Chapter 3

Exploring Proteins and Proteomes

Dr. Jaroslava Miksovska

Chapter objectives:

Protein purification and isolation techniques

Protein sequencing

Immunology for protein identification

Protein identification using mass spectrometry

Solid state peptide synthesis

Tertiary structure determination

Proteome: a set of proteins expressed in a given type of cell or organism, at a given time, under defined conditions. Changes with cell tissue development

Genome: set of genes present in the cell, unvarying characteristic of the cell

Proteome: varying characteristic of the cell (stimulus, pathological and physiological conditions tissue type, cell age).

Proteins can be post-translationally modified (hydroxylation, methylation, glycosylation, phosphorylation)

Proteins associate with other proteins forming inter-protein complexes with specific conditions

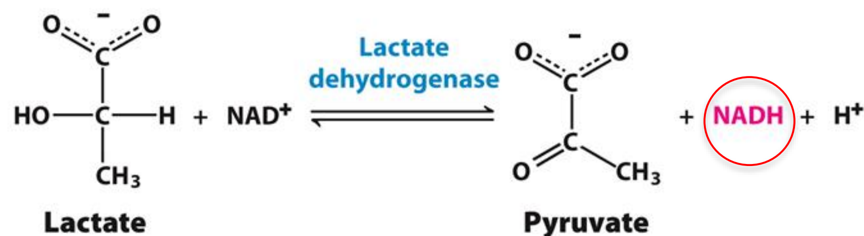
To fully understand protein function we need to characterize isolated/purified proteins

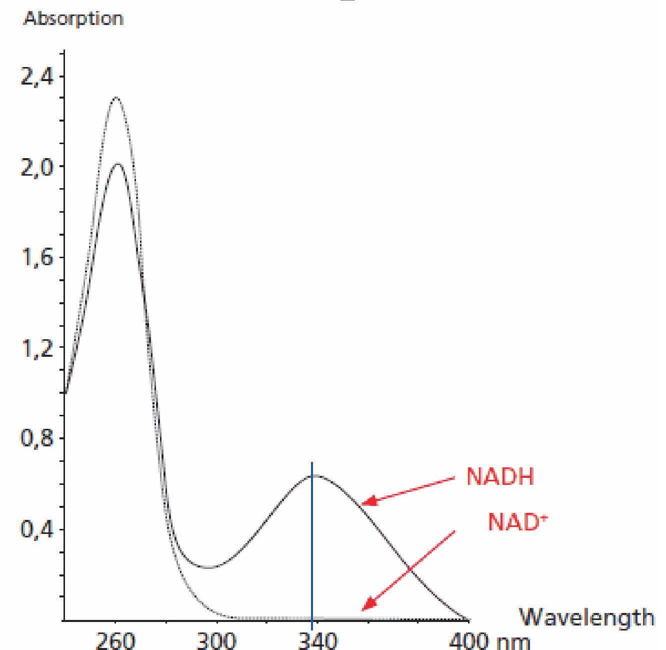
Proteins can be purified on the basis of differences in their chemical and physical properties.

Protein purification requires a test, or assay, that determines whether the protein of interest is present.

Enzymatic assays measure the change in concentration of product substrate of byproduct .

Enzymatic essay for lactate dehydrogenase





Protein purification steps (soluble proteins)

1. Homogenization of cell or tissue: ultrasound; French press, beads, homogenizer….
2. Fractionalization (removal of membranes organelles etc). Differential centrifugation

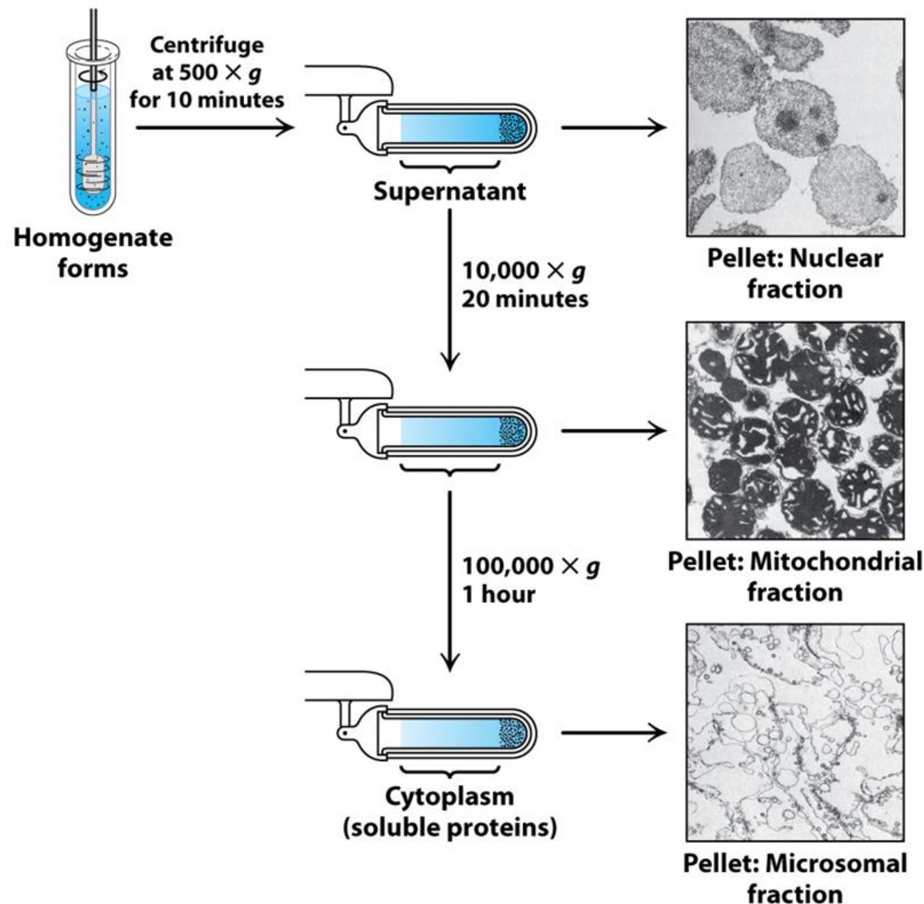


Figure 3.1
Biochemistry, Eighth Edition
© 2015 Macmillan Education [Photographs courtesy of Dr. S. Fleischer and Dr. B. Fleischer]

The fractions are essayed for its activity and usually only one fraction has desirably activity

Protein purification

Proteins are separated and purified based on their size, charge, solubility and specific binding affinity .

Salting out:
solubility of protein varies with salt concentration. Different proteins precipitate at different salt concentration.
For example: fibrinogen precipitates at 0.8 M ammonium sulfate
serum albumin precipitate at 2.4 M ammonium sulfate

Dialysis:
separation of proteins from small molecules (salt, buffer exchange). The smaller molecules migrate into surrounding solution.

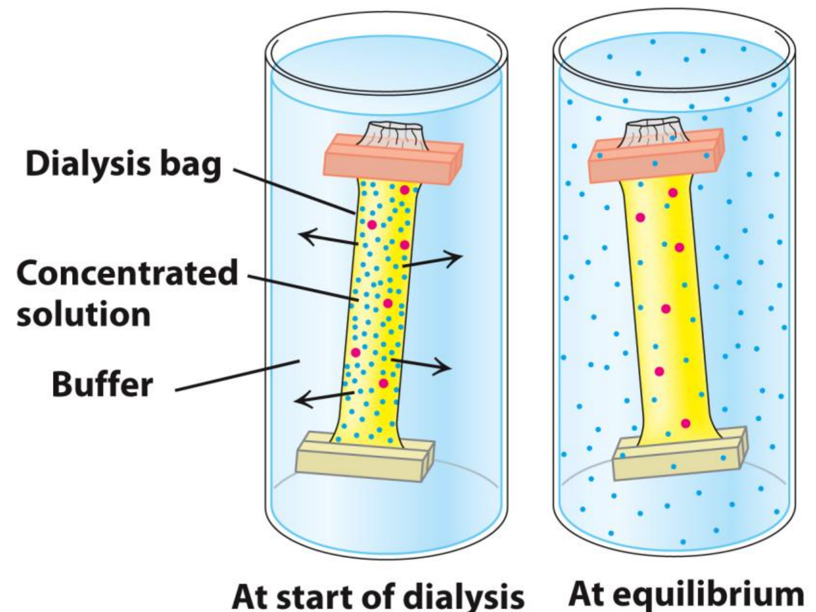


Figure 3.2
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Size exclusion (gel filtration) chromatography

Separation based on the size
Column of porous beads (dextran, agarose, acrylamide)
Large molecule migrates faster than small molecules which penetrates into the beads.

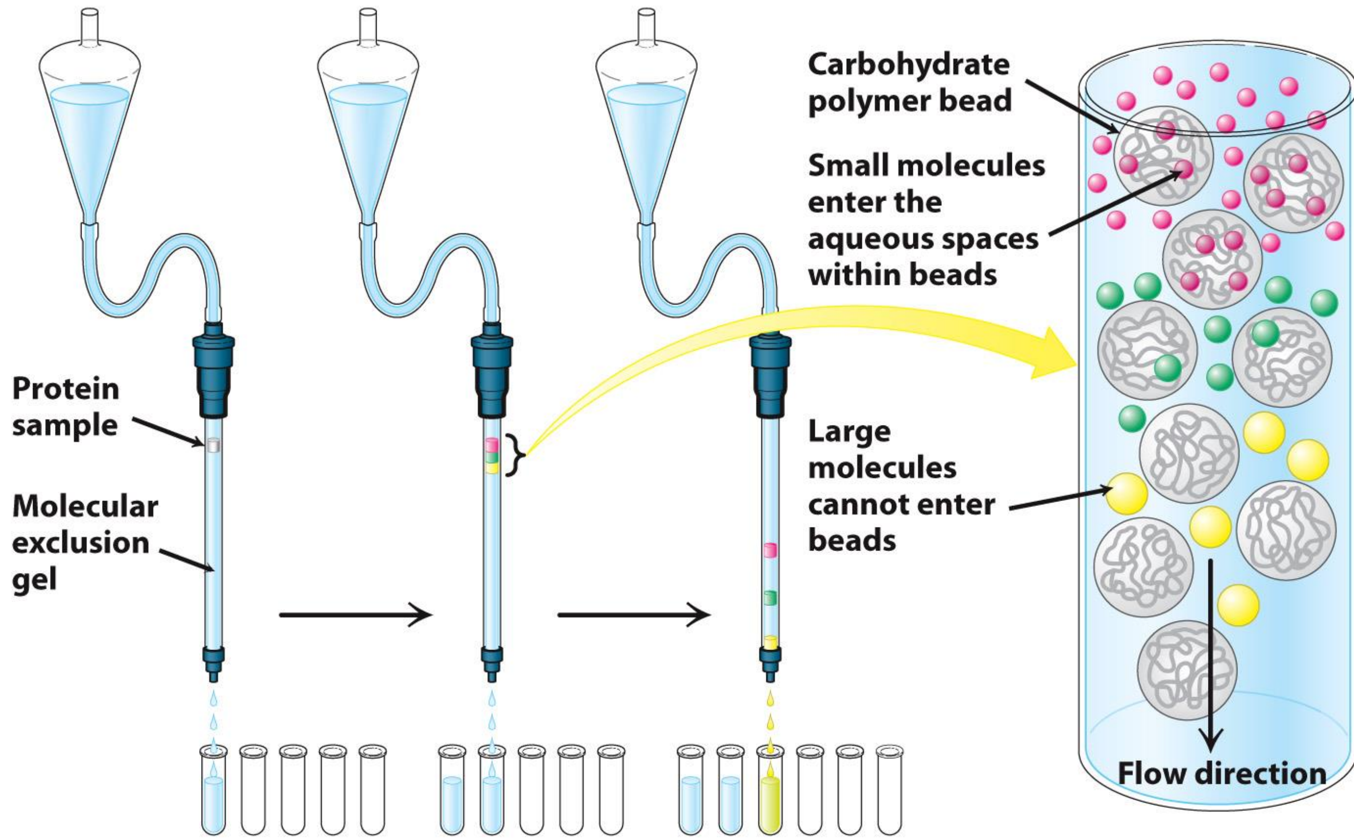


Figure 3.3
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Ion exchange chromatography

Proteins are separated on the bases of the net charge

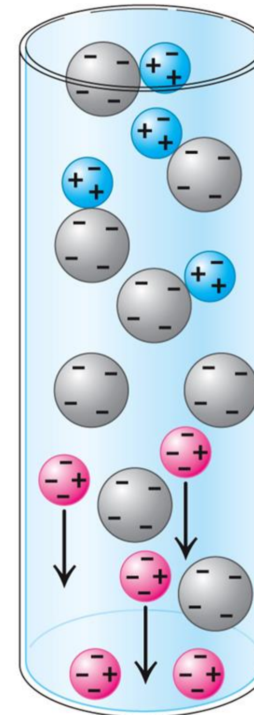
Charged molecules bind to the oppositely charged groups that were immobilized on matrix.

Anion exchanger: Negatively charged proteins bind to cationic groups on the matrix

Cation exchanger: positively charged molecules bind to anionic groups on the matrix.

The binding affinity of proteins depends on **pH**, and presence of other ions.

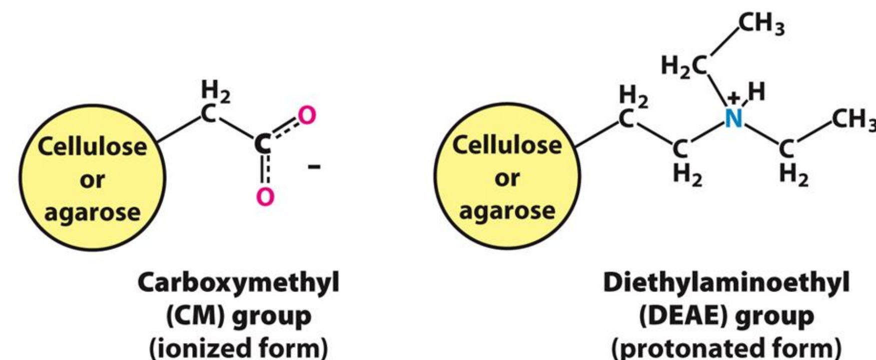
Tightly bind proteins are **eluted** from column using **eluent** - buffer with higher salt concentration of pH that decreases the affinity.



Positively charged protein binds to negatively charged bead

Negatively charged protein flows through

Figure 3.4
Biochemistry, Eighth Edition
© 2015 Macmillan Education



Unnumbered 3 p70
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Affinity chromatography

Many proteins bind to specific ligands with a very high affinity

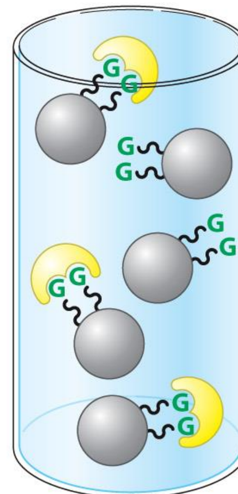
A molecule (a ligand) that specifically binds to a protein is covalently attached to an inert matrix.

Mixture of proteins pass through such column and only proteins with high affinity for a ligand binds to the matrix whereas other proteins are washed through the column with the buffer.

Affinity chromatography has a high separation power (does not separate proteins based on small differences in size of charge).

Protein is released by adding the excess of the ligand

Glucose-binding protein attaches to glucose residues (G) on beads



Addition of glucose (G)

Glucose-binding proteins are released on addition of glucose

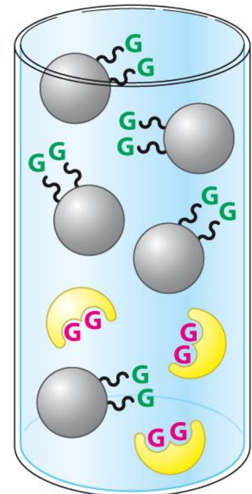


Figure 3.5
Biochemistry, Eighth Edition
© 2015 Macmillan Education

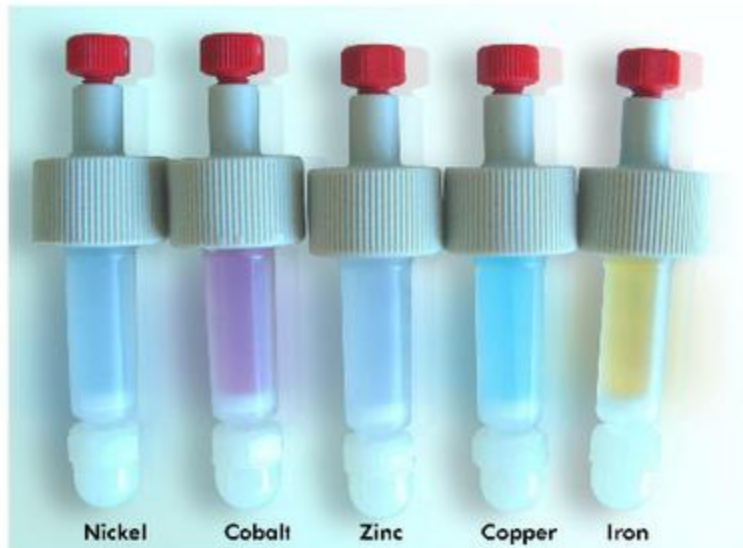
Immobilized metal ion affinity chromatography

Immobilized metal ion affinity chromatography (IMAC) is based on the specific interaction of certain amino acids with the metals, especially histidine.

