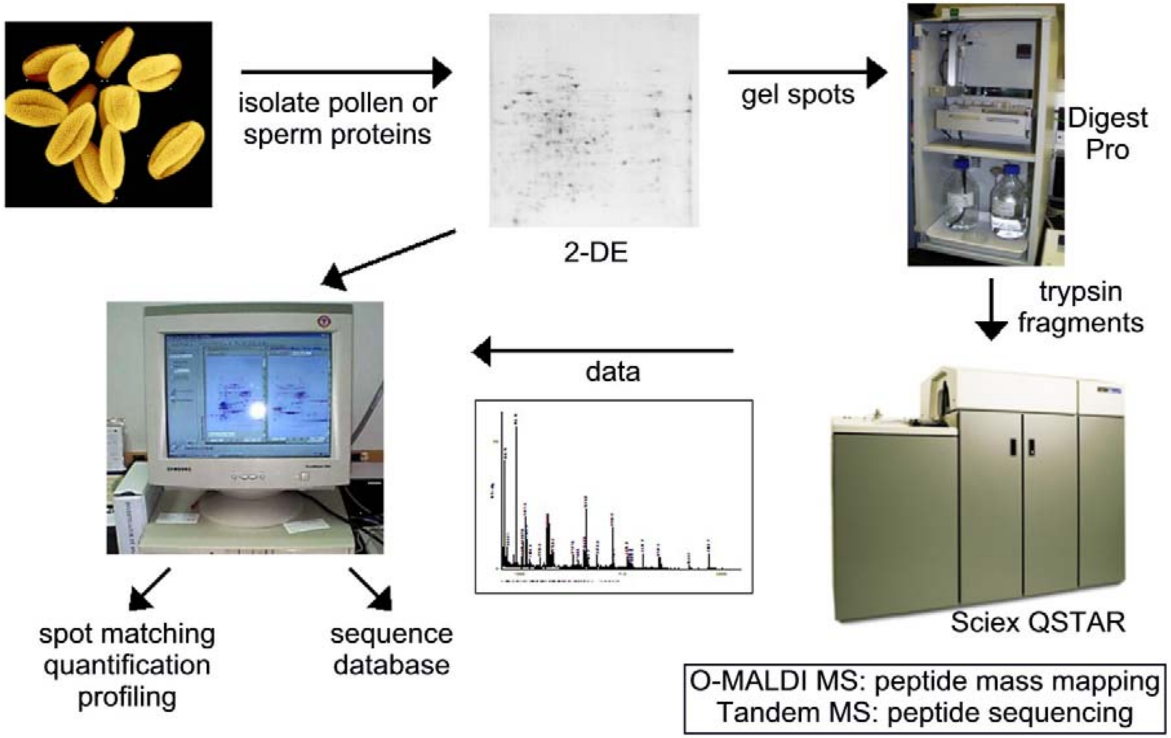


Figure 5-11 Fundamentals of Biochemistry, 2/e



isolate pollen or sperm proteins

2-DE

gel spots

Digest Pro

trypsin fragments

Sciex QSTAR

O-MALDI MS: peptide mass mapping
Tandem MS: peptide sequencing

data

spot matching
quantification
profiling

sequence
database

Ultracentrifugation

Analytical
Preparative
Density gradient centrifugation (zonal centrifugation): sucrose, Percoll
Equilibrium density gradient centrifugation: CsCl

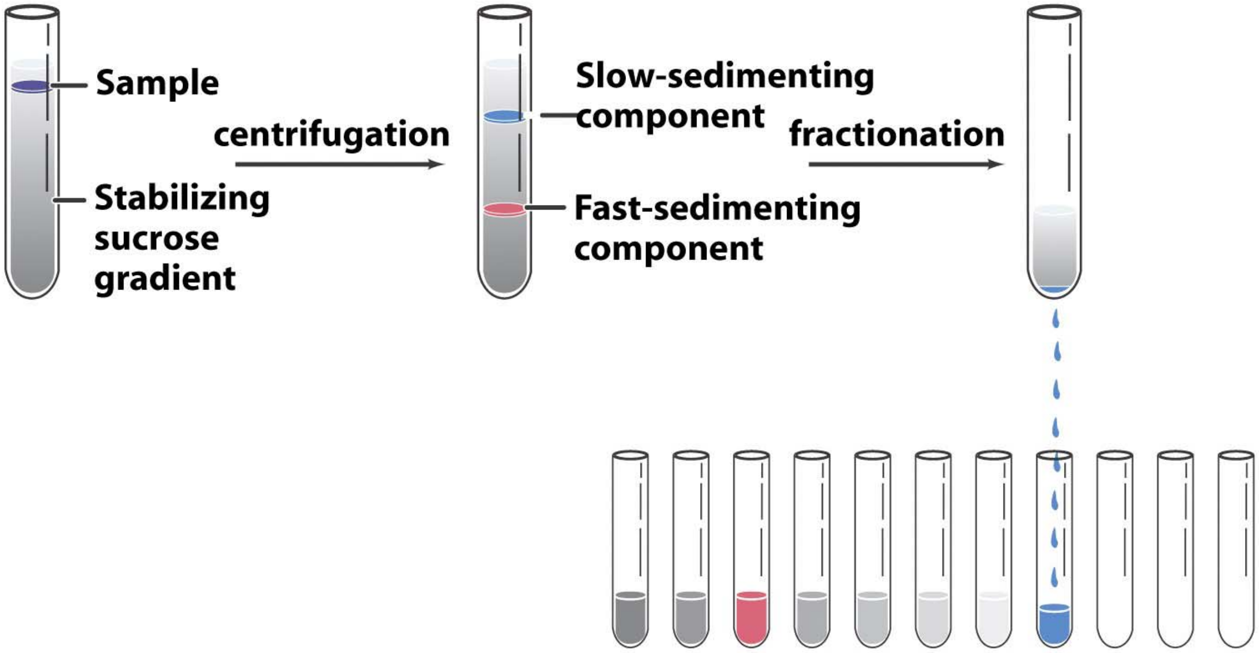
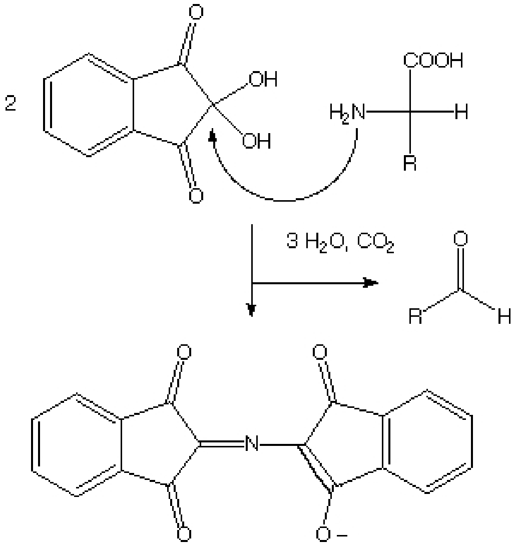
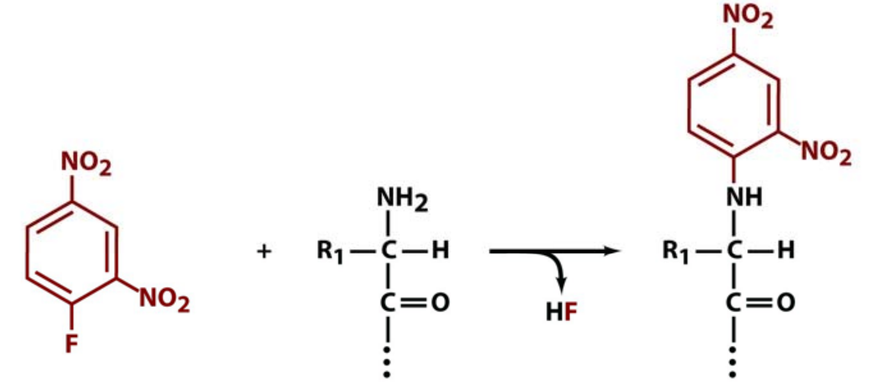


Figure 5-12 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Staining of amino acids and polypeptides





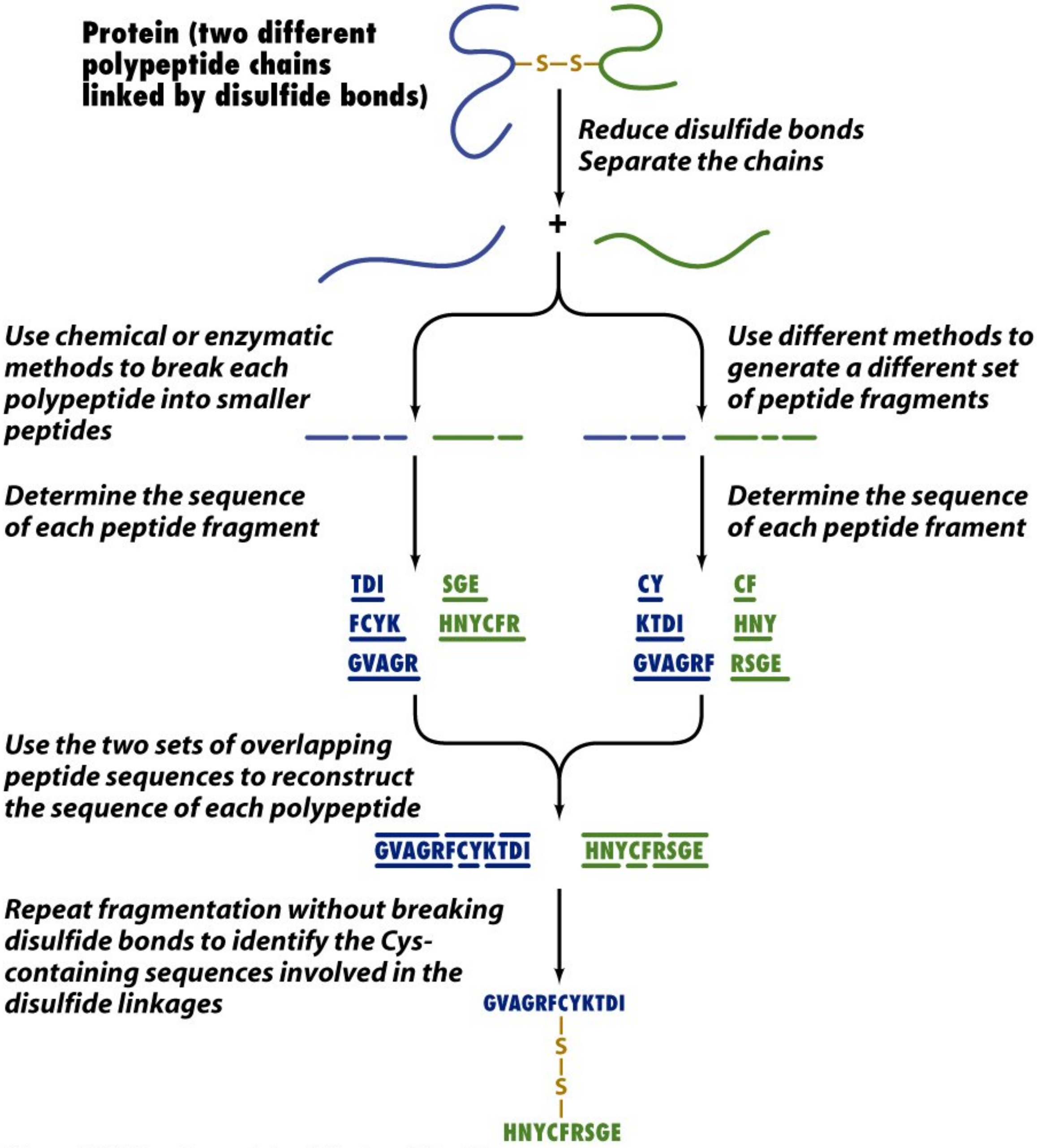
2,4-Dinitrofluorobenzene (DNFB)

Polypeptide
N-terminal

DNP Polypeptide

Box 5-2 Fundamentals of Biochemistry, 2/e

Protein sequencing



Protein (two different polypeptide chains linked by disulfide bonds)

S–S

Reduce disulfide bonds
Separate the chains

+

Use chemical or enzymatic methods to break each polypeptide into smaller peptides

Use different methods to generate a different set of peptide fragments

Determine the sequence of each peptide fragment

Determine the sequence of each peptide frament

TDI
FCYK
GVAGR

SGE
HNYCFR

CY
KTDI
GVAGRF

CF
HNY
RSGE

Use the two sets of overlapping peptide sequences to reconstruct the sequence of each polypeptide

GVAGRFCYKTDI

HNYCFRSGE

Repeat fragmentation without breaking disulfide bonds to identify the Cys-containing sequences involved in the disulfide linkages

GVAGRFCYKTDI
S
S
HNYCFRSGE

Figure 5-13 Fundamentals of Biochemistry, 2/e

Preliminary steps

N-terminal analysis: dansy-chloride
  differentiate subunits
Disulfide bond cleavage
    Reduction with reducing reagent (2-mercaptoethanol or dithiothreitol): cystine to 2 cysteine
    Alkylatio with iodoacetate: Cys to S-carboxymethyl-Cys
Determination of amino acid composition: complete hydrolysis with HCl