Protein is modified by a covalent attachment of histidine residues

Nickel, Cobat or Copper are used for purifying His-Tagged proteins and phosphorylated proteins are separated out using iron or zinc metal ions.

Chelators most widely used for IMAC are **Nitrilotriacetic acid (NTA)** or **Iminodiacetic acid (IDA)**.



Hydrophobic Interaction Chromatography

Hydrophobic interactions between the chromatographic matrix and proteins
Matrix is lightly substituted with phenyl groups .

High Performance Liquid Chromatography

Column is made from a fine beads in metal column.

Very fine beads allow more interactions and thus greater resolving power, but flow rates through such columns are too slow.

Adequate flow rates are achieved by applying pressure.

Increased speed and resolution

Protein is detected by monitoring absorbance (280 nm or 220 nm)

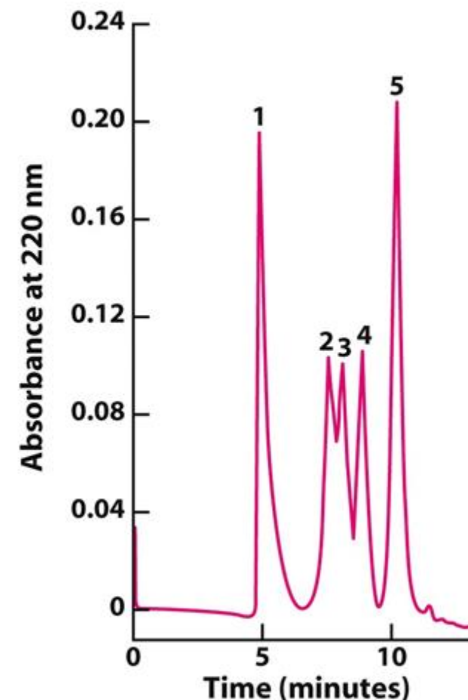
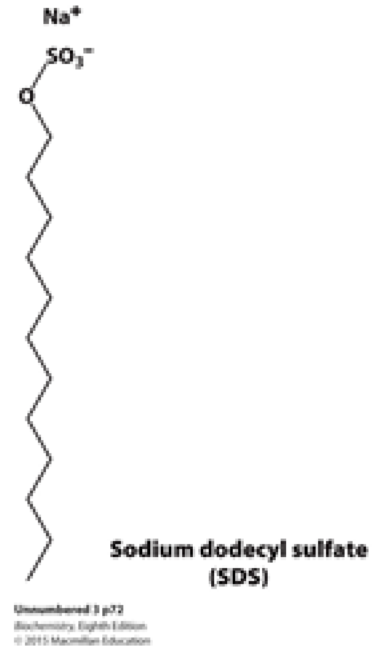


Figure 3.6
Biochemistry, Seventh Edition
© 2012 W. H. Freeman and Company

Gel electrophoresis

Migration of proteins in an electrical field.
Velocity of proteins (v) is proportional to:
electric field strength (E),
the charge on the protein (z)
frictional coefficient (f).

$$v = Ez/f$$



frictional coefficient is a function of
the radius (r) that depends on mass
and shape of the protein
density of the medium (η).

$$f = 6\pi\eta r$$

When the migration occurs in a gel, it is called gel electrophoresis.

Sodium dodecyl sulfate-**p**oly**a**crylamide **g**el **e**lectrophoresis (SDS-PAGE)
accurate determination of mass.
SDS denatures proteins, and for most proteins, 1 molecule of SDS binds for every two amino acids. Thus, proteins have the same charge to mass ratio and migrate in the gel on the basis of mass only in the direction of the anode.

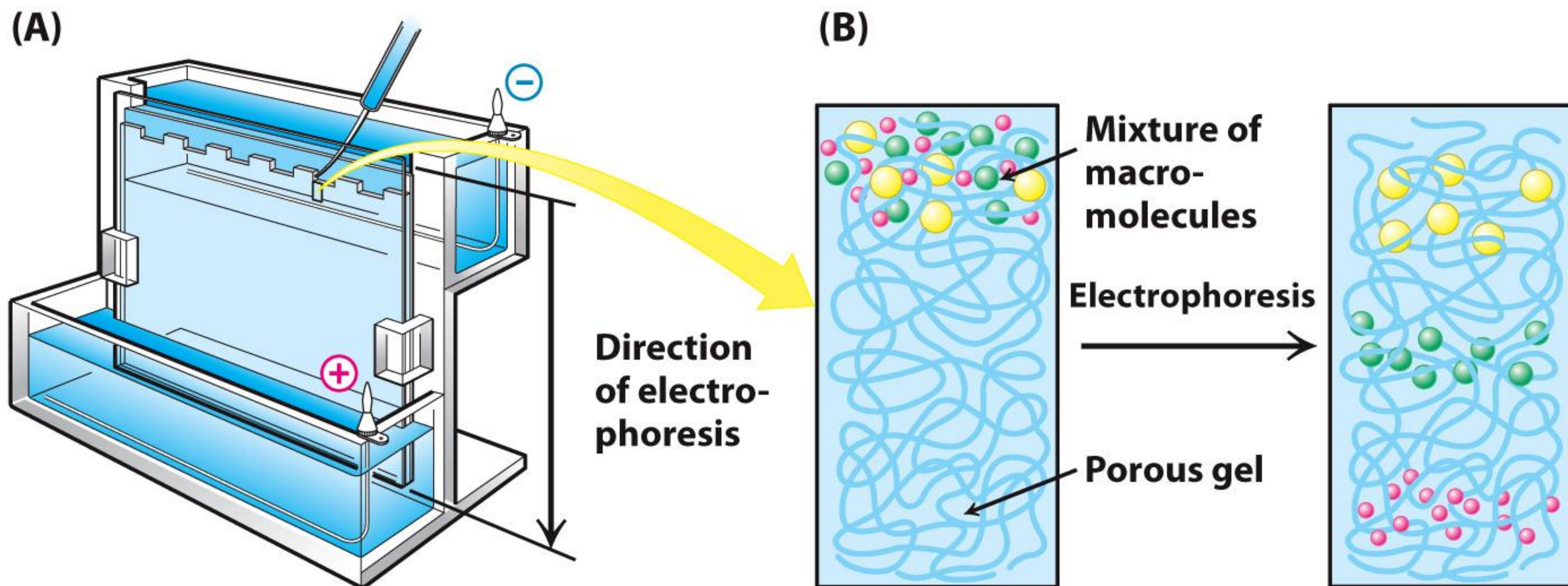


Figure 3.7
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Example of SDS page electrophoresis

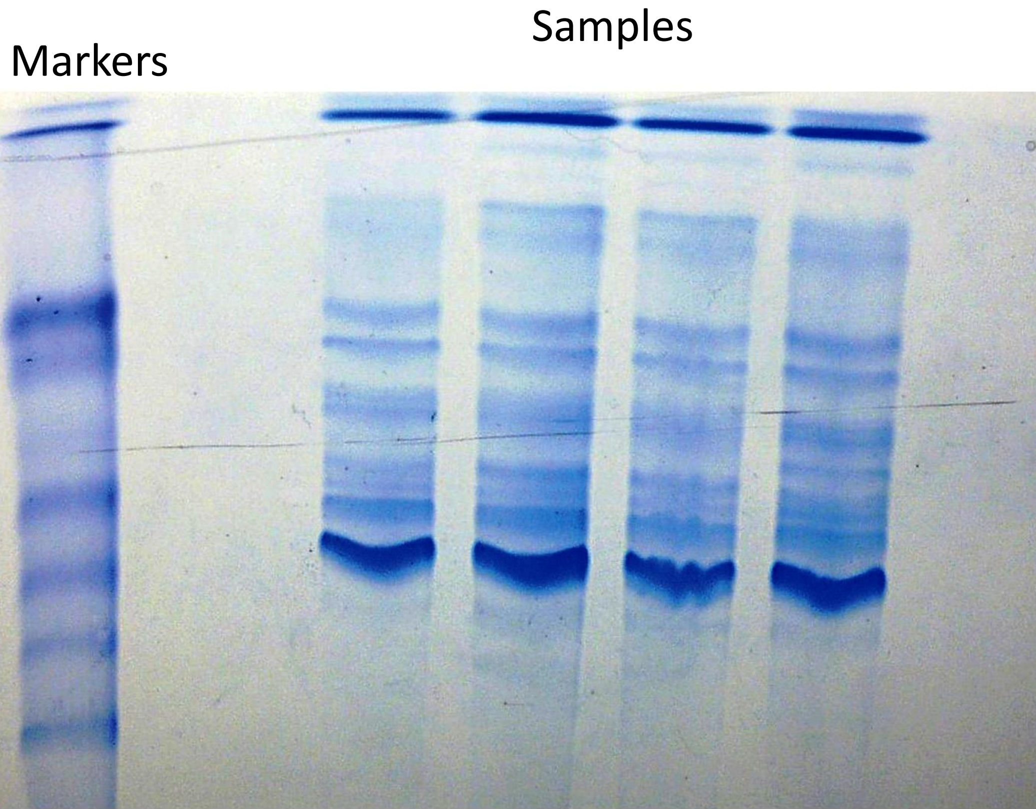


Figure 3.9
Biochemistry, Eighth Edition
© Dr. Robert Farrell

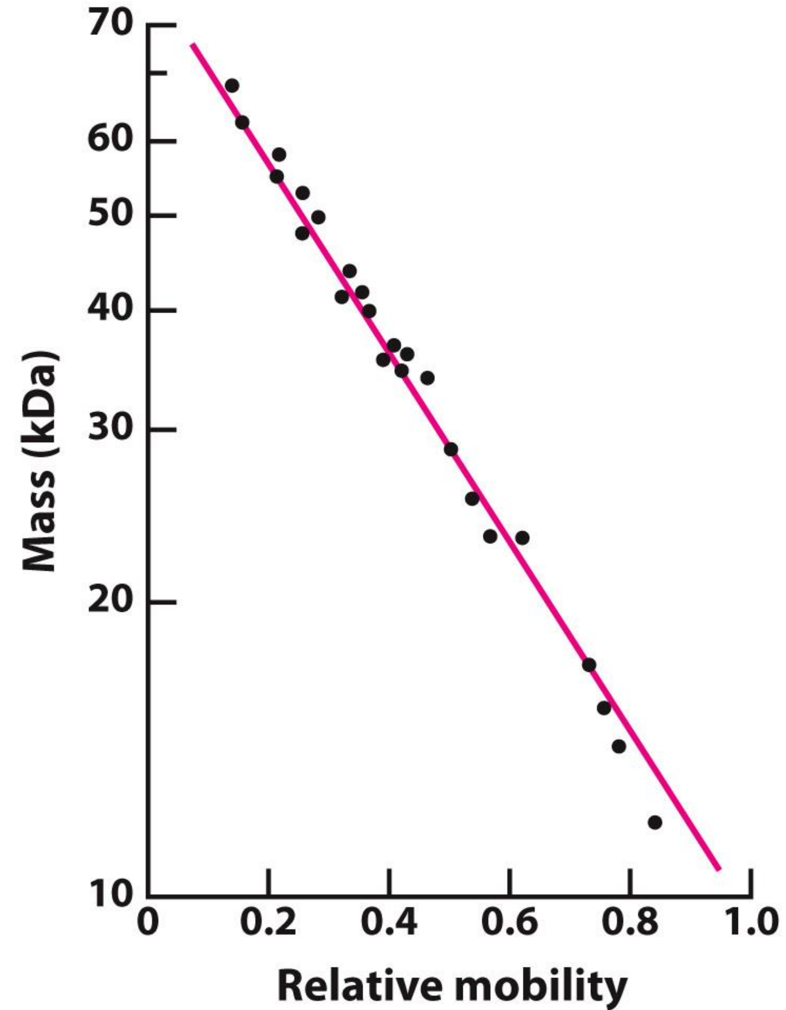


Figure 3.10
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Protein visualization:
Coomassie blue , silver or radioactive labels incorporated into proteins

Isoelectric focusing

Net protein charge depends on pH

Isoelectric point (pI) is a pH at which a molecule **carries a zero electric charge**. Proteins will be positively charged below their pI value and negatively charged above their pI value.

Proteins are separated based on in a gel on the basis of their relative amounts of acidic and basic amino acids. If a mixture of proteins is placed in a gel with a pH gradient and an electrical field is applied, proteins will migrate to their isoelectric point, the pH at which they have no net charge.

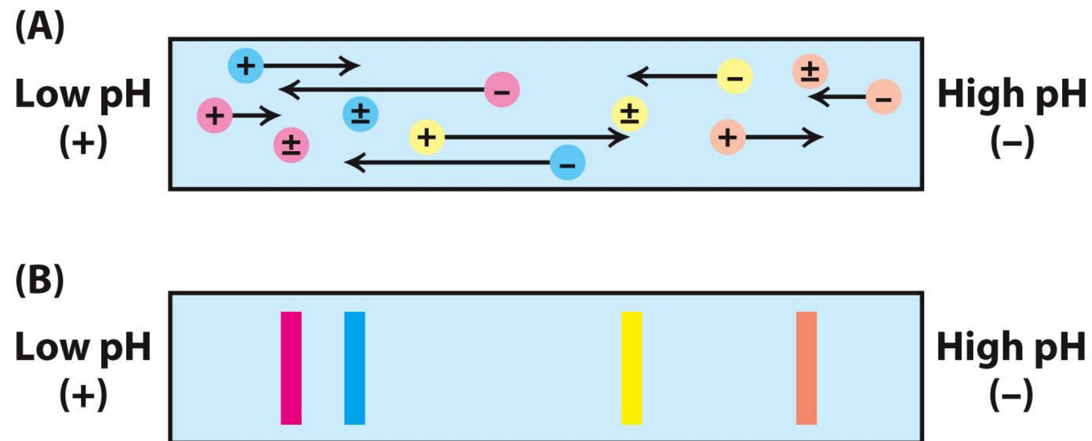


Figure 3.11
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Proteins are separated in narrow bands

Two dimensional electrophoresis

In two-dimensional gel electrophoresis, proteins are separated in one direction by isoelectric focusing. This gel is then attached to an SDS-PAGE gel and electrophoresis is performed at a 90° angle to the direction of the isoelectric focusing separation.