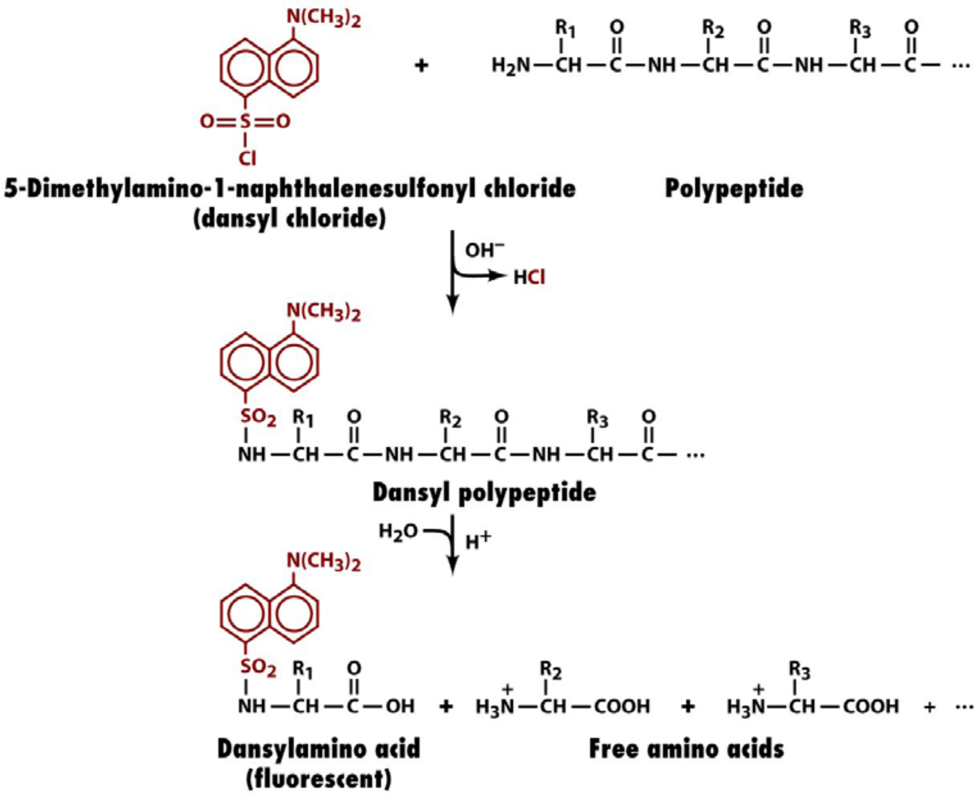


Figure 5-14 Fundamentals of Biochemistry, 2/e

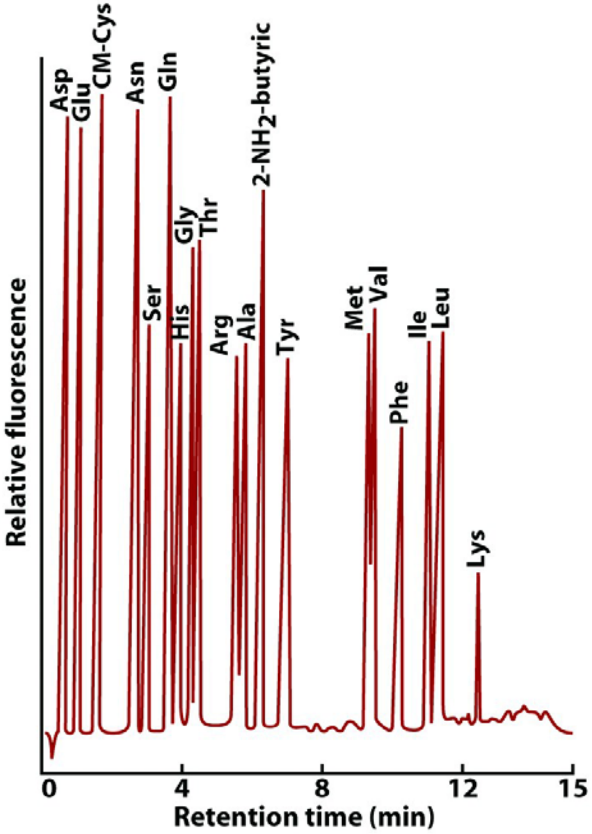
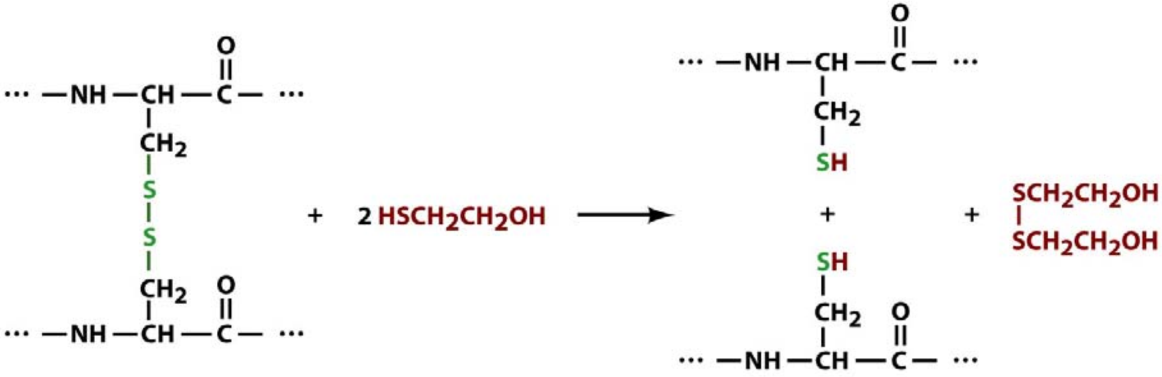


Figure 5-15 Fundamentals of Biochemistry, 2/e



Cystine **2-Mercaptoethanol** **Cysteine**

Unnumbered figure pg 112a Fundamentals of Biochemistry, 2/e

$$\text{Cys}-\text{CH}_2-\text{SH} + \text{ICH}_2\text{COO}^- \longrightarrow \text{Cys}-\text{CH}_2-\text{S}-\text{CH}_2\text{COO}^- + \text{HI}$$

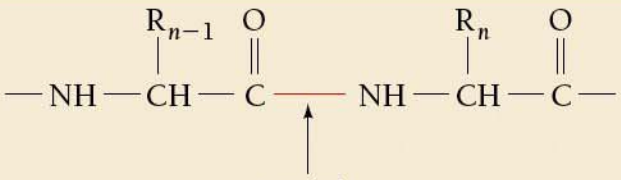
Cysteine **Iodoacetate** ***S*-Carboxymethylcysteine (CM-Cys)**

Unnumbered figure pg 112b Fundamentals of Biochemistry, 2/e

Polypeptide cleavage

Endopeptidase
exopeptidase

Table 5-3 Specificities of Various Endopeptidases



$$-NH-\overset{\overset{R_{n-1}}{|}}{CH}-\overset{\overset{O}{\|}}{C}-NH-\overset{\overset{R_n}{|}}{CH}-\overset{\overset{O}{\|}}{C}-$$

Scissile
peptide bond

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	*Bacillus thermoproteolyticus*	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	*Staphylococcus aureus*	R_{n-1} = Glu	