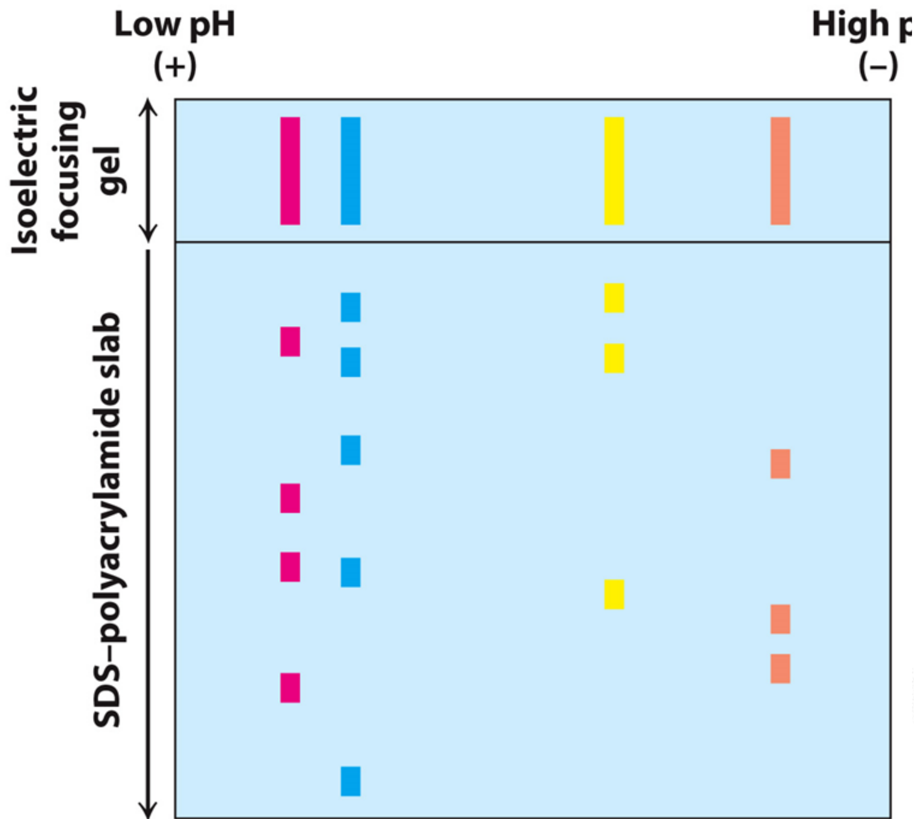


Figure 3.12a
Biochemistry, Eighth Edition
© 2015 Macmillan Education

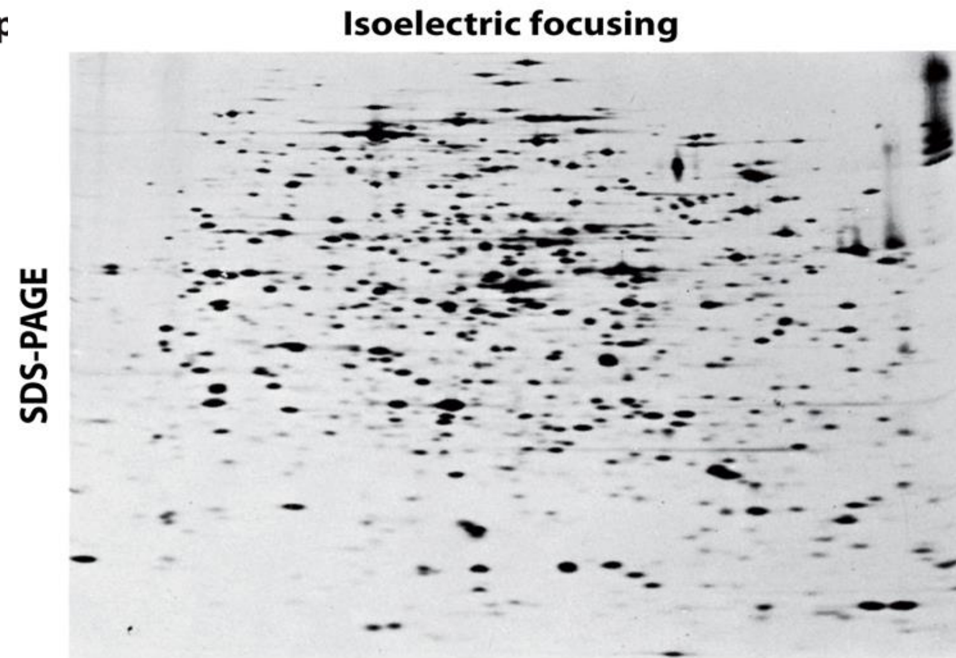
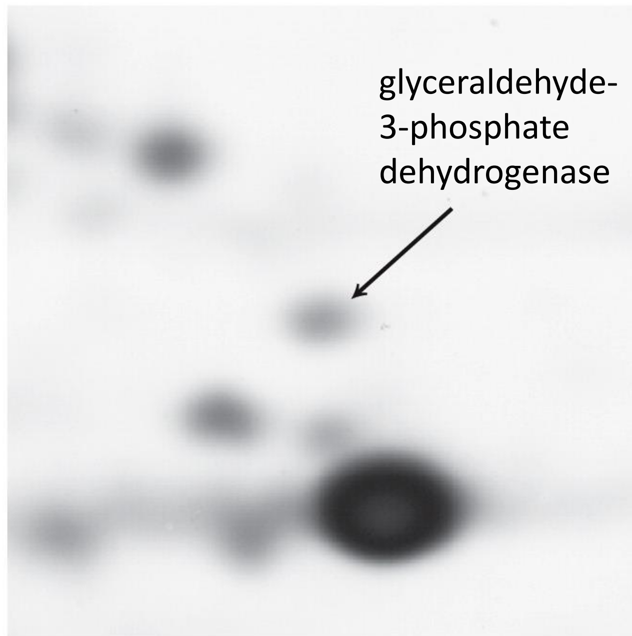


Figure 3.12b
Biochemistry, Eighth Edition
Courtesy of Dr. Patrick H. O'Farrell

Up to 5000 proteins can be separated
Great tool for proteomics
Protein can be extracted and further characterized by MS



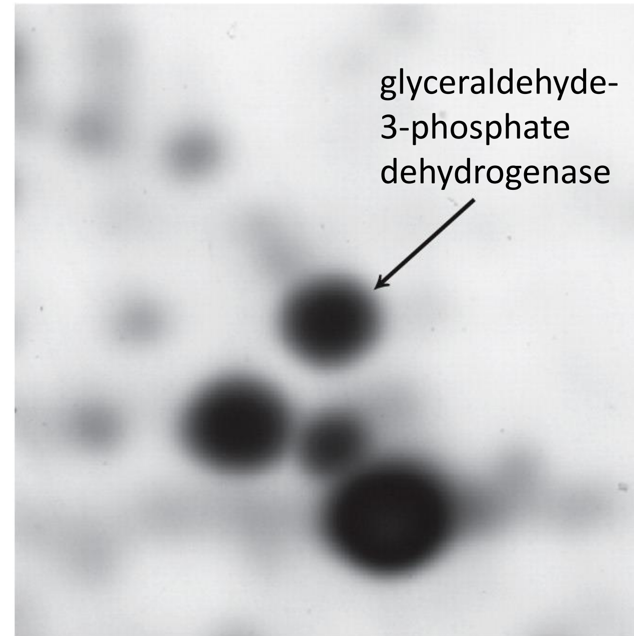


Figure 3.13
Biochemistry, Eighth Edition
Courtesy of Lin Quinsong © 2010, The American Society for Biochemistry and Molecular Biology

Samples of normal colon mucosa and colorectal tumor tissue

The effectiveness of a purification scheme is measured by calculating the specific activity after each separation technique.

Specific activity is the ratio of enzyme activity to protein concentration. Specific activity should increase with each step of the purification procedure.

SDS-PAGE allows a visual evaluation of the purification scheme.

TABLE 3.1 Quantification of a purification protocol for a fictitious protein

Step	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Yield (%)	Purification level
Homogenization	15,000	150,000	10	100	1
Salt fractionation	4,600	138,000	30	92	3
Ion-exchange chromatography	1,278	115,500	90	77	9
Gel-filtration chromatography	68.8	75,000	1,100	50	110
Affinity chromatography	1.75	52,500	30,000	35	3,000

Table 3.1
Biochemistry, Eighth Edition
© 2015 Macmillan Education

The effectiveness of the purification scheme can be probed by SDS PAGE

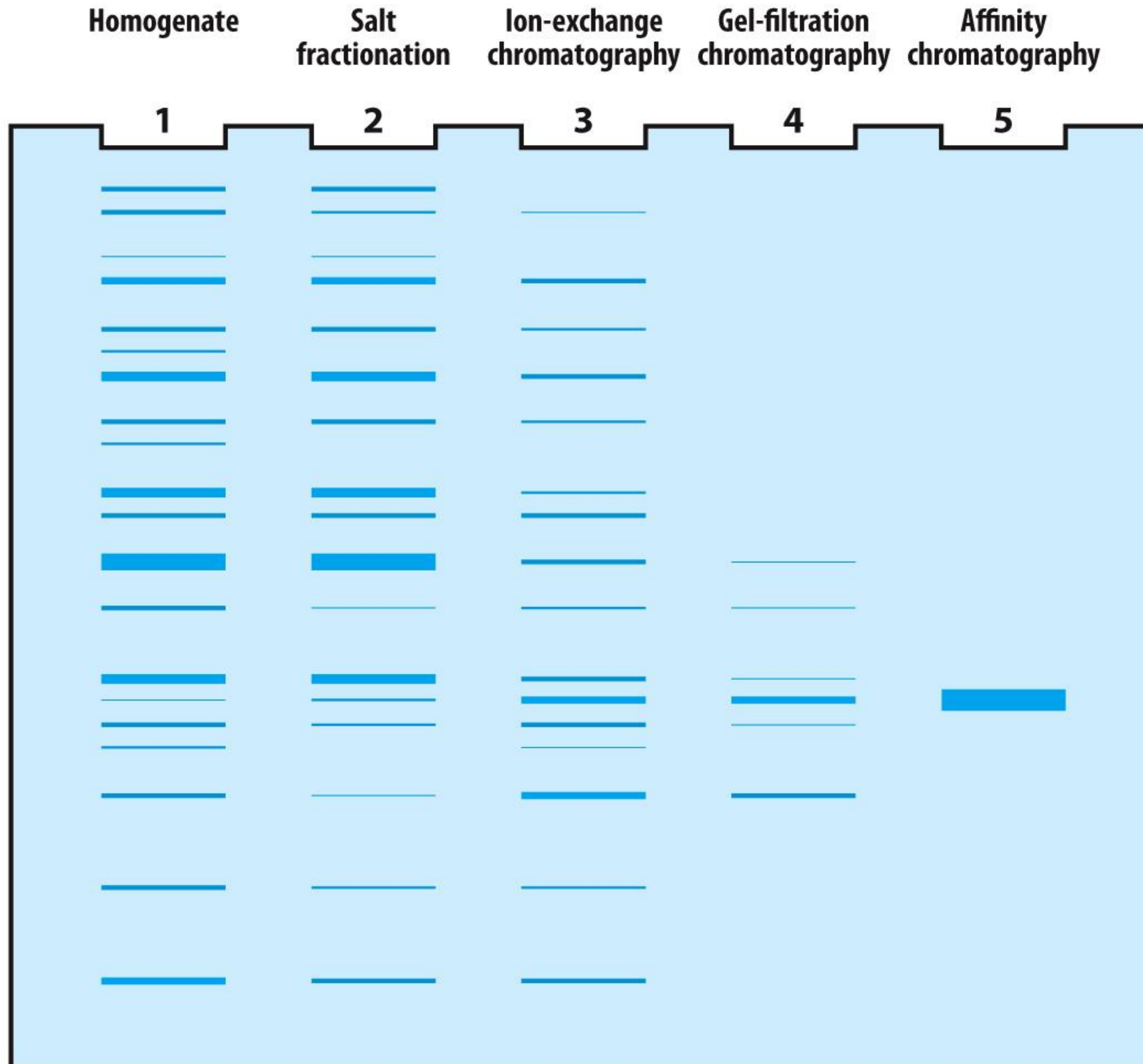


Figure 3.14
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Ultracentrifugation

Macromolecules when subjected to accelerations (centrifugal force) sediment. The rate of sedimentation in an ultracentrifuge is related to the mass

The sedimentation coefficient (sedimentation velocity per unit of centrifugal force) can be expressed as:

$$s = m(1 - \overline{\nu}\rho)/f$$

m =mass,
$\overline{\upsilon}$ = the partial specific volume (or one over the mass of the particle,
ρ =density of the medium,
f =the frictional coefficient of the particle.

Sedimentation coefficient of proteins is usually expressed in Svedberg units (S), equal to 10^{-13} s

Zonal ultracentrifugation

Centrifugation is carried out in a solution in which the concentration of inert substance increases from the top to the bottom

Use of the density gradient increases the resolving power of centrifugation by preventing convective flow.

The macromolecules move through the gradient based on their buoyant density .

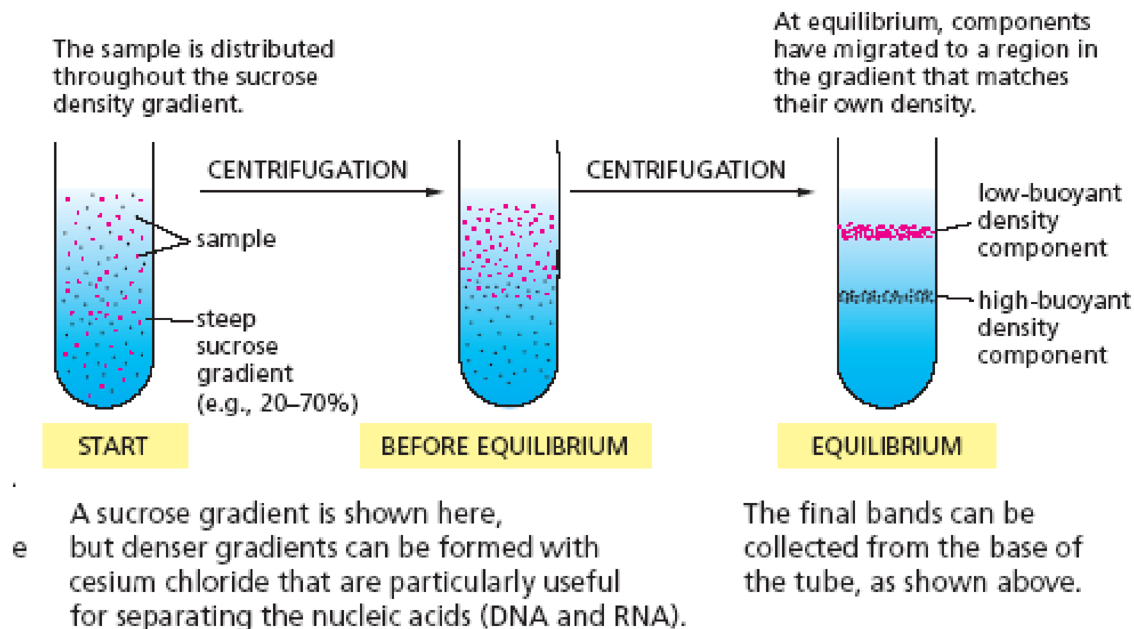


TABLE 3.2 **S values and molecular weights of sample proteins**

Protein	S value (Svedberg units)	Molecular weight
Pancreatic trypsin inhibitor	1	6,520
Cytochrome *c*	1.83	12,310
Ribonuclease A	1.78	13,690
Myoglobin	1.97	17,800
Trypsin	2.5	23,200
Carbonic anhydrase	3.23	28,800
Concanavalin A	3.8	51,260
Malate dehydrogenase	5.76	74,900
Lactate dehydrogenase	7.54	146,200

Source: T. Creighton, *Proteins,* 2d ed. (W. H. Freeman and Company, 1993), Table 7.1.

Table 3.2
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Recombinant proteins

1. Large quantities of proteins can be obtained.

2. Proteins can be modified with affinity tags that allow purification of the protein or visualization of the protein in the cell.

3. Proteins with modified primary structure can be generated. Allows for determination of the role of individual amino acid residues in enzymatic reations

Immunology

Important tools for protein identification and purification

Antibody: a protein synthesized in response to the presence of a foreign substance called an **antigen**.

The antibody recognizes a particular structural feature on the antigen called the antigenic determinant or **epitope**.

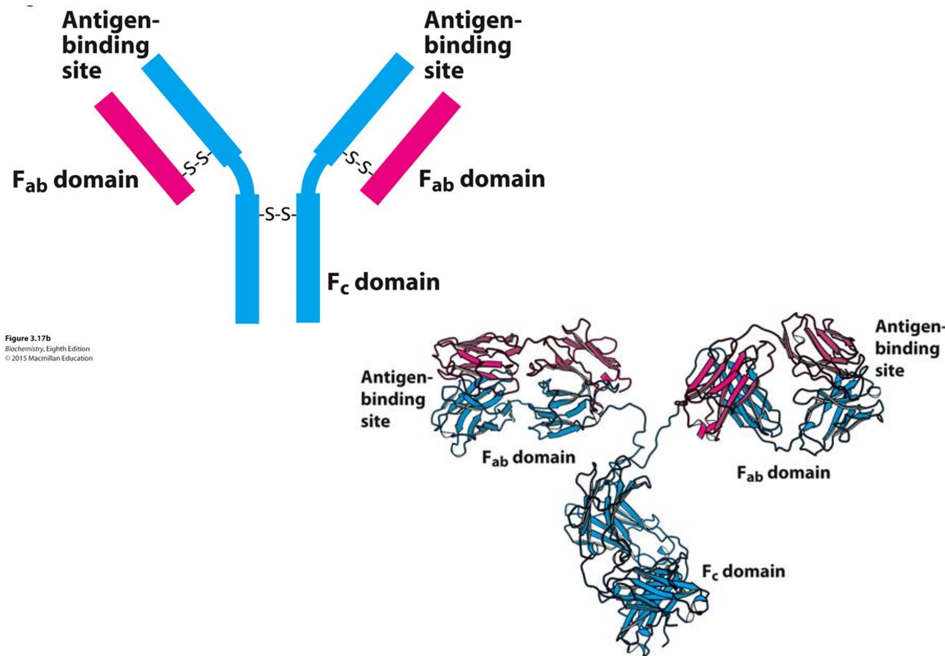


Figure 3.17b
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Figure 3.17a
Biochemistry, Eighth Edition
© 2015 Macmillan Education

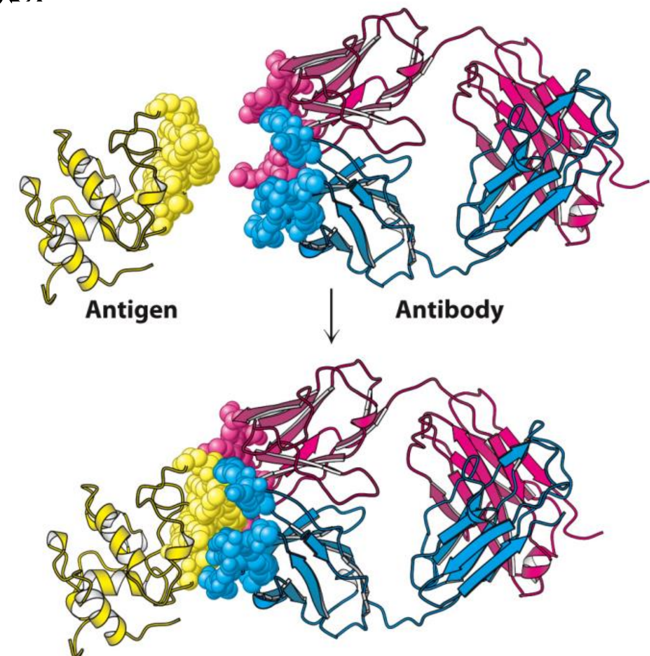


Figure 3.18
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Monoclonal and Polyclonal Antibodies

Monoclonal antibodies :

identical

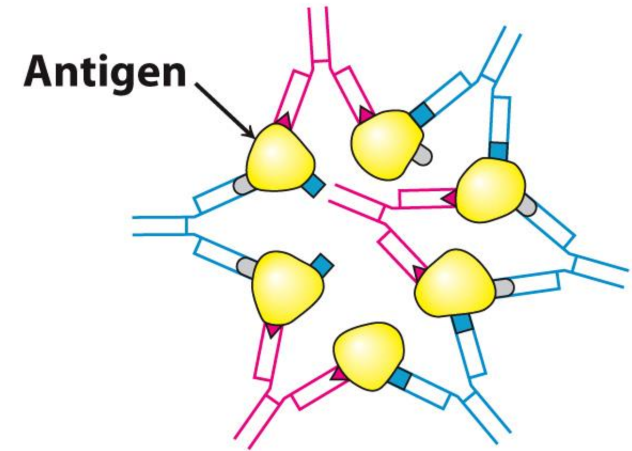
produced by a clone of the cell derived from a single anti-body producing cell

recognize one specific epitope

Polyclonal antibodies:

derived from multiple antibody producing cells populations

they recognize different surface feature of the same antigen



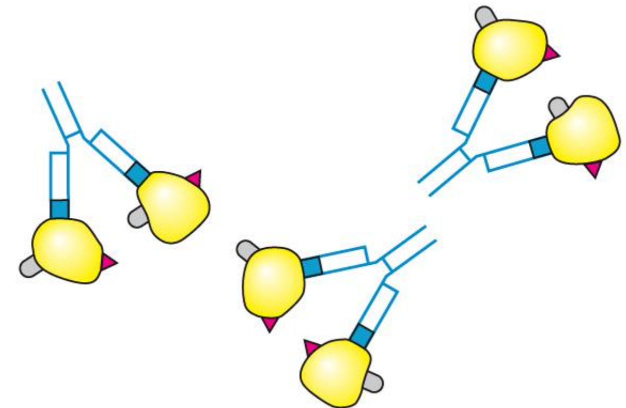


Figure 3.19
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Production of monoclonal antibodies

Antibody producing cells spleen cells have a short life spans.

Myeloma cells are imortal cells. They are derived from multiple myeloma – malignant disorder of antibody producing cells. These cells produce large amount of immunoglobulin of one type.

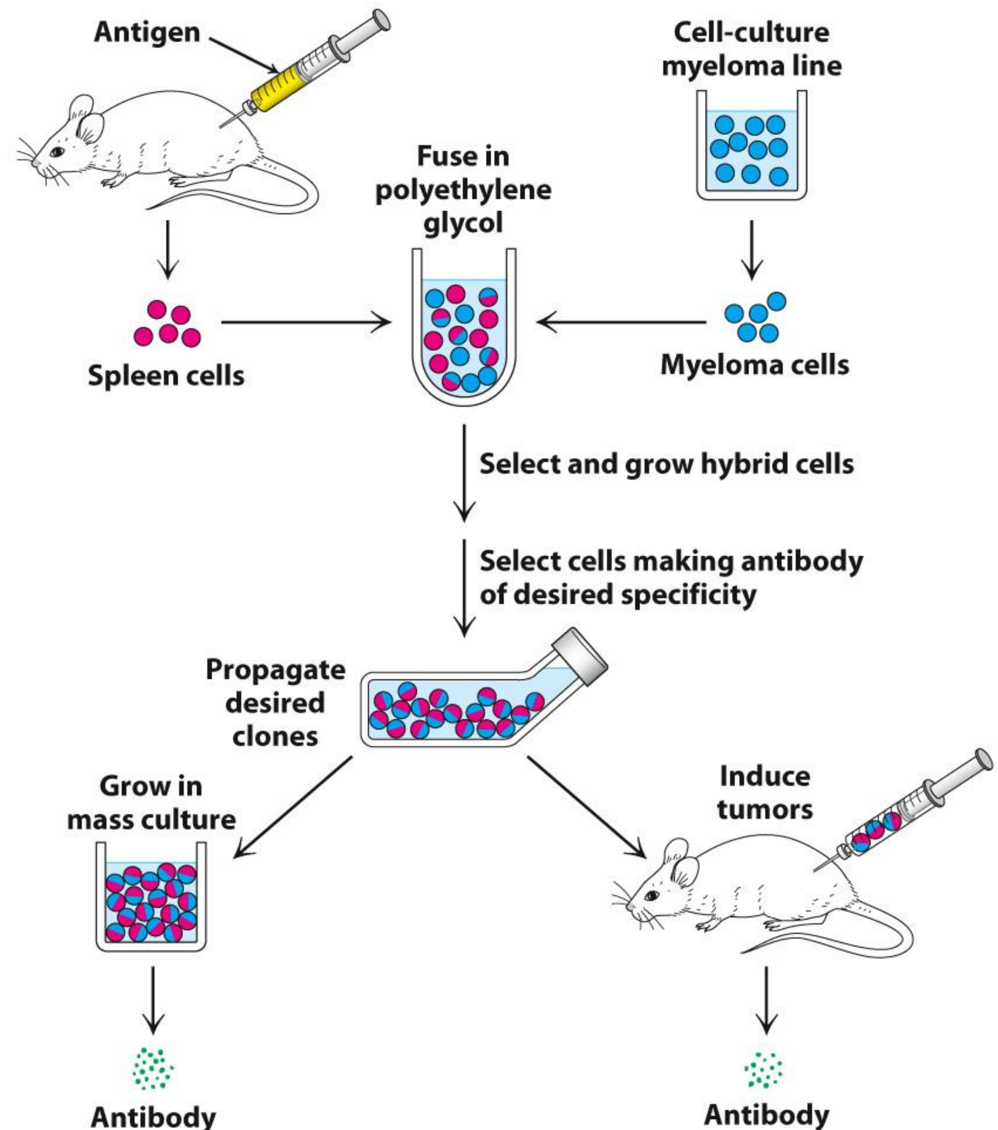
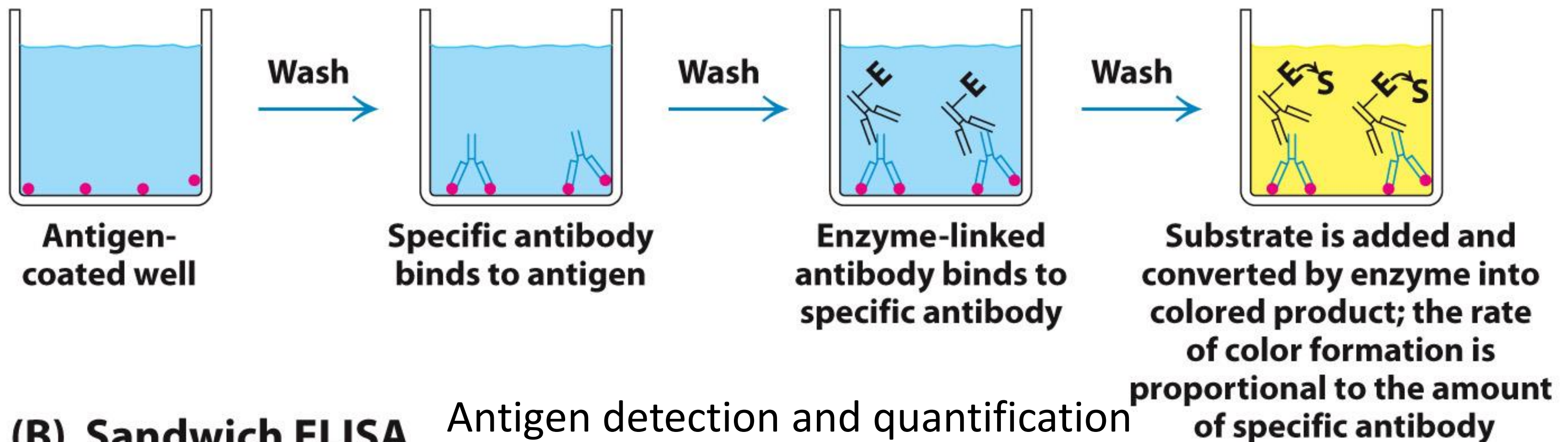


Figure 3.20
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Application of antibodies

1. Protein quantification

(A) Indirect ELISA Antibody quantification



(B) Sandwich ELISA Antigen detection and quantification

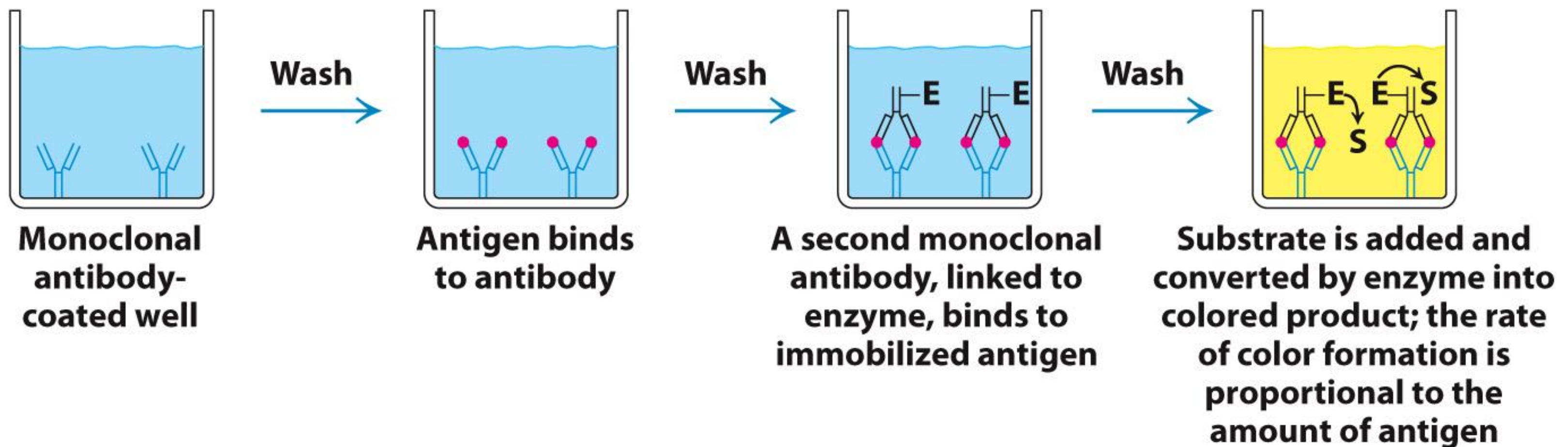


Figure 3.22
Biochemistry, Eighth Edition

Western blotting

detection of protein(s) in mixture

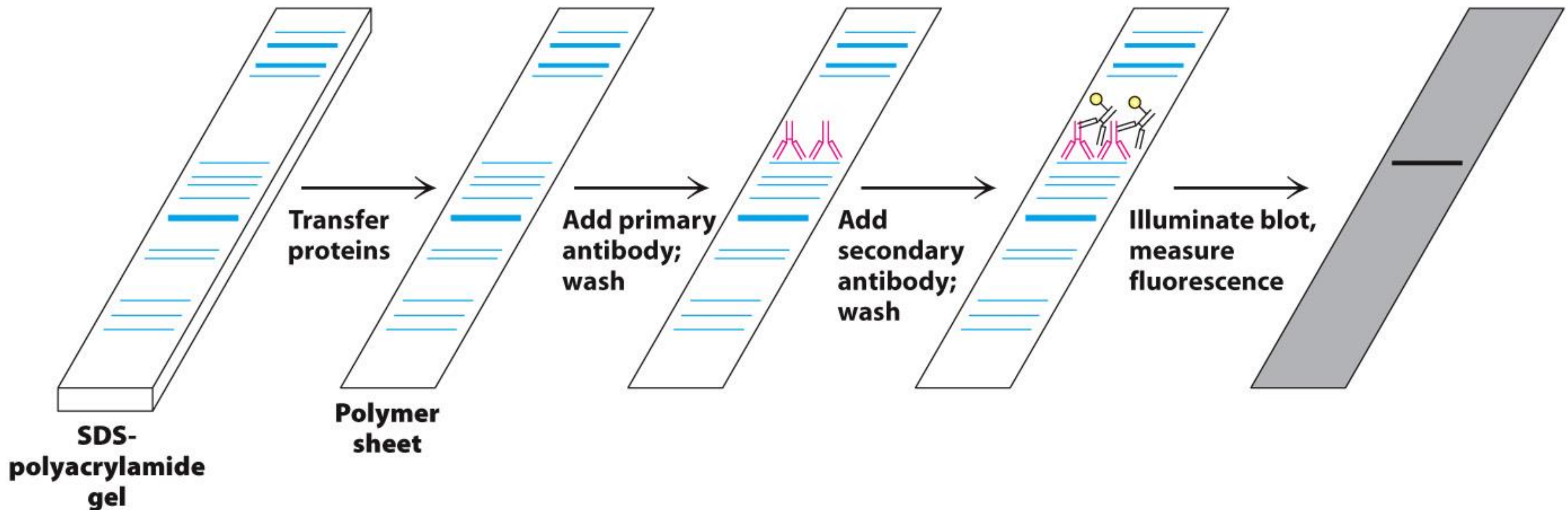


Figure 3.23
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Detection of a very small quantity of the protein