Table 5-3 Fundamentals of Biochemistry, 2/e

CNBr cleavage: C-side of Met residue

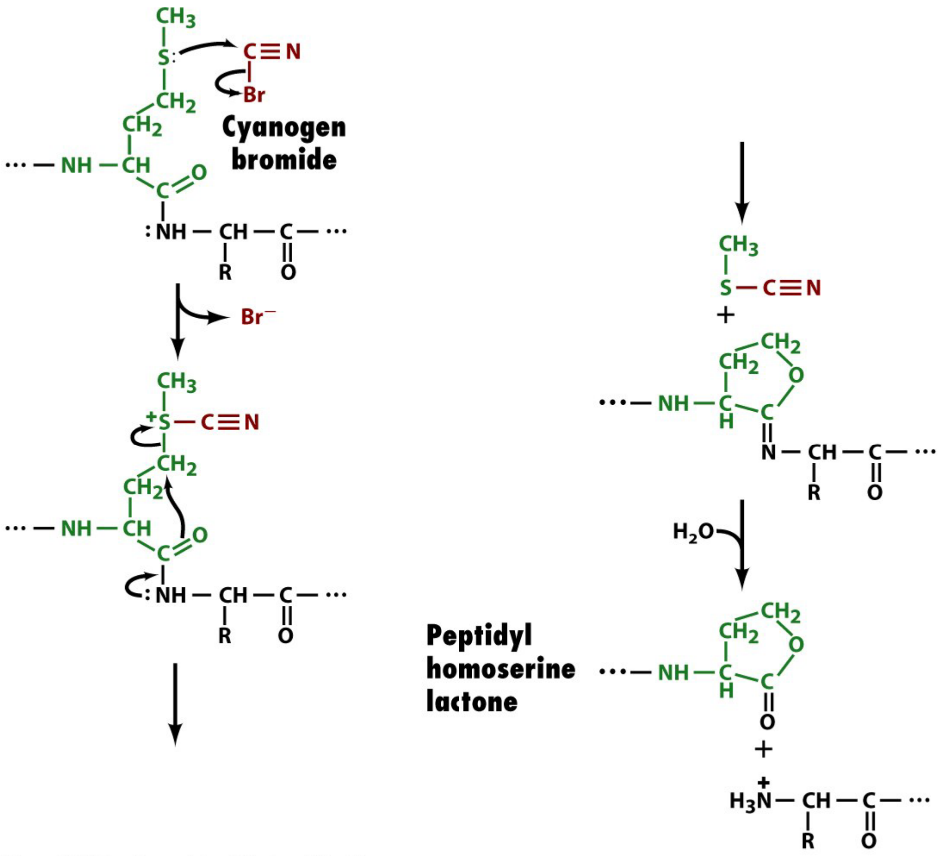


Figure 5-16 Fundamentals of Biochemistry, 2/e

Edman degradation

Sequential cleavage and identification of N-terminal residues

Solid-phase sequencing

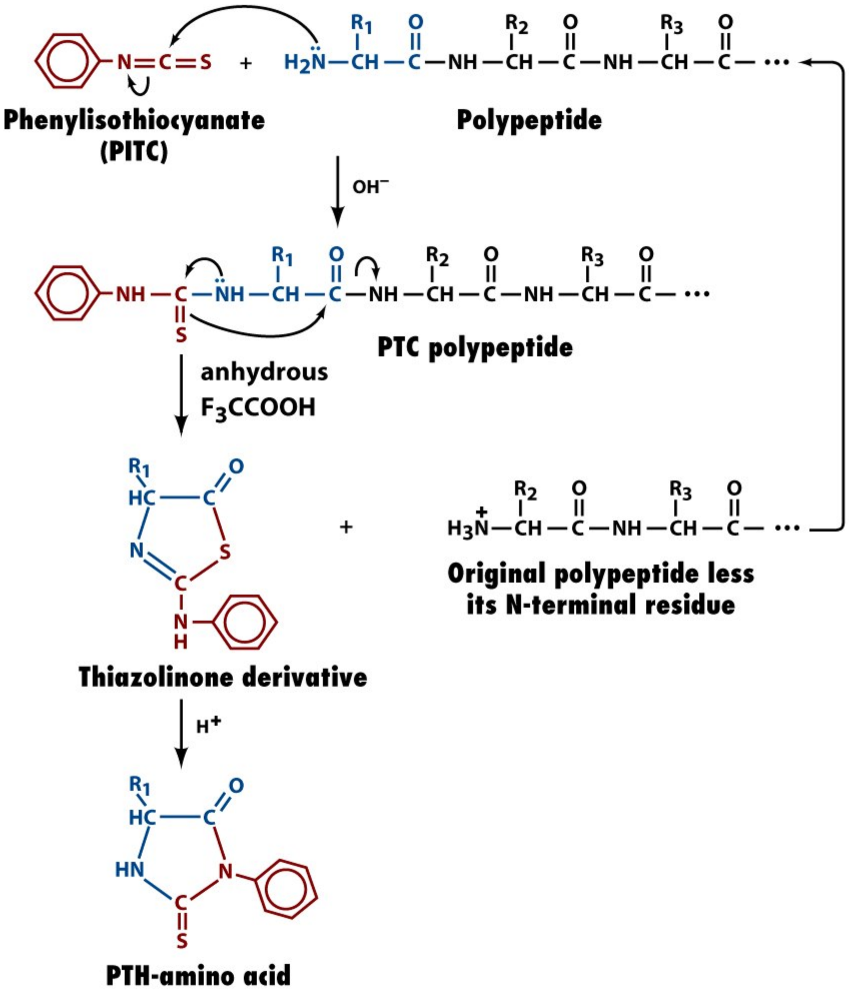


Figure 5-17 Fundamentals of Biochemistry, 2/e

Sequencing by mass spectrometry

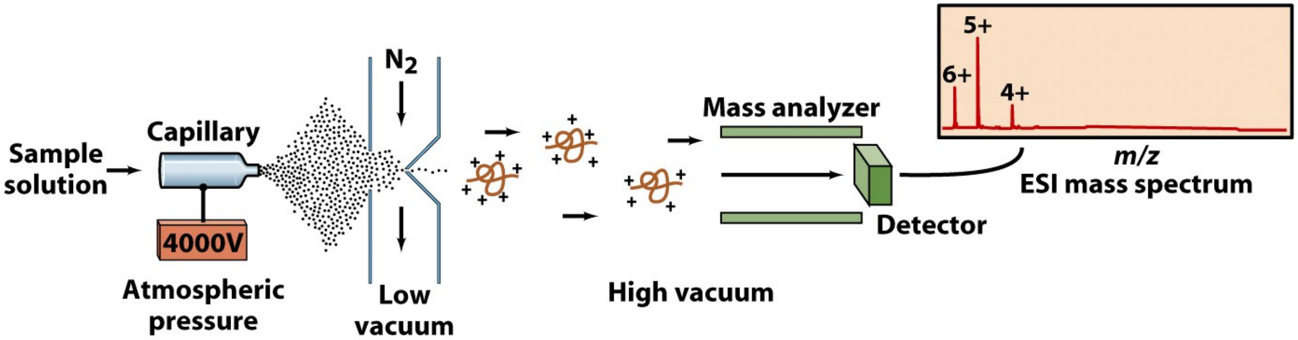


Figure 5-18a Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

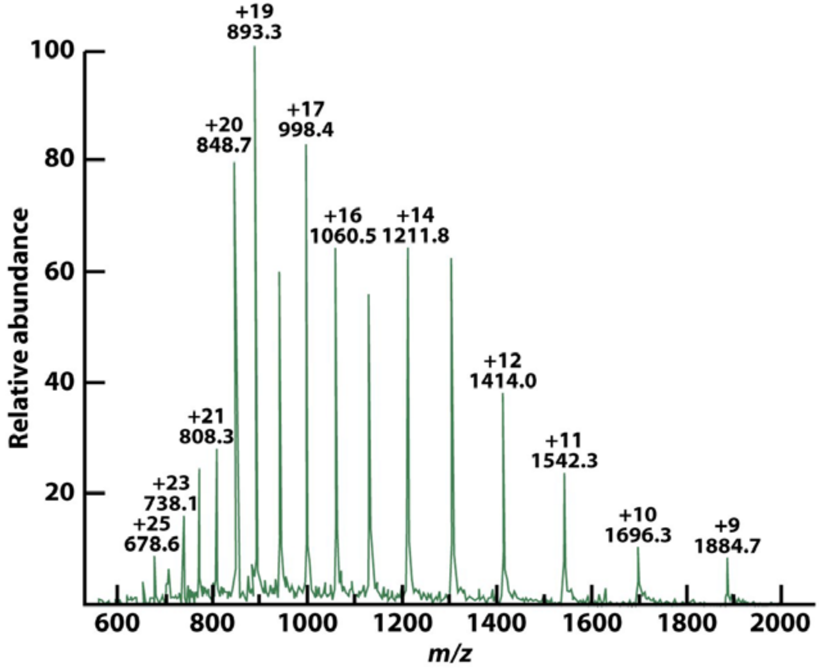


Figure 5-18b Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

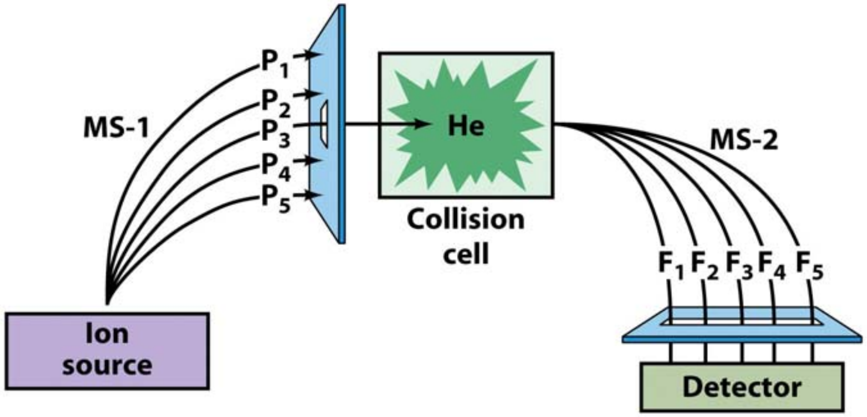


Figure 5-19 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons

Reconstructing the protein’s sequence

Localization of disulfide bond

Cleave protein with its disulfide bonds intact

Alkylation and sequencing

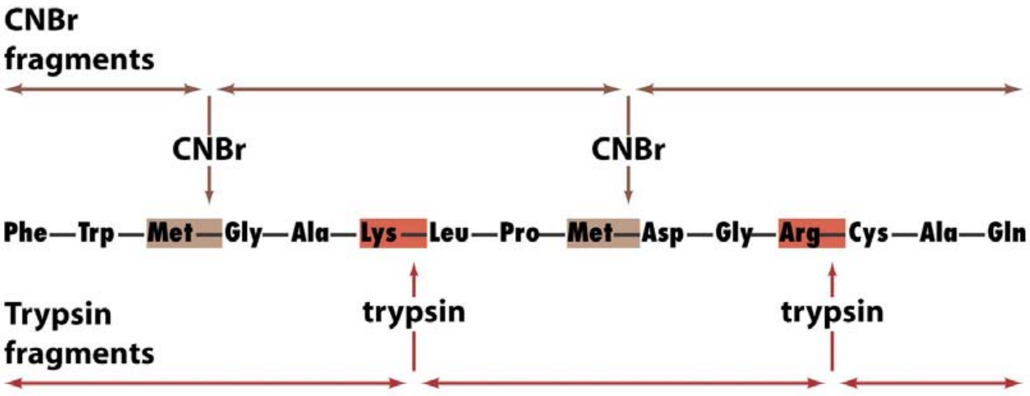


Figure 5-20 Fundamentals of Biochemistry, 2/e

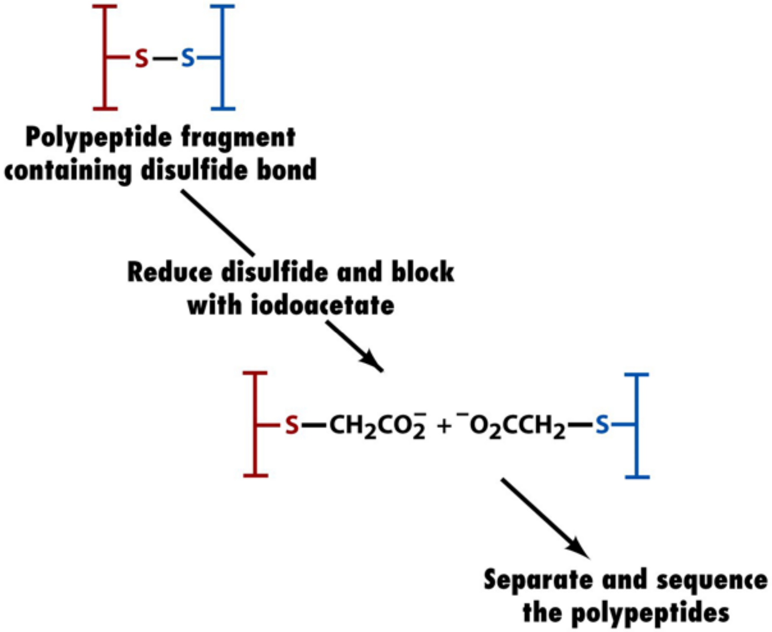


Figure 5-21 Fundamentals of Biochemistry, 2/e

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

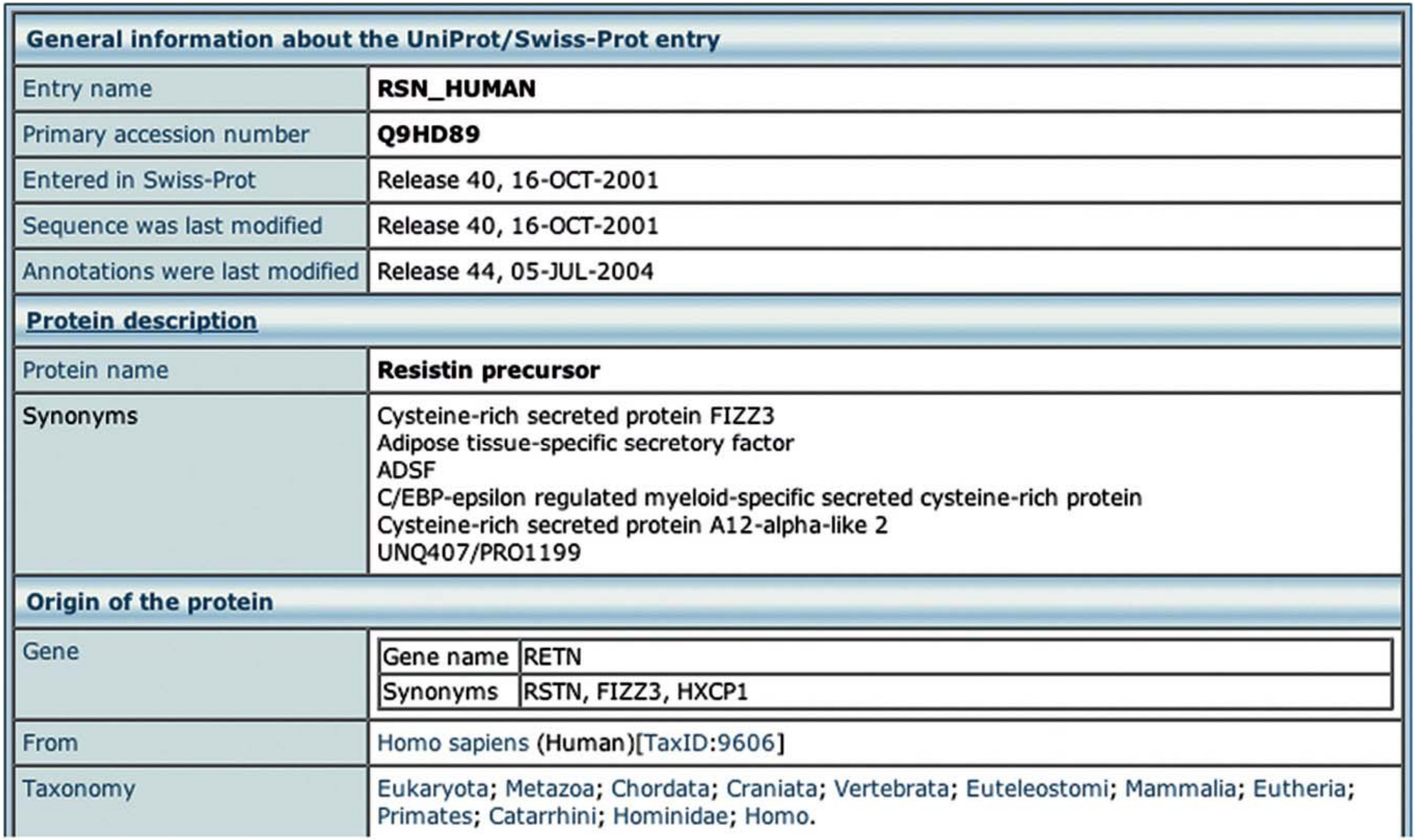
Data Banks Containing Protein Sequences

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

Table 5-4 Fundamentals of Biochemistry, 2/e



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

Darwinian evolution: many mutations are deleterious and disappear
some mutations are beneficial and survive

Protein sequence evolution

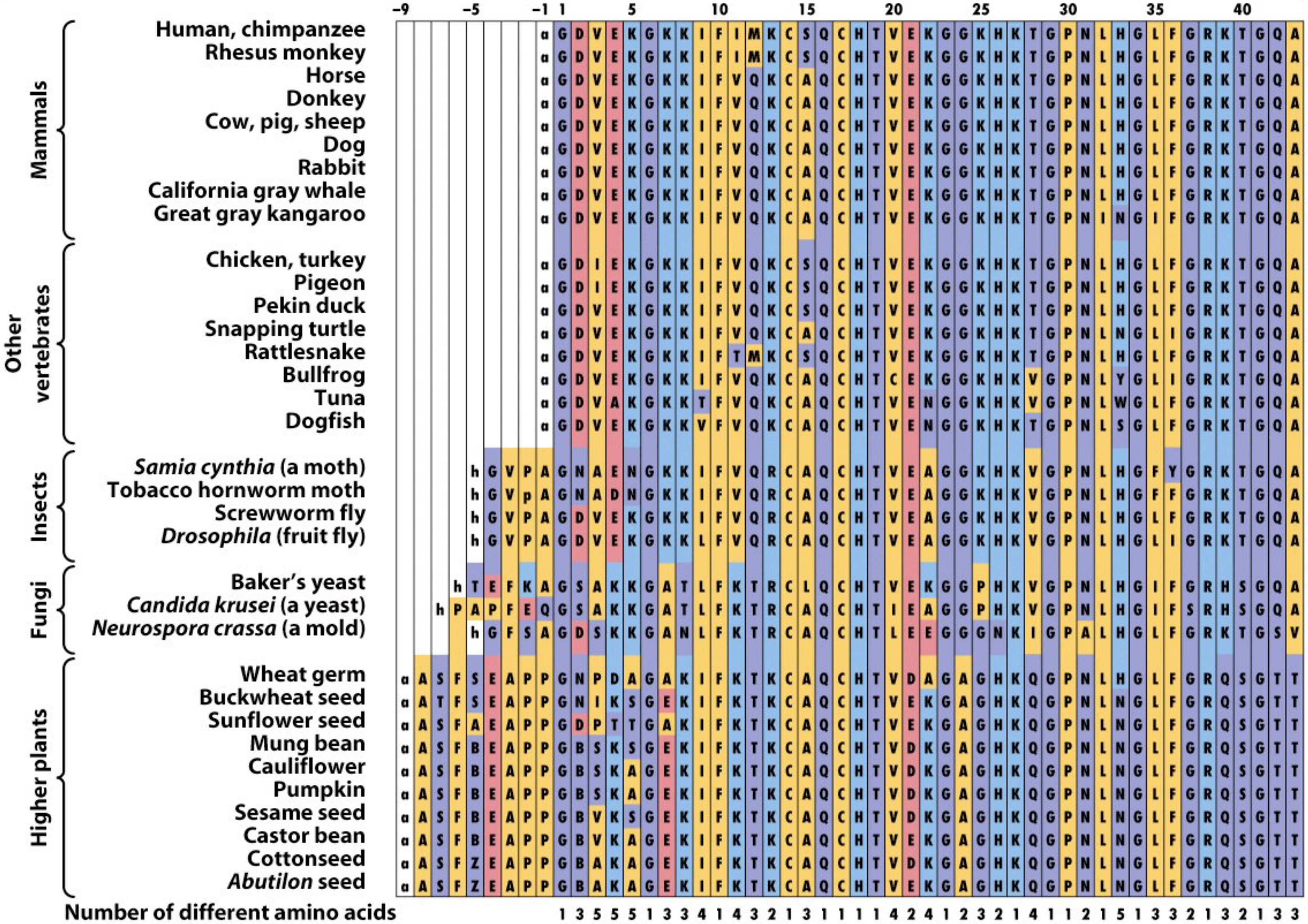
- Sequence similarity
  - Multiple alignment
  - Evolution from a common ancestor
  - Evolutionary conservative: the meaning of biological importance