Enable identification of protein in a complex mixture

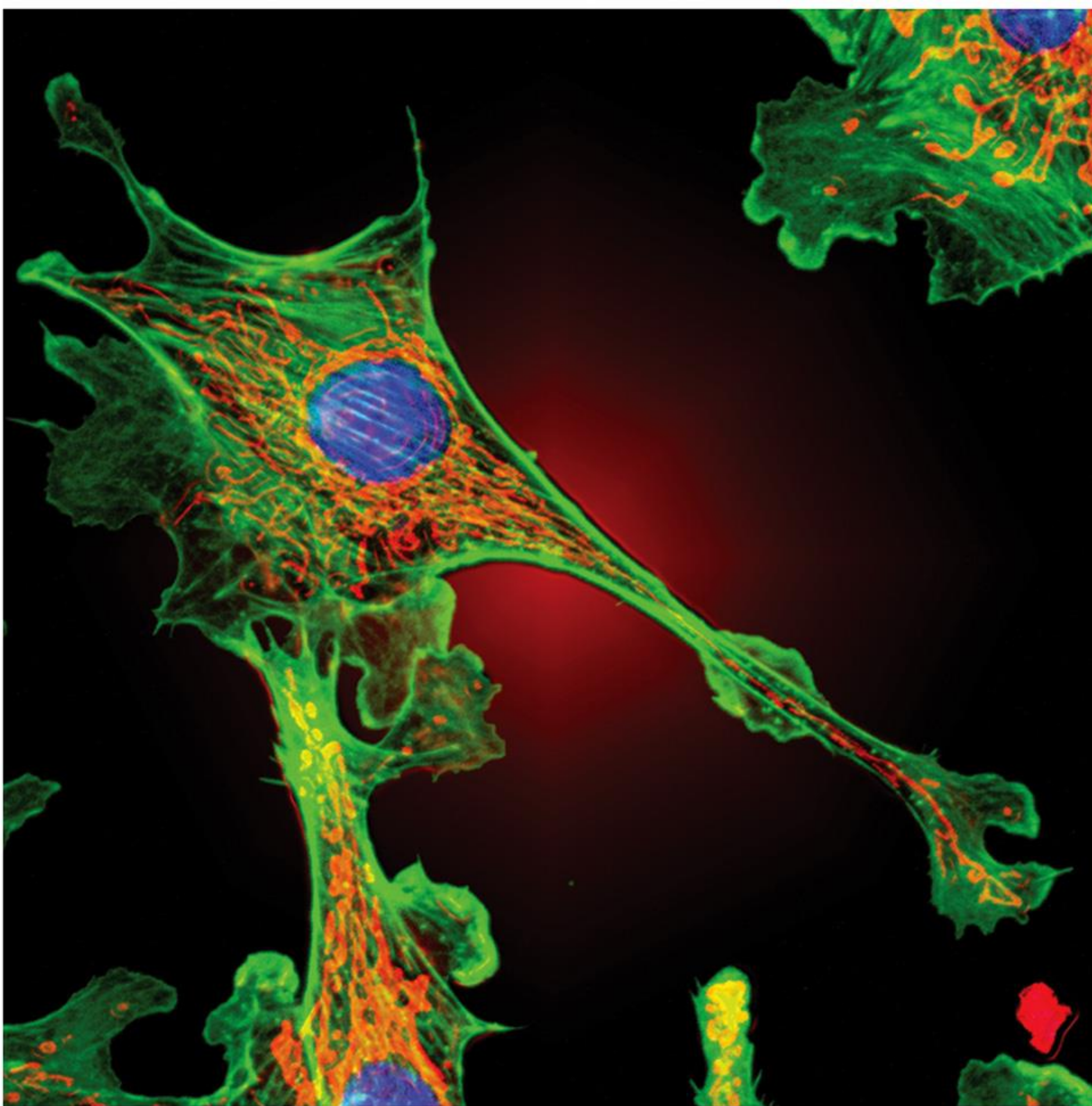


Figure 3.24
Biochemistry, Eighth Edition
David Becker/Science Source

Antibodies with attached fluorescenct probes for identifications of proteins in vivo

Fluorescence micrograph of a cell shows actin filaments stained green using an antibody specific to actin

Mass spectrometry for identification of proteins and proteomics

Mass spectrometry allows the highly accurate and sensitive detection of the mass of the molecule of interest, or analyte.

Mass spectrometers convert the analyte into gas-phase ions. The mass-to-charge ratio (m/z) can be determined.

Mass spectrometry became a method of choice for peptide/ protein sequencing

Mass spectrometers consist of three components: an ion source, a mass analyzer and a detector.

Analyte has to be converted into gas-phase ions (ionization).

MALDI (matrix assisted laser desorption ionization)

Soft ionization technique

Protein is mixed with a matrix (3,5-dimethoxy-4-hydroxycinnamic acid) and solvent, spotted on a metal plate. Upon solvent vaporization protein co-crystalizes with the matrix.
The laser energy is absorbed by the matrix leading to matrix desorption and ionization. The matrix is then thought to transfer protons to the analyte molecules (e.g., protein molecules), thus charging the analyte

An ion observed after this process will consist of the initial neutral molecule [M] with ions added or removed. $[M+H]^+$ in the case of an added proton, $[M+Na]^+$ in the case of an added sodium ion, or $[M-H]^-$ in the case of a removed proton. MALDI is capable of creating singly charged ions or multiply charged ions ($[M+nH]^{n+}$) depending on the nature of the matrix, the laser intensity, and/or the voltage used.

Mass analyzer:

Analytes ions are separated based on their mass/charge ratio (m/z)

Time of flight mass analyzer (TOF):

Ions are accelerated through an elongated chamber with a fixed electric potential

For ions with the same charge the smaller ions moves faster than larger ions

The mass can be determined based on the time required to reach the detector

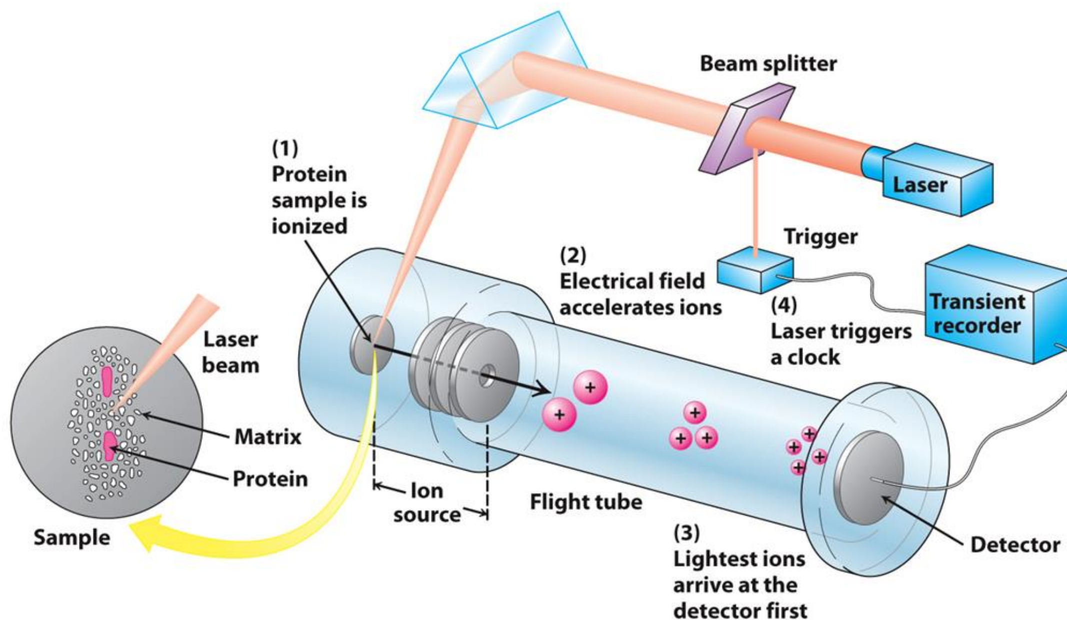


Figure 3.27
Biochemistry, Eighth Edition
© 2015 Macmillan Education

MALDI-TOF mass spectrum of insulin and β-lactoglobulin

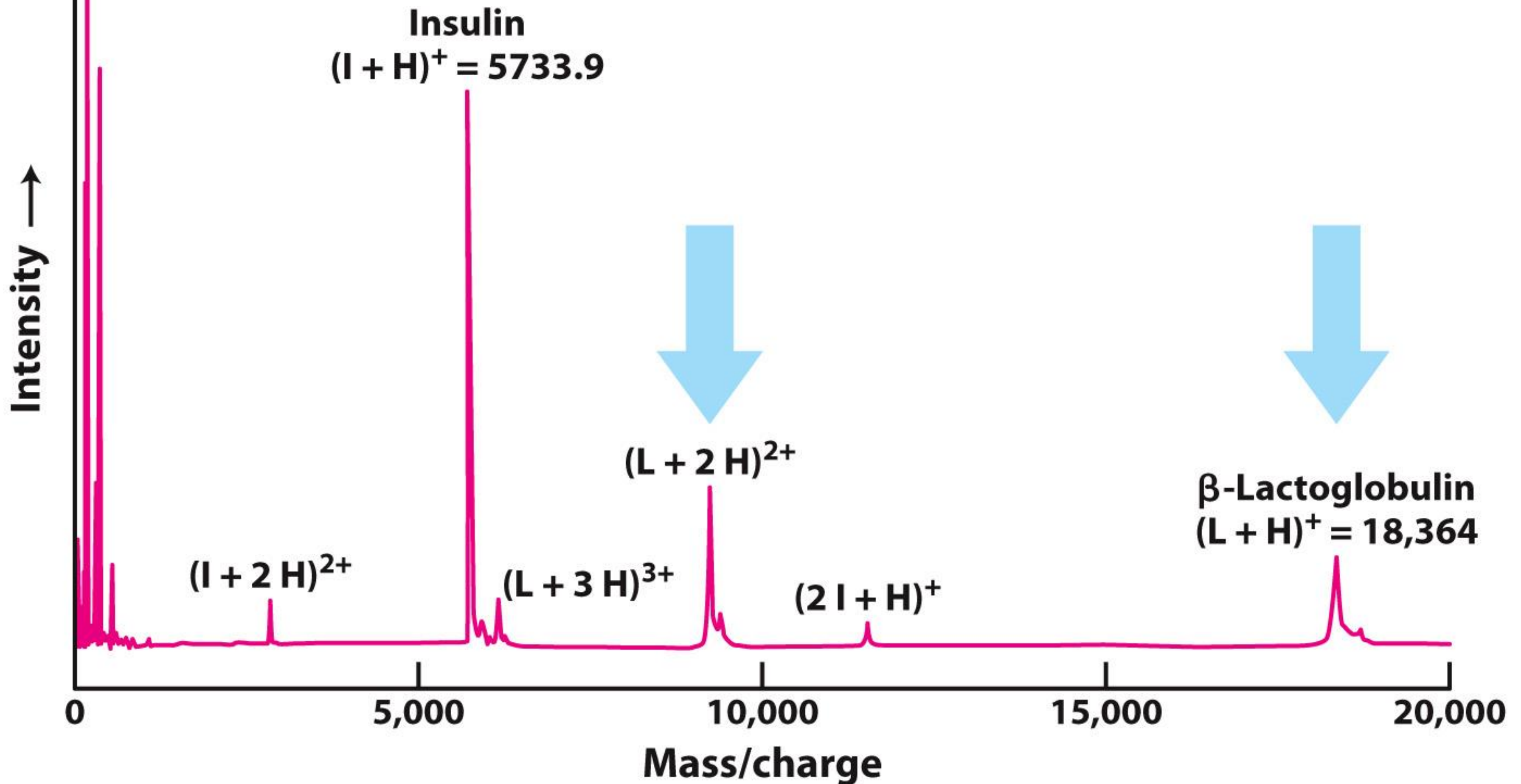


Figure 3.28
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Protein sequence determination

Edman degradation

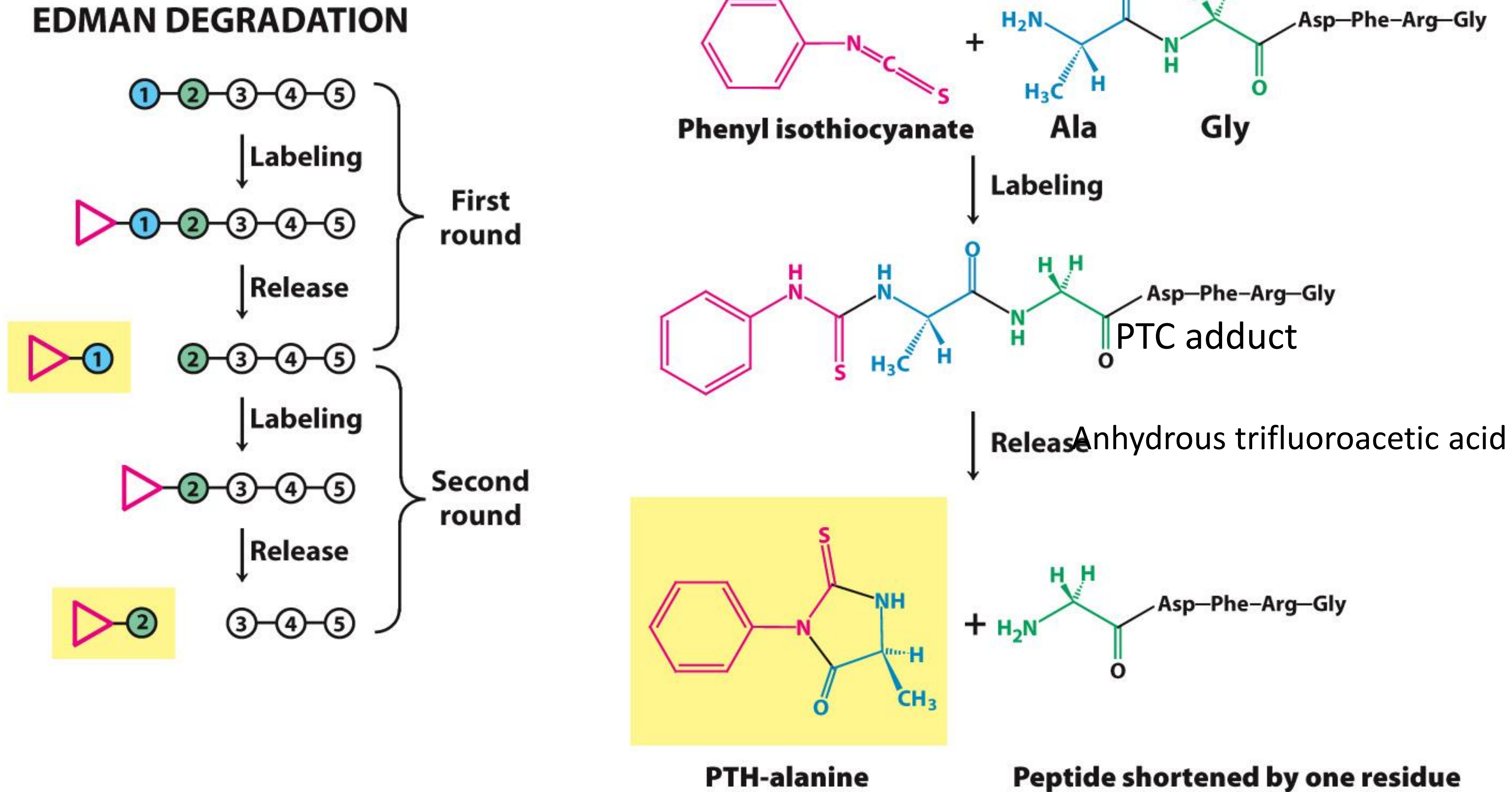


Figure 3.29
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Phenyl thiohydantoin (PTH) amino acid can be identified using chromatography

Up to 100 residues peptide can be sequenced

Sequencing by mass spectrometry

Tandem mass spectrometry facilitates protein sequence determination

Ions of proteins are broken into smaller peptides by bombardment with inert gas (He or Ne) and analyzed using MS

The fragments of peptide are bombarded with inert gas and analyzed is second mass analyzer. The products ions are formed such way that a single amino acid residue is cleaved from the sequence either at N- or C- terminus.