Sequence comparisons provide information on protein structure and function

- Homologous proteins
  - Invariant residues: essential
  - Conservatively substituted: similar side chains
  - Hypervariable: nonspecific

If cyt.c is important, why not identical?
- the random nature of mutational processes: neutral drift
- depends on tolerable or untolerable

TABLE 5-5 Amino Acid Sequences of Cytochrome *c* from 38 species[a]



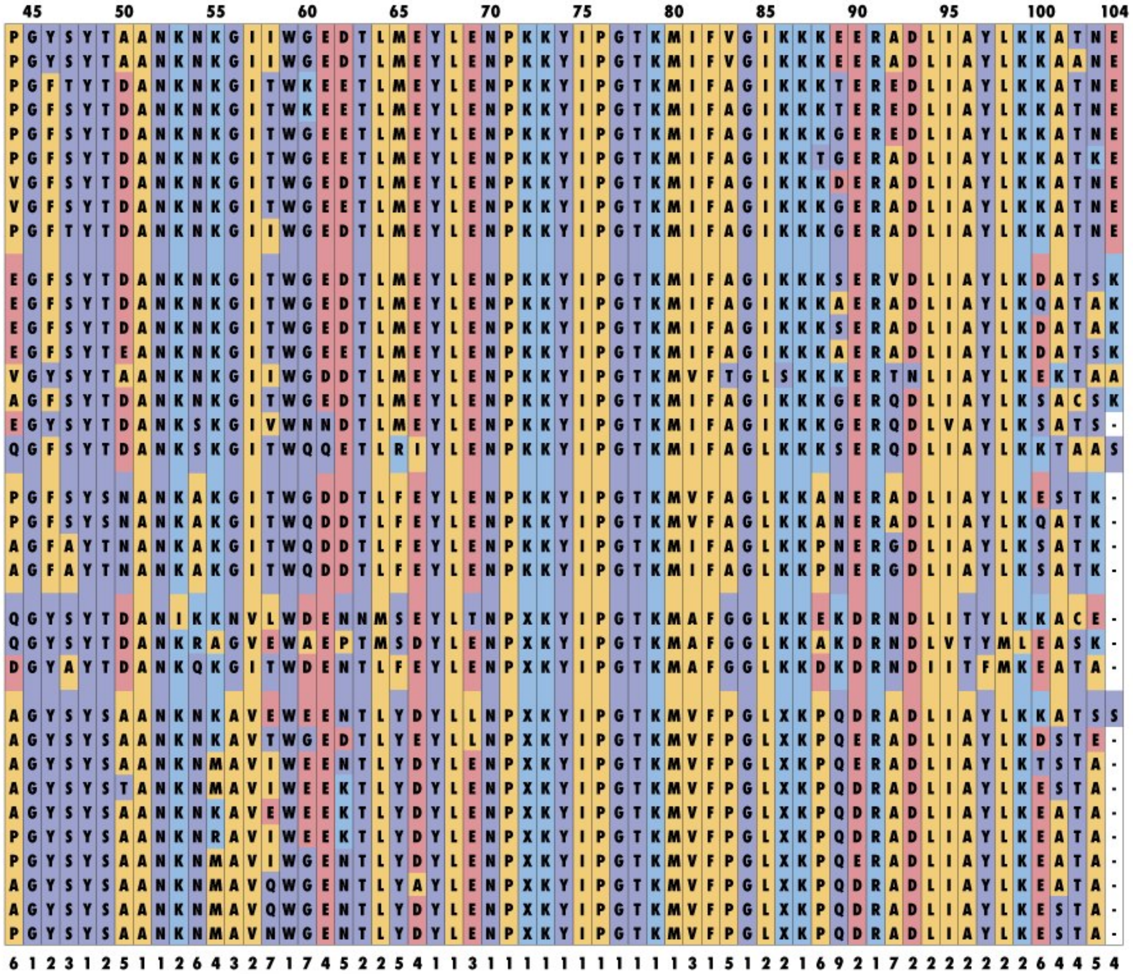
		-9				-5				-1	1				5					10					15					20					25					30					35					40			
Mammals	Human, chimpanzee									a	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Rhesus monkey									a	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Horse									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Donkey									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Cow, pig, sheep									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Dog									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Rabbit									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	California gray whale									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Great gray kangaroo									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	I	N	G	I	F	G	R	K	T	G	Q	A
Other vertebrates	Chicken, turkey									a	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Pigeon									a	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Pekin duck									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Snapping turtle									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	N	G	L	I	G	R	K	T	G	Q	A
	Rattlesnake									a	G	D	V	E	K	G	K	K	I	F	T	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Bullfrog									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	C	E	K	G	G	K	H	K	V	G	P	N	L	Y	G	L	I	G	R	K	T	G	Q	A
	Tuna									a	G	D	V	A	K	G	K	K	T	F	V	Q	K	C	A	Q	C	H	T	V	E	N	G	G	K	H	K	V	G	P	N	L	W	G	L	F	G	R	K	T	G	Q	A
	Dogfish									a	G	D	V	E	K	G	K	K	V	F	V	Q	K	C	A	Q	C	H	T	V	E	N	G	G	K	H	K	T	G	P	N	L	S	G	L	F	G	R	K	T	G	Q	A
Insects	*Samia cynthia* (a moth)					h	G	V	P	A	G	N	A	E	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	F	Y	G	R	K	T	G	Q	A
	Tobacco hornworm moth					h	G	V	p	A	G	N	A	D	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	F	F	G	R	K	T	G	Q	A
	Screwworm fly					h	G	V	P	A	G	D	V	E	K	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Drosophila (fruit fly)					h	G	V	P	A	G	D	V	E	K	G	K	K	L	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	I	G	R	K	T	G	Q	A
Fungi	Baker's yeast				h	T	E	F	K	A	G	S	A	K	K	G	A	T	L	F	K	T	R	C	L	Q	C	H	T	V	E	K	G	G	P	H	K	V	G	P	N	L	H	G	I	F	G	R	H	S	G	Q	A
	Candida krusei (a yeast)			h	P	A	P	F	E	Q	G	S	A	K	K	G	A	T	L	F	K	T	R	C	A	Q	C	H	T	I	E	A	G	G	P	H	K	V	G	P	N	L	H	G	I	F	S	R	H	S	G	Q	A
	Neurospora crassa (a mold)					h	G	F	S	A	G	D	S	K	K	G	A	N	L	F	K	T	R	C	A	Q	C	H	T	L	E	E	G	G	G	N	K	I	G	P	A	L	H	G	L	F	G	R	K	T	G	S	V
Higher plants	Wheat germ	a	A	S	F	S	E	A	P	P	G	N	P	D	A	G	A	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	A	G	A	G	H	K	Q	G	P	N	L	H	G	L	F	G	R	Q	S	G	T	T
	Buckwheat seed	a	A	T	F	S	E	A	P	P	G	N	I	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Sunflower seed	a	A	S	F	A	E	A	P	P	G	D	P	T	T	G	A	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Mung bean	a	A	S	F	B	E	A	P	P	G	B	S	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Cauliflower	a	A	S	F	B	E	A	P	P	G	B	S	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Pumpkin	a	A	S	F	B	E	A	P	P	G	B	S	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Sesame seed	a	A	S	F	B	E	A	P	P	G	B	V	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Castor bean	a	A	S	F	B	E	A	P	P	G	B	V	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Cottonseed	a	A	S	F	Z	E	A	P	P	G	B	A	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Abutilon seed	a	A	S	F	Z	E	A	P	P	G	B	A	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Number of different amino acids										1	3	5	5	5	1	3	3	4	1	4	3	2	1	3	1	1	1	1	4	2	4	1	2	3	2	1	4	1	1	2	1	5	1	3	3	2	1	3	2	1	3	3

---

[a] The 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 indi-cates 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
P	G	Y	S	Y	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	Y	S	Y	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	A	N	E
P	G	F	T	Y	T	D	A	N	K	N	K	G	I	T	W	K	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	K	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	T	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	T	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	K	E
V	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	D	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
V	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
P	G	F	T	Y	T	D	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	V	D	L	I	A	Y	L	K	D	A	T	S	K
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	Q	A	T	A	K
E	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	A	D	L	I	A	Y	L	K	D	A	T	A	K
E	G	F	S	Y	T	E	A	N	K	N	K	G	I	T	W	G	E	E	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	D	A	T	S	K
V	G	Y	S	Y	T	A	A	N	K	N	K	G	I	I	W	G	D	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	T	G	L	S	K	K	K	E	R	T	N	L	I	A	Y	L	K	E	K	T	A	A
A	G	F	S	Y	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	Q	D	L	I	A	Y	L	K	S	A	C	S	K
E	G	Y	S	Y	T	D	A	N	K	S	K	G	I	V	W	N	N	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	Q	D	L	V	A	Y	L	K	S	A	T	S	-
Q	G	F	S	Y	T	D	A	N	K	S	K	G	I	T	W	Q	Q	E	T	L	R	I	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	K	S	E	R	Q	D	L	I	A	Y	L	K	K	T	A	A	S
P	G	F	S	Y	S	N	A	N	K	A	K	G	I	T	W	G	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	A	G	L	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	E	S	T	K	-
P	G	F	S	Y	S	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	V	F	A	G	L	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	Q	A	T	K	-
A	G	F	A	Y	T	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	P	N	E	R	G	D	L	I	A	Y	L	K	S	A	T	K	-
A	G	F	A	Y	T	N	A	N	K	A	K	G	I	T	W	Q	D	D	T	L	F	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	L	K	K	P	N	E	R	G	D	L	I	A	Y	L	K	S	A	T	K	-
Q	G	Y	S	Y	T	D	A	N	I	K	K	N	V	L	W	D	E	N	N	M	S	E	Y	L	T	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	E	K	D	R	N	D	L	I	T	Y	L	K	K	A	C	E	-
Q	G	Y	S	Y	T	D	A	N	K	R	A	G	V	E	W	A	E	P	T	M	S	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	A	K	D	R	N	D	L	V	T	Y	M	L	E	A	S	K	-
D	G	Y	A	Y	T	D	A	N	K	Q	K	G	I	T	W	D	E	N	T	L	F	E	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	A	F	G	G	L	K	K	D	K	D	R	N	D	I	I	T	F	M	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	E	W	E	E	N	T	L	Y	D	Y	L	L	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	K	A	T	S	S
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	T	W	G	E	D	T	L	Y	E	Y	L	L	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	D	S	T	E	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	I	W	E	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	T	S	T	A	-
A	G	Y	S	Y	S	T	A	N	K	N	M	A	V	I	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	K	A	V	E	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	R	A	V	I	W	E	E	K	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	M	A	V	I	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	E	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	Q	W	G	E	N	T	L	Y	A	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	A	T	A	-
A	G	Y	S	Y	S	A	A	N	K	N	M	A	V	Q	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
P	G	Y	S	Y	S	A	A	N	K	N	M	A	V	N	W	G	E	N	T	L	Y	D	Y	L	E	N	P	X	K	Y	I	P	G	T	K	M	V	F	P	G	L	X	K	P	Q	D	R	A	D	L	I	A	Y	L	K	E	S	T	A	-
6	1	2	3	1	2	5	1	1	2	6	4	3	2	7	1	7	4	5	2	2	5	4	1	1	3	1	1	1	1	1	1	1	1	1	1	1	3	1	5	1	2	2	1	6	9	2	1	7	2	2	2	2	2	2	2	6	4	4	5	4

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

Constructing phylogenetic trees

Evolutionary relationship

Computer program to construct phylogeny

Quantitative measure of the degree of relatedness

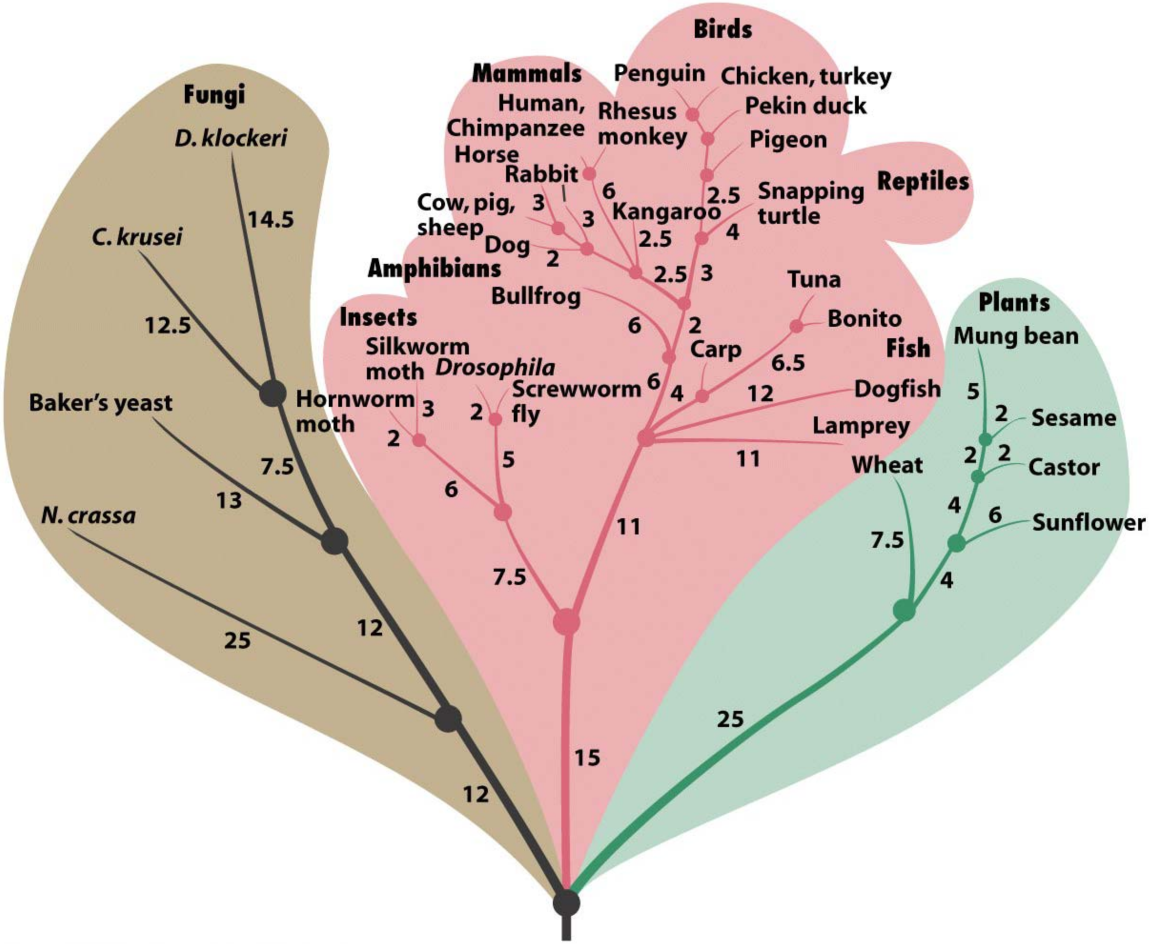


Figure 5-23 Fundamentals of Biochemistry, 2/e

Protein evolve at characteristic rates

- Protein sequence difference plotted against the time when the species diverged
  - The linear nature of mutation rate depending on proteins
  - Mutations accumulate at a constant rate over a geological time scale

- Protein sequences do not reveal the complete story of evolution
  - The proteins of human and chimpanzee are >99% identical
  - Why then so different between them?
    - Differences in regulation and expression (time and space)
    - Rearrangements and duplications of genes, giving rise to new proteins