A set of fragments with one or more AA removed from original peptide is obtained

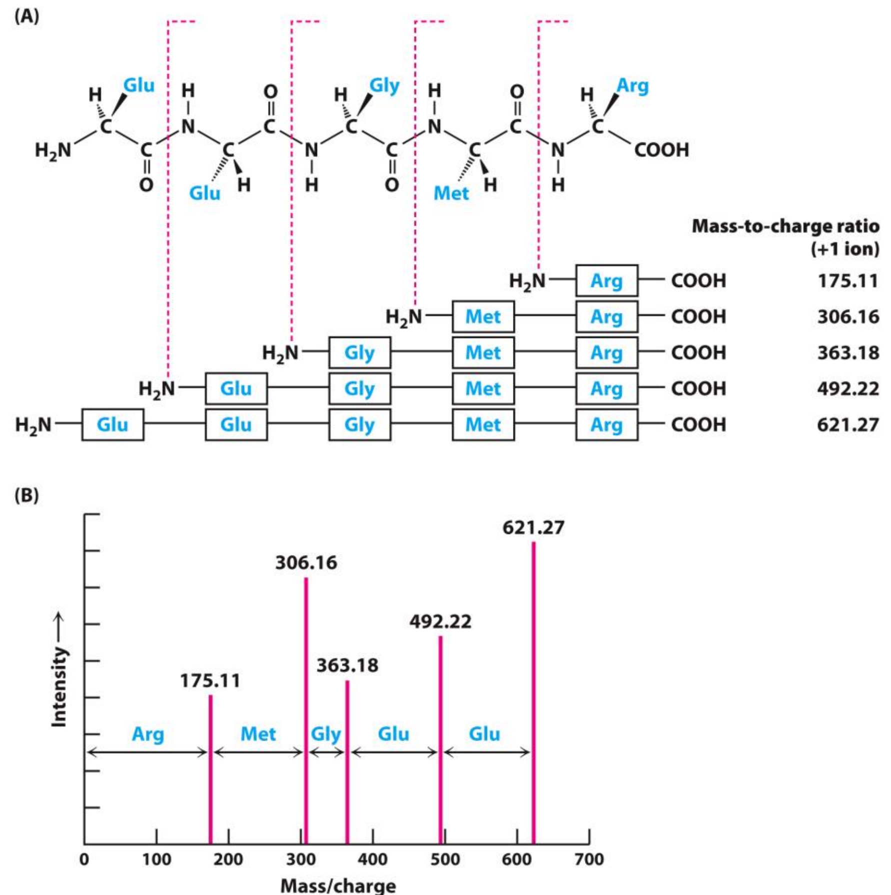
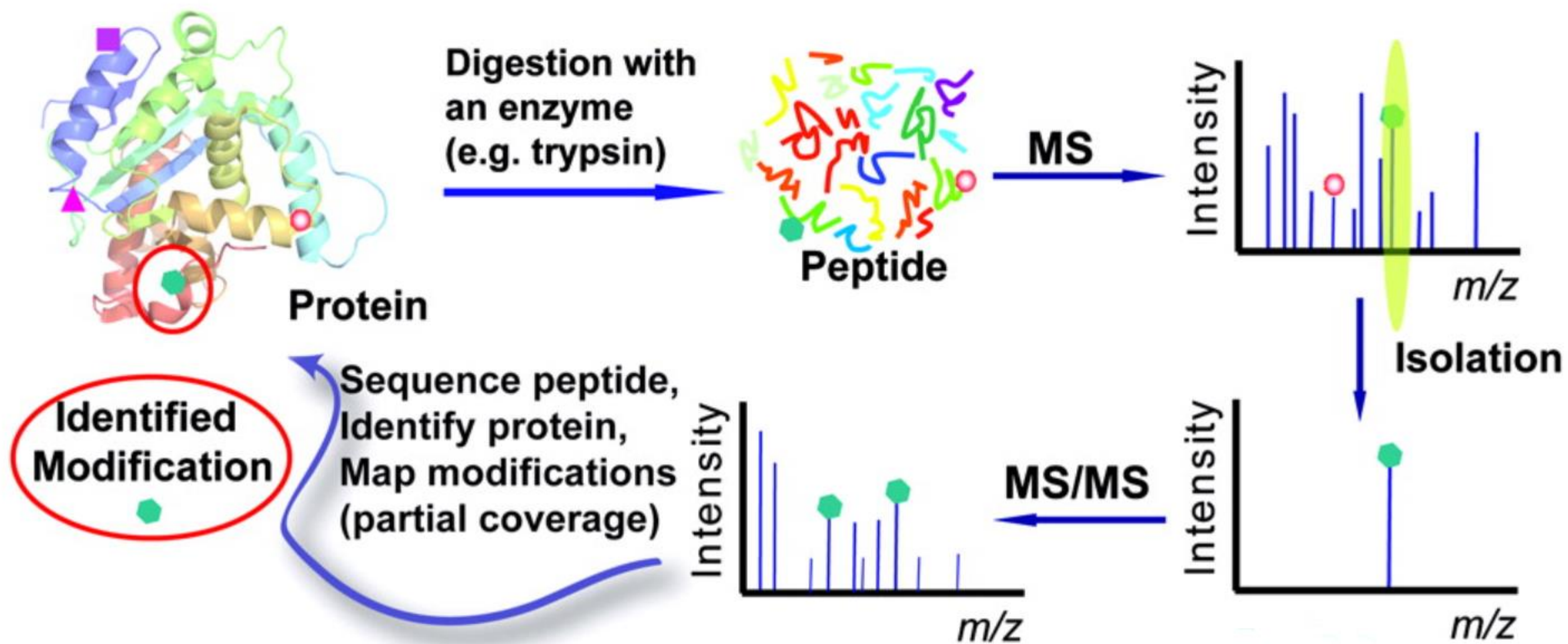


Figure 3.30
Biochemistry, Eighth Edition
© 2015 Macmillan Education


Digestion with
an enzyme
(e.g. trypsin)
Peptide
MS
Intensity
m/z
Protein
Isolation
Identified
Modification
Sequence peptide,
Identify protein,
Map modifications
(partial coverage)
Intensity
m/z
MS/MS
Intensity
m/z

Proteins are cleaved into smaller peptides chemically or enzymatically

Because the reactions of the Edman degradation and mass spectrometry procedure are not 100% effective, it is not possible to sequence polypeptides longer than 50 amino acids.

In order to sequence the entire protein, the protein is chemically or enzymatically cleaved to yield peptides of fewer than 50 amino acids.

The peptides are then ordered by performing a different cleavage procedure in order to generate overlap peptides.

Table 3.3 Specific cleavage of polypeptides

Reagent	Cleavage site
Chemical cleavage	
Cyanogen bromide	Carboxyl side of methionine residues
O-Iodosobenzoate	Carboxyl side of tryptophan residues
Hydroxylamine	Asparagine–glycine bonds
2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues
Enzymatic cleavage	
Trypsin	Carboxyl side of lysine and arginine residues
Clostripain	Carboxyl side of arginine residues
Staphylococcal protease	Carboxyl side of aspartate and glutamate residues (glutamate only under certain conditions)
Thrombin	Carboxyl side of arginine
Chymotrypsin	Carboxyl side of tyrosine, tryptophan, phenylalanine, leucine, and methionine
Carboxypeptidase A	Amino side of C-terminal amino acid (not arginine, lysine, or proline)

Table 3.3
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Extra steps are required for the determination of the disulfide bonds.

Disulfide bonds are cleaved by the addition of a reducing agent.

Disulfide bond reformation is prevented by the alkylation of the cysteine residues.

Modified peptides are then sequenced

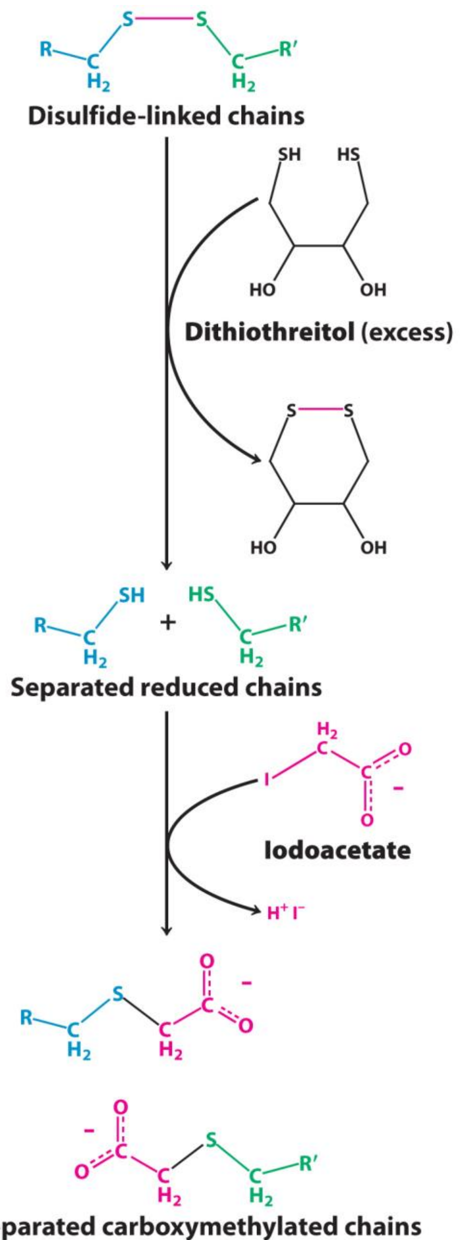


Figure 3.32
Biochemistry, Eighth Edition
© 2015 Macmillan Education

The amino acid sequence of a protein provides valuable information

1. Amino acid sequences of proteins can be compared to identify similarities.

2. Comparison of the sequence of the same protein from different species yields evolutionary information.

3. Amino acid sequence searches can reveal the presence of internal repeats.

4. Sequencing information can identify signals that determine the location of the protein or processing signals.

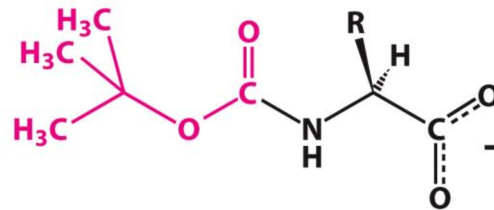
5. Sequence information can be used to generate antibodies for the protein.

6. Amino acid sequence can be used to generate DNA probes specific for the gene encoding the protein.

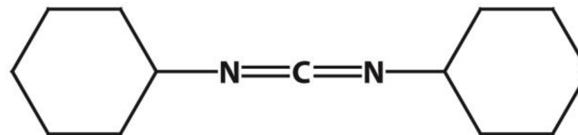
Peptide solid state synthesis

Synthetic peptide are synthesized in a step-wise fashion with the carboxyl terminus of the growing chain attached to an inert matrix.

t-Boc blocks the amino group of the incoming amino acid and DCC facilitates peptide bond formation.



***t*-Butyloxycarbonyl amino acid**
(*t*-Boc amino acid)



Dicyclohexylcarbodiimide
(DCC)

Unnumbered 3 p93
Biochemistry, Eighth Edition
© 2015 Macmillan Education

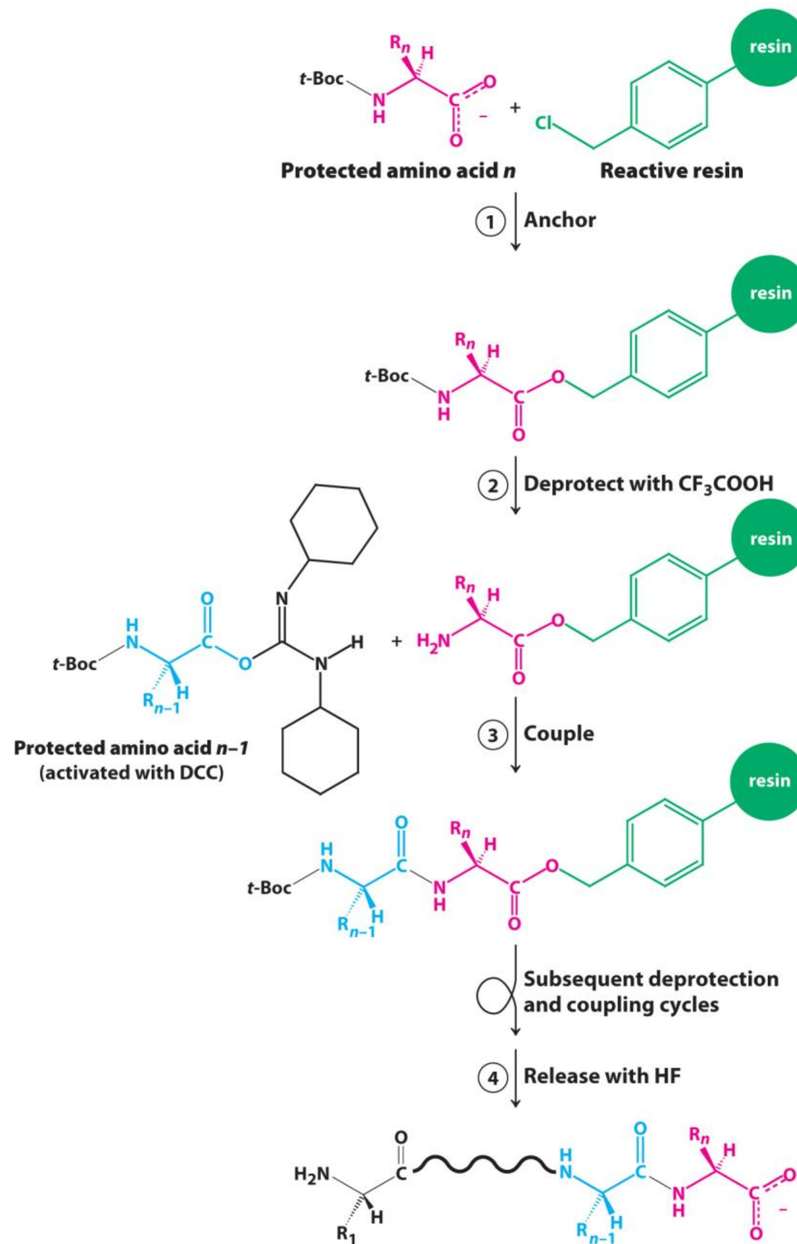


Figure 3.37
Biochemistry, Eighth Edition
© 2015 Macmillan Education

X-ray crystallography

X-ray crystallography reveals three-dimensional structure in atomic detail

Crystals of proteins are irradiated with x-rays.

1. Electrons of the atoms scatter x-rays.
2. The scattered waves recombine.
3. The way in which the scattered x-rays recombine reveals the atomic arrangement.

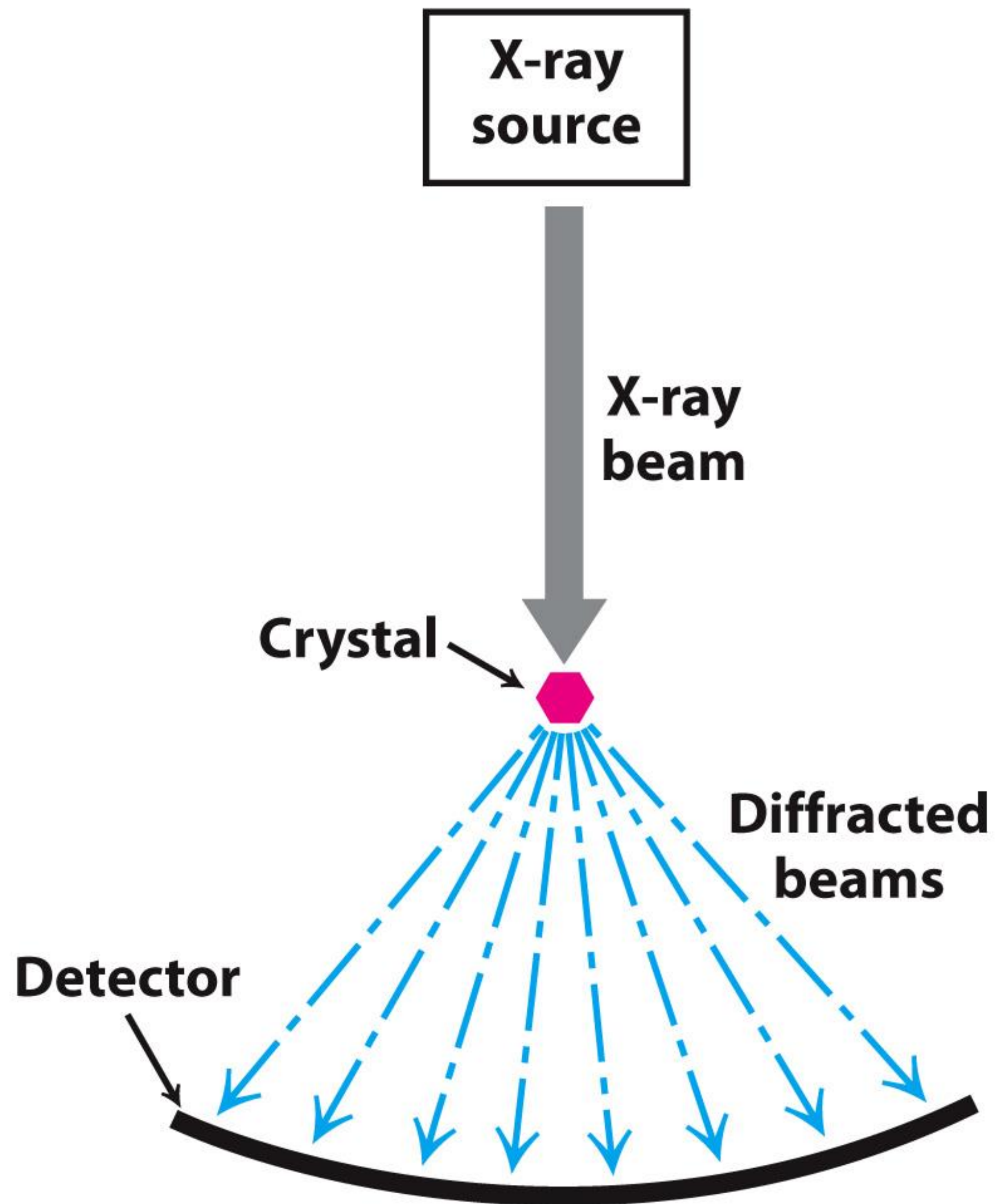


Figure 3.38
Biochemistry, Eighth Edition
© 2015 Macmillan Education

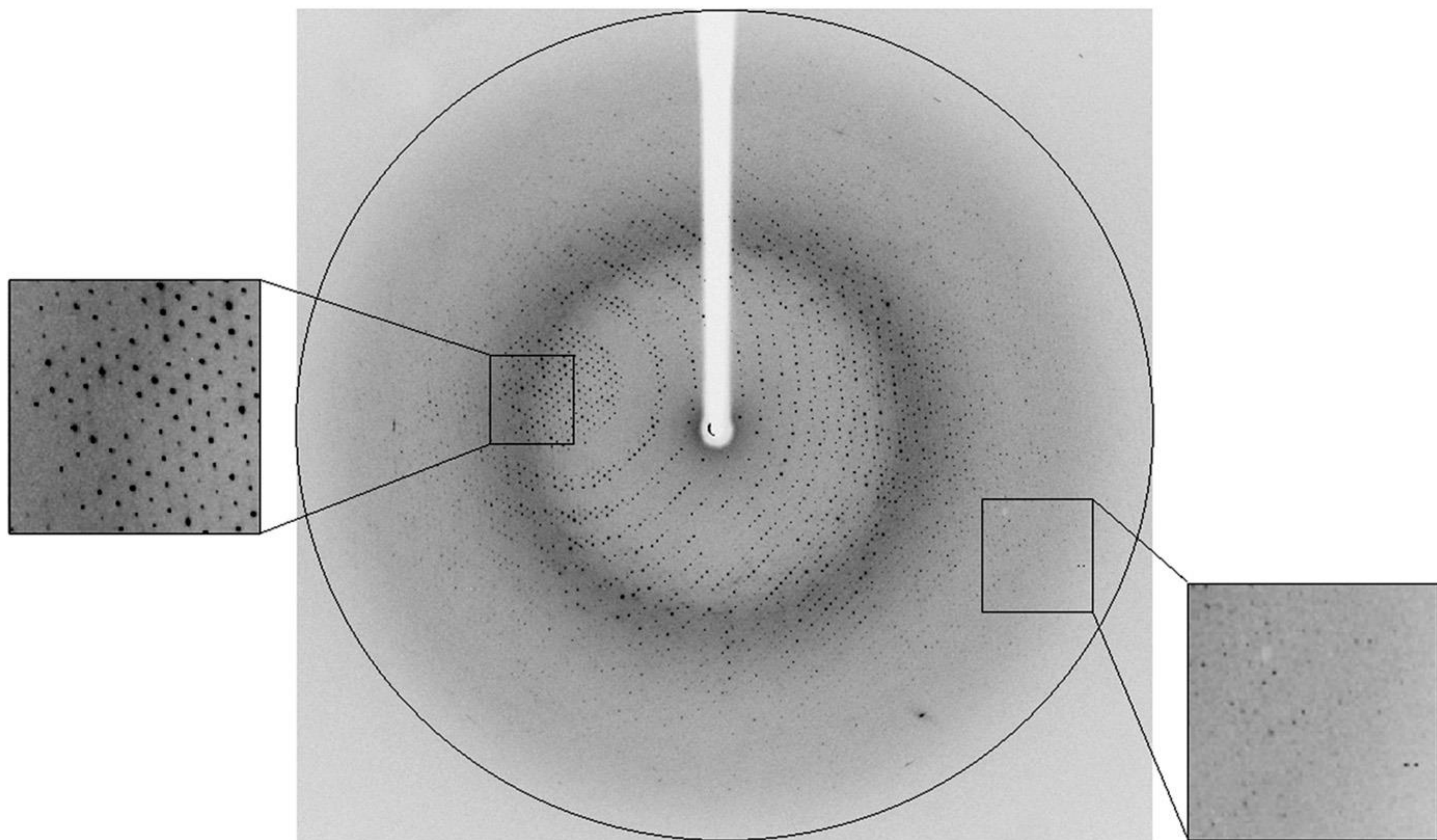
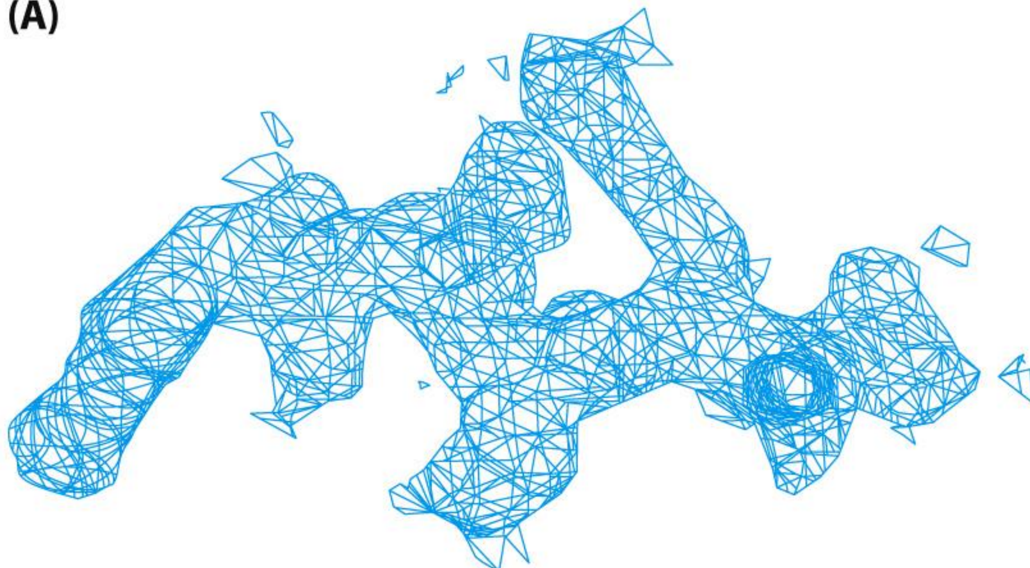


Figure 3.39
Biochemistry, Eighth Edition
S. Lansky, O. Alalouf, V. Solomon, A. Alhassid, L. Govada, N. E. Chayan, H. Belrhali, Y. Shoham, and G. Shoham. *Acta Cryst.* F69:430–434, 2013, Fig. 2.



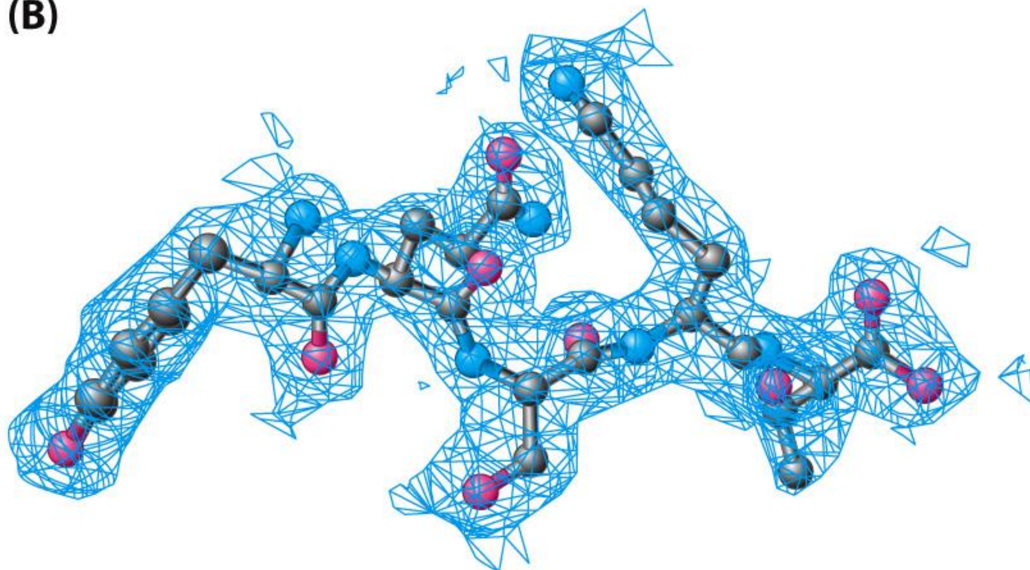


Figure 3.40

Biochemistry, Eighth Edition

© 2015 Macmillan Education

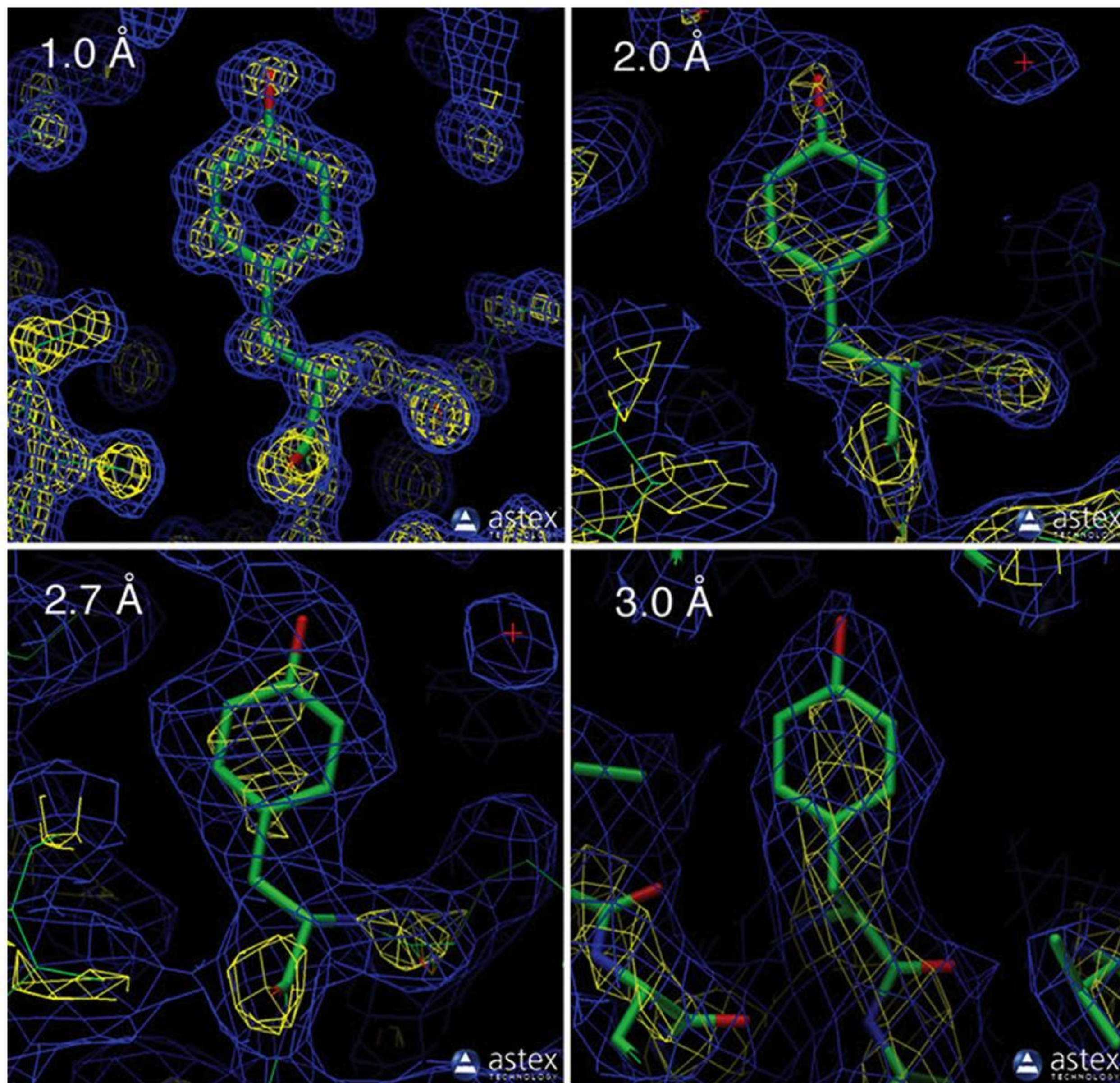


Figure 3.42

Biochemistry, Eighth Edition

Data from www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/resolution.html

Nuclear magnetic resonance spectroscopy can reveal the structures of proteins in solution

NMR is based on the fact that certain atomic nuclei are intrinsically magnetic and can exist in two spin states when an external magnetic field is applied.

The nuclei of the sample absorb electromagnetic radiation at different frequencies termed chemical shifts.

The chemical shifts depend on the environment of the nuclei, and the environment depends on protein structure.

One-dimensional NMR reveals changes to a particular chemical group that are close in the primary sequence.

Two-dimensional NMR (NOSEY) displays groups that are in close proximity.

Table 3.4 Biologically important nuclei giving NMR signals

Nucleus	Natural abundance (% by weight of the element)
^{1}H	99.984
^{2}H	0.016
^{13}C	1.108
^{14}N	99.635
^{15}N	0.365
^{17}O	0.037
^{23}Na	100.0
^{25}Mg	10.05
^{31}P	100.0
^{35}Cl	75.4
^{39}K	93.1

Table 3.4
***Biochemistry*, Seventh Edition**
© 2012 W. H. Freeman and Company

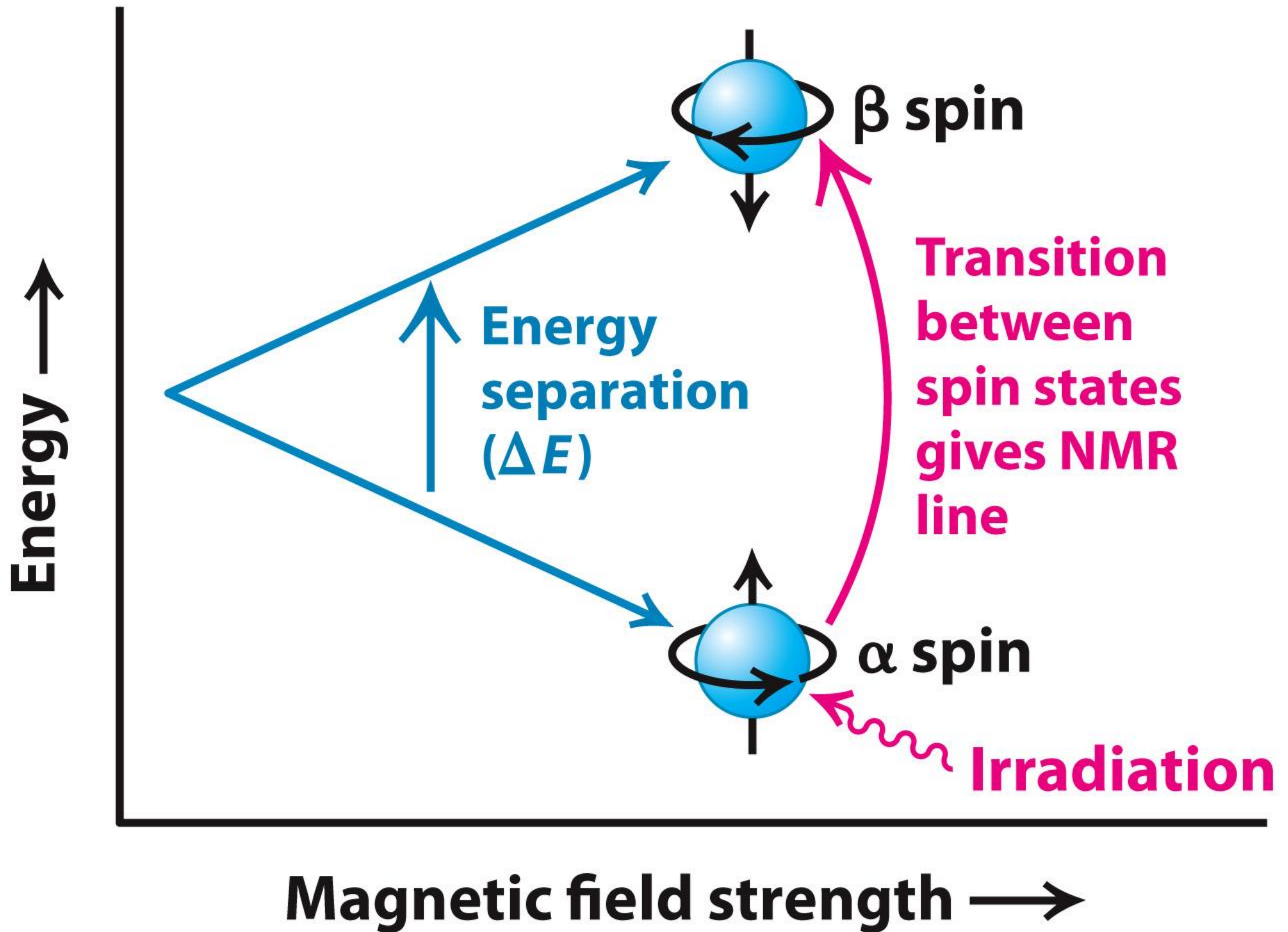


Figure 3.43
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Energy of two orientation of a nucleous spin depends on the magnetic field