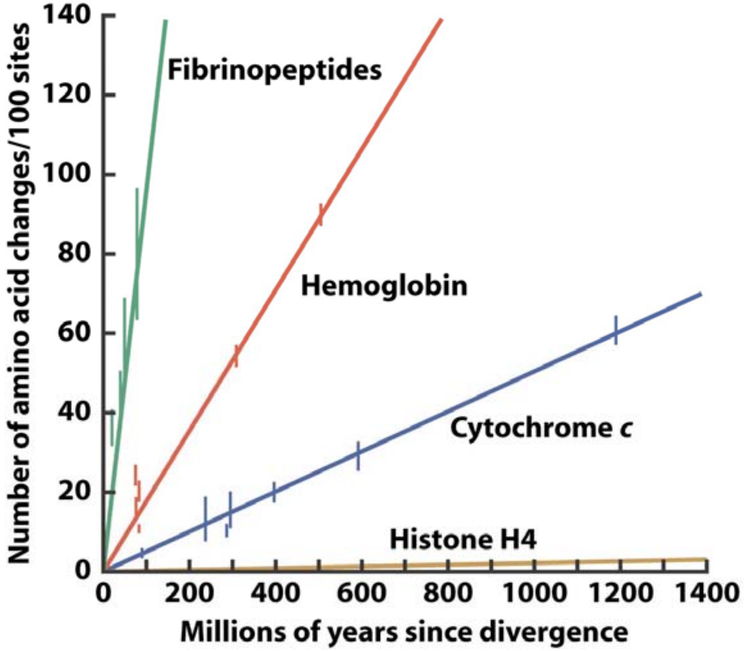


Figure 5-24 Fundamentals of Biochemistry, 2/e

Gene duplication and protein families

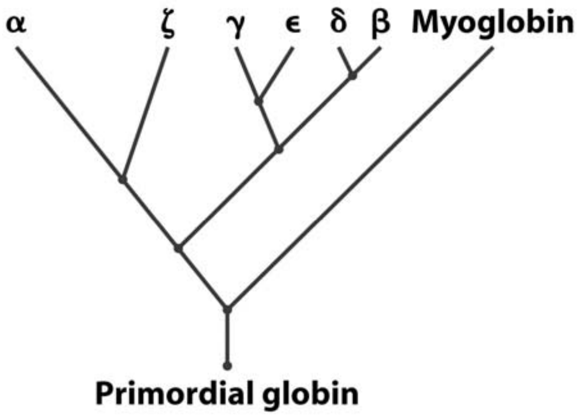
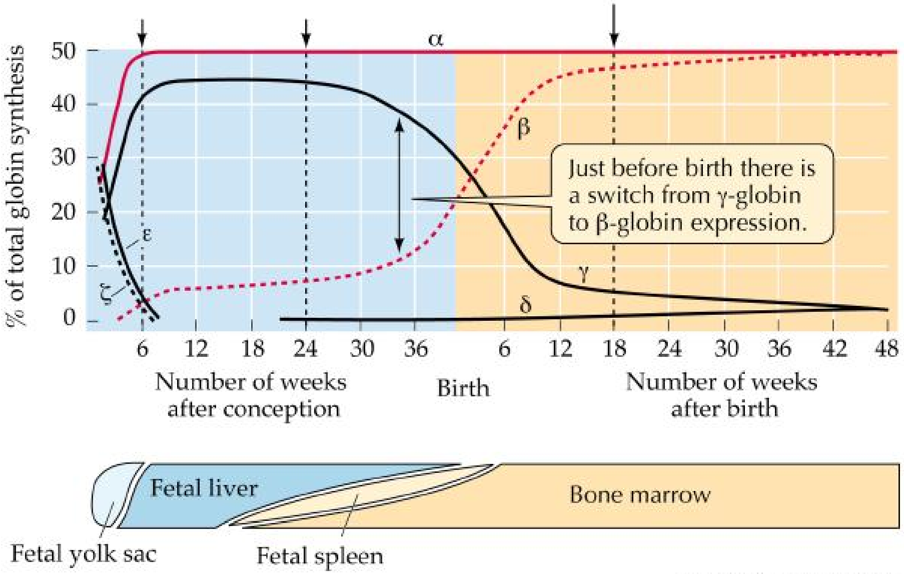


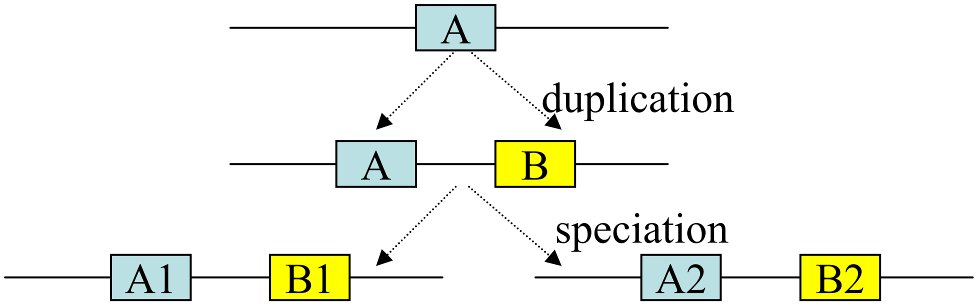
Figure 5-25 Fundamentals of Biochemistry, 2/e
© 2006 John Wiley & Sons



© 1998 Sinauer Associates, Inc.

Homologous genes: derived from a common ancestor
- orthologs: separated by speciation, similar functions and structure
- paralogs: separated by a duplication event, different but related functions
- xenologs: lateral transfer between two organisms

***pseudogenes



Protein modules

Modules of ~40-100 amino acid residues

exon shuffling

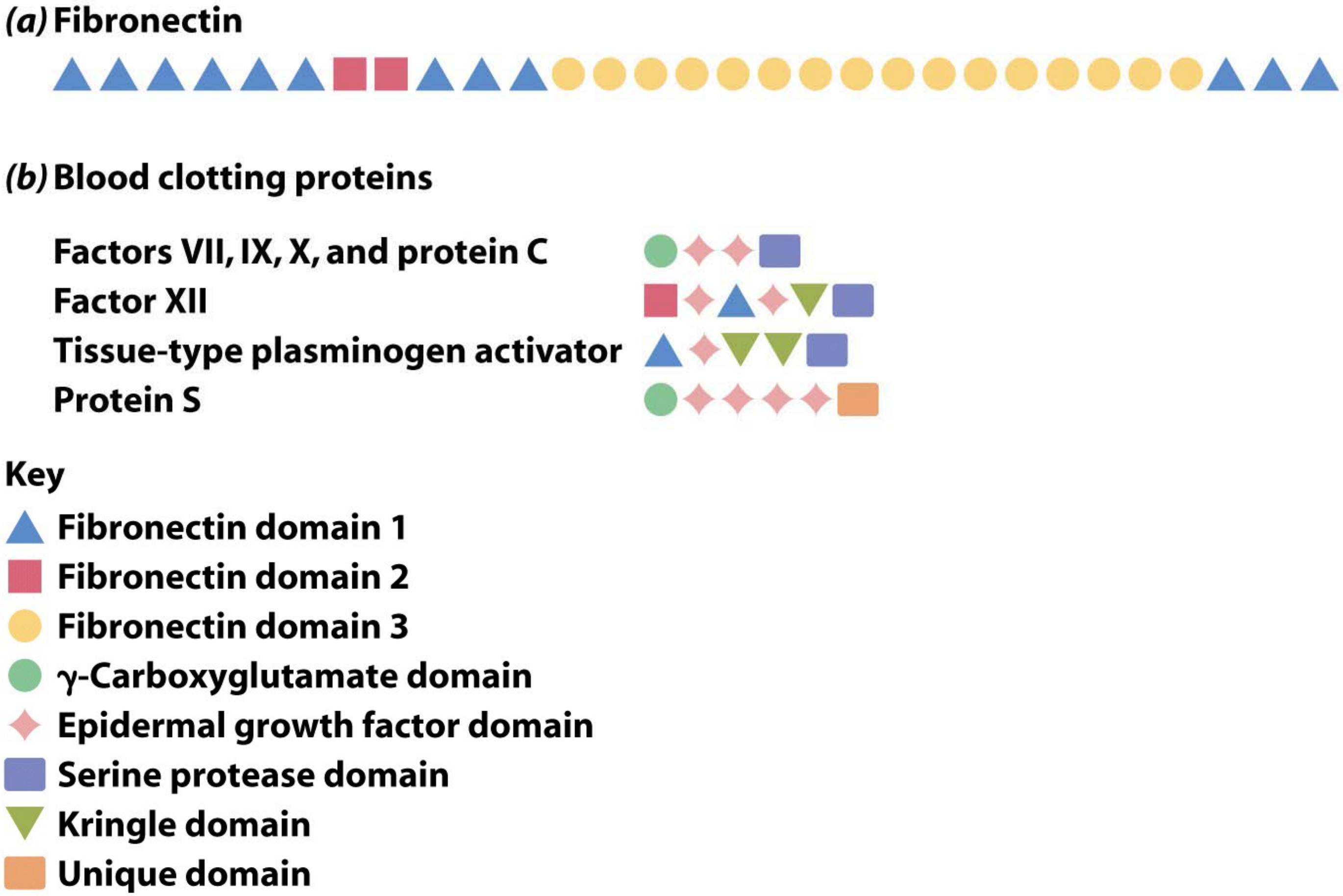


Figure 5-26 Fundamentals of Biochemistry, 2/e