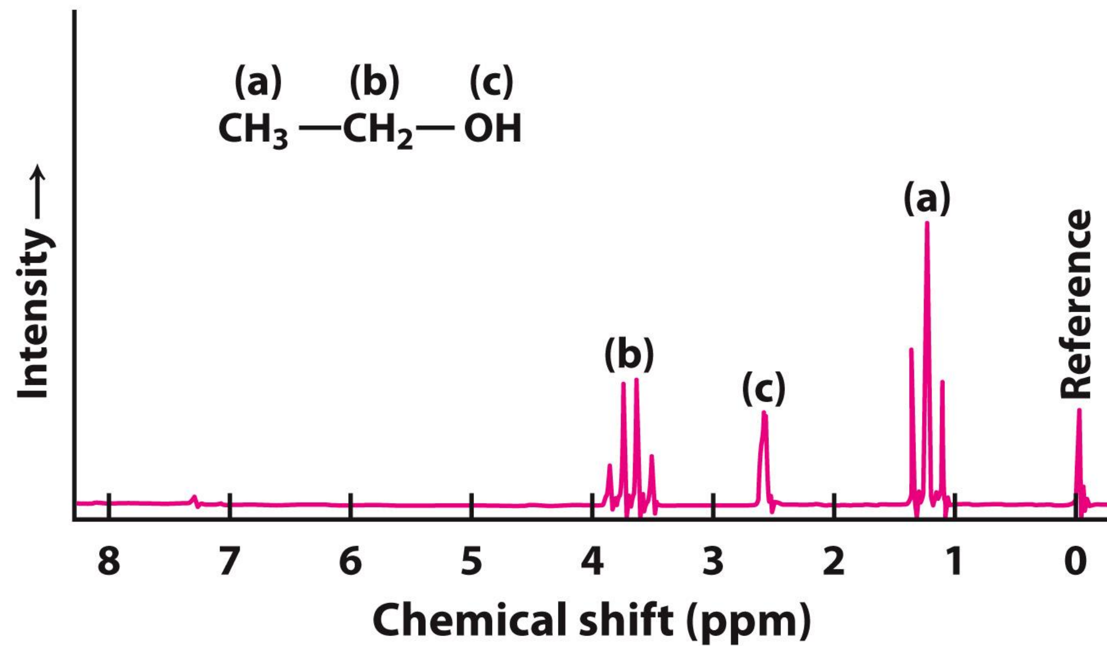


Figure 3.44a
Biochemistry, Eighth Edition
© 2015 Macmillan Education

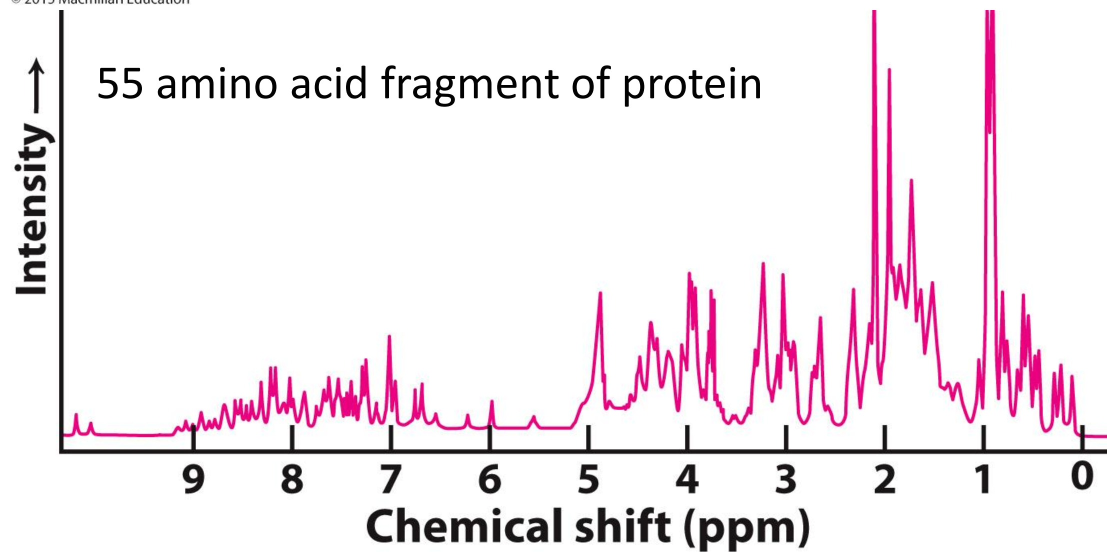


Figure 3.44b
Biochemistry, Eighth Edition
© 2015 Macmillan Education

2 D NMR

Takes advantage of Nuclear Overhauser effect (NOE) : the interaction between the nuclei if proportional to the $1/r^6$. Magnetisation is transfer from an excited nuclei to the unexcited nuclei if they are closer than 5 A.

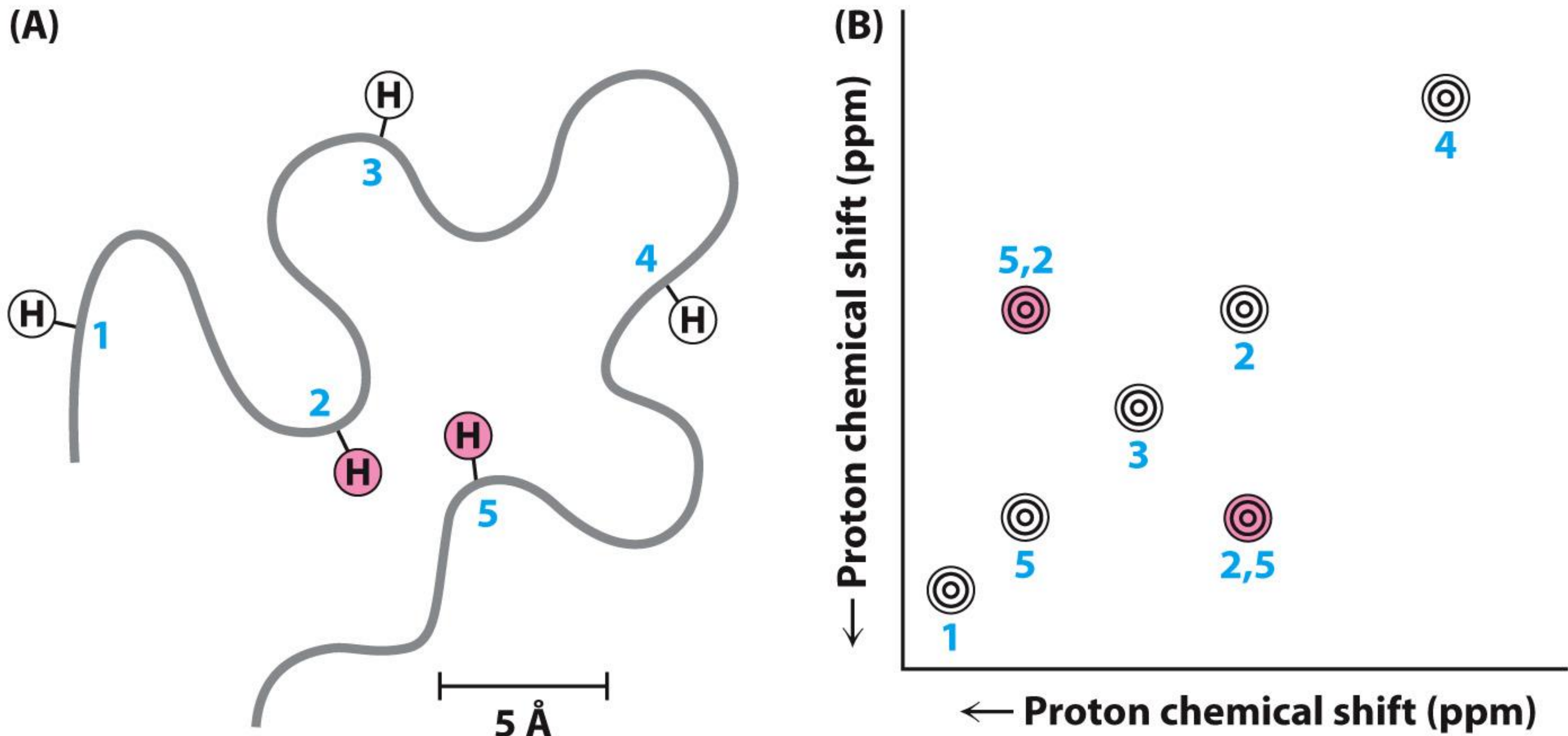


Figure 3.45
Biochemistry, Eighth Edition
© 2015 Macmillan Education

2,5 and 5,2 are cross peaks

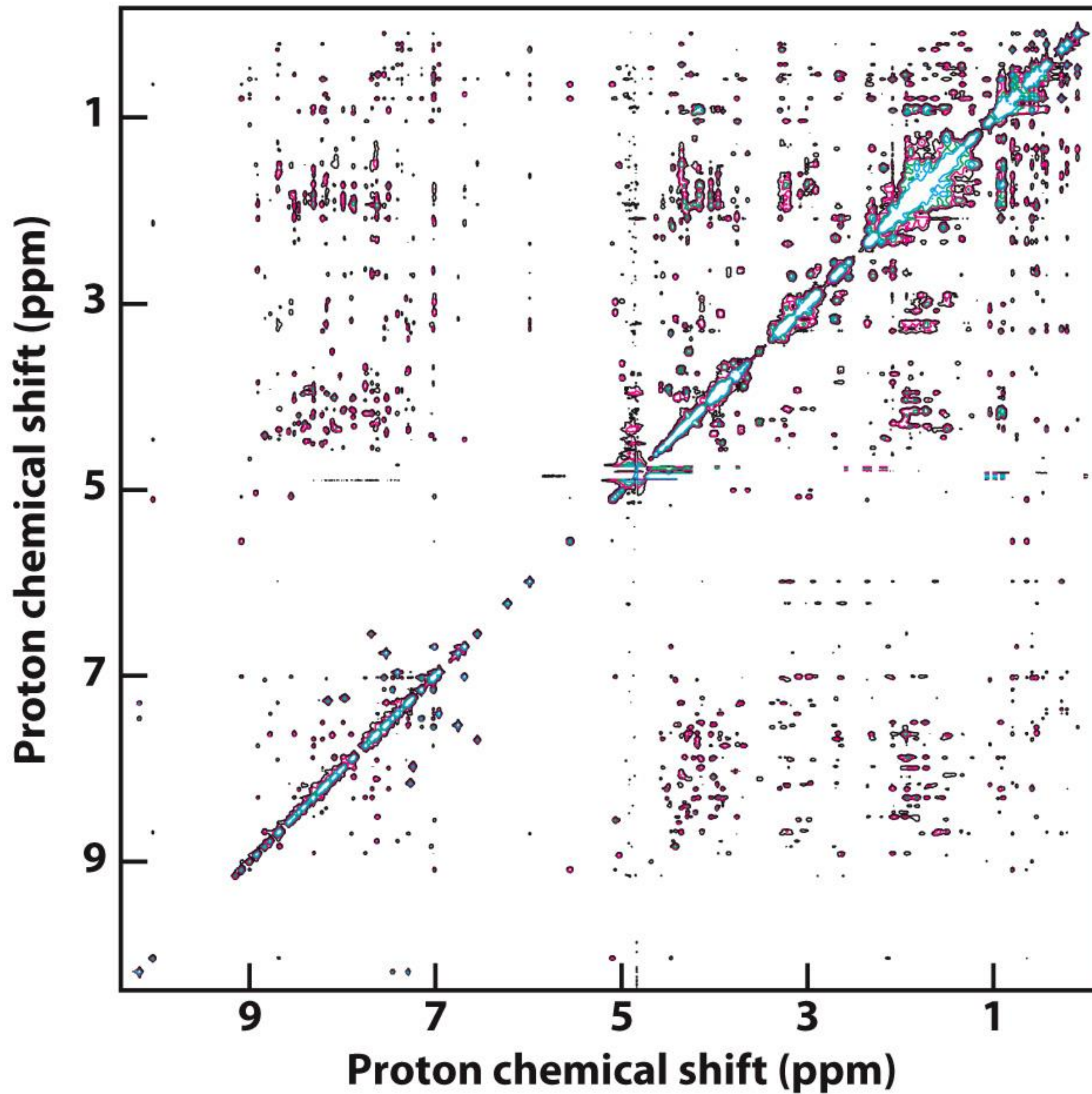
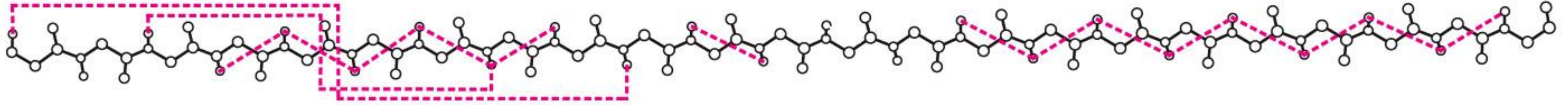


Figure 3.46
Biochemistry, Eighth Edition
Courtesy of Dr. Barbara Amann and Dr. Wesley McDermott



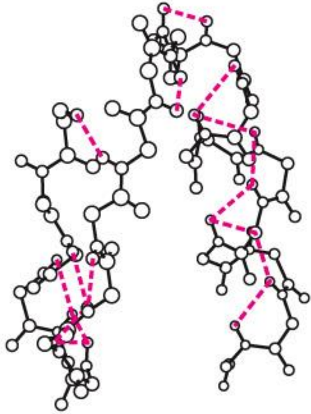


Figure 3.47
Biochemistry, Eighth Edition
© 2015 Macmillan Education

Calculated structure

28 amino acid domain from a zinc-finger-DNA-binding protein

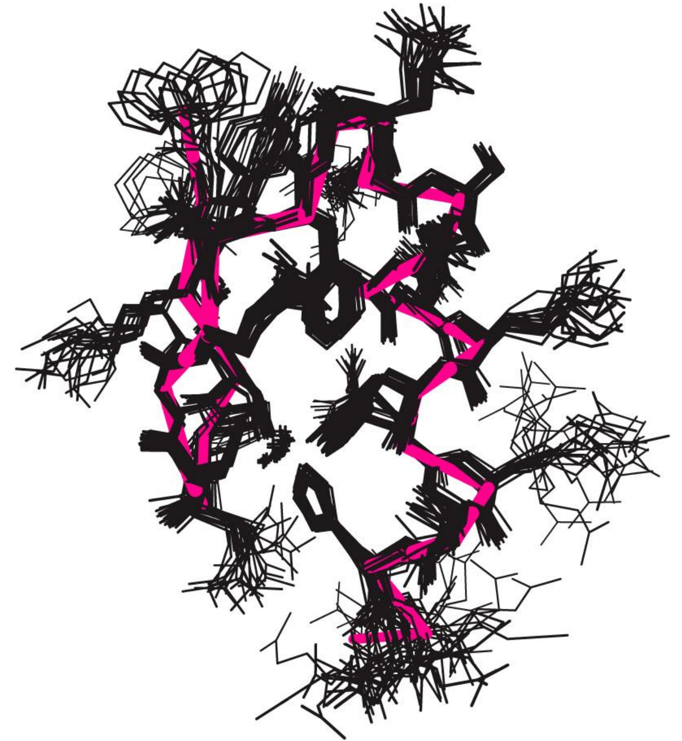


Figure 3.48
Biochemistry, Eighth Edition
Courtesy of Dr. Barbara Amann