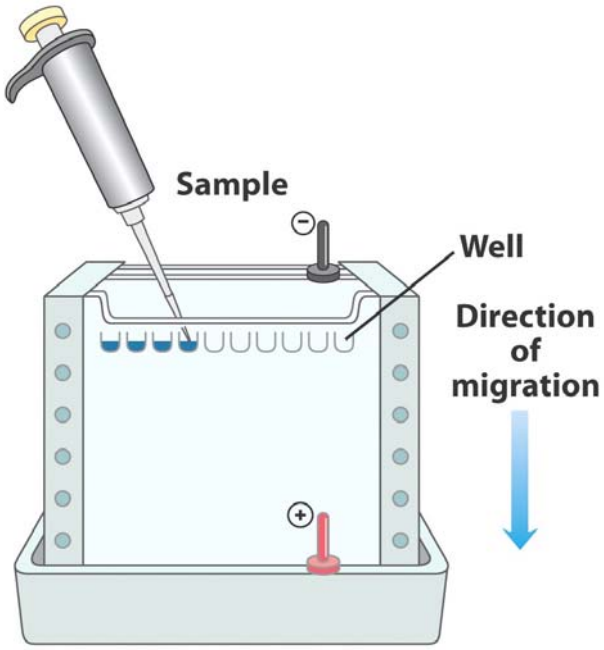
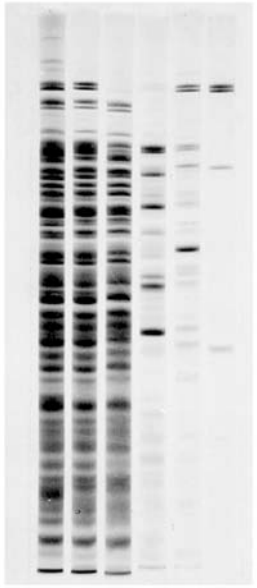
*So.....*

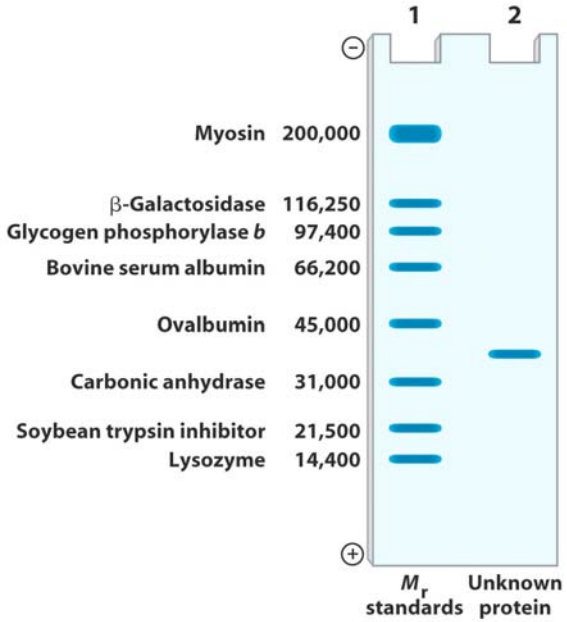
*Let us say you have an impure solution containing a protein of interest.*

*Q: How do you (a) analyze what you have and (b) purify what you want?*

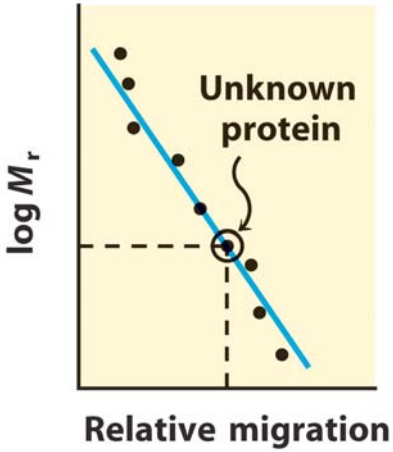
# Polyacrylamide Gel Electrophoresis (PAGE)



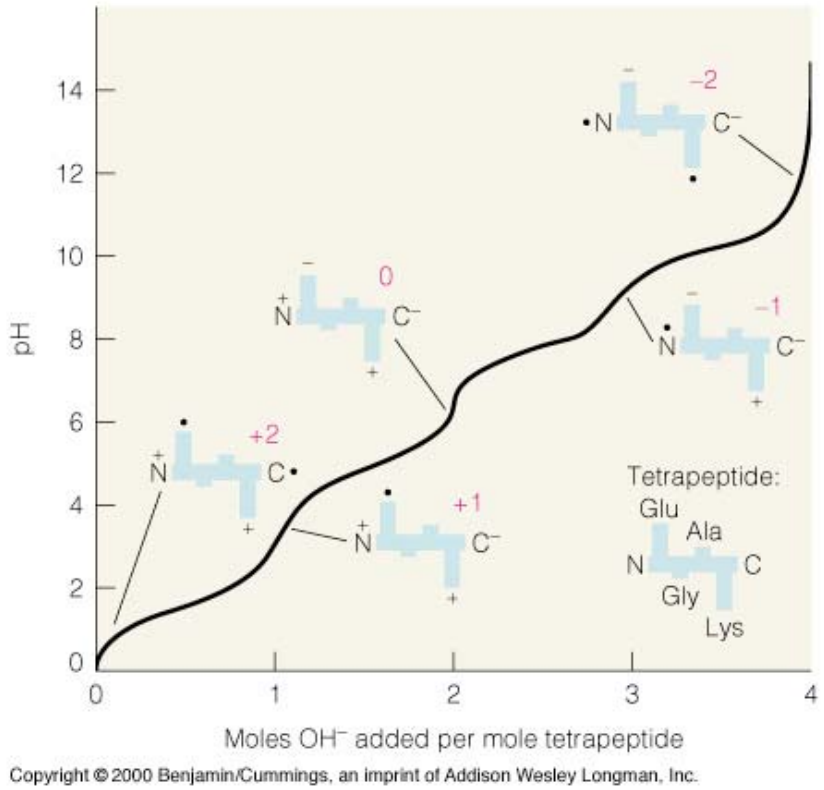






Note: proteins are usually mixed with a detergent, sodium dodecylsulfate (SDS), and a tracking dye to make the sample. The SDS binds to the protein and gives it size-dependent negative charge and consistent hydrodynamic properties.



# Polyampholyte Character of a Tetrapeptide and Isoelectric Points





| Group | pKa |
|---|---|
| $\alpha$-$NH_3^+$ | 9.7 |
| Glu $\gamma$-COOH | 4.2 |
| Lys $\varepsilon$-$NH_3^+$ | 10.0 |
| $\alpha$-COOH | 2.2 |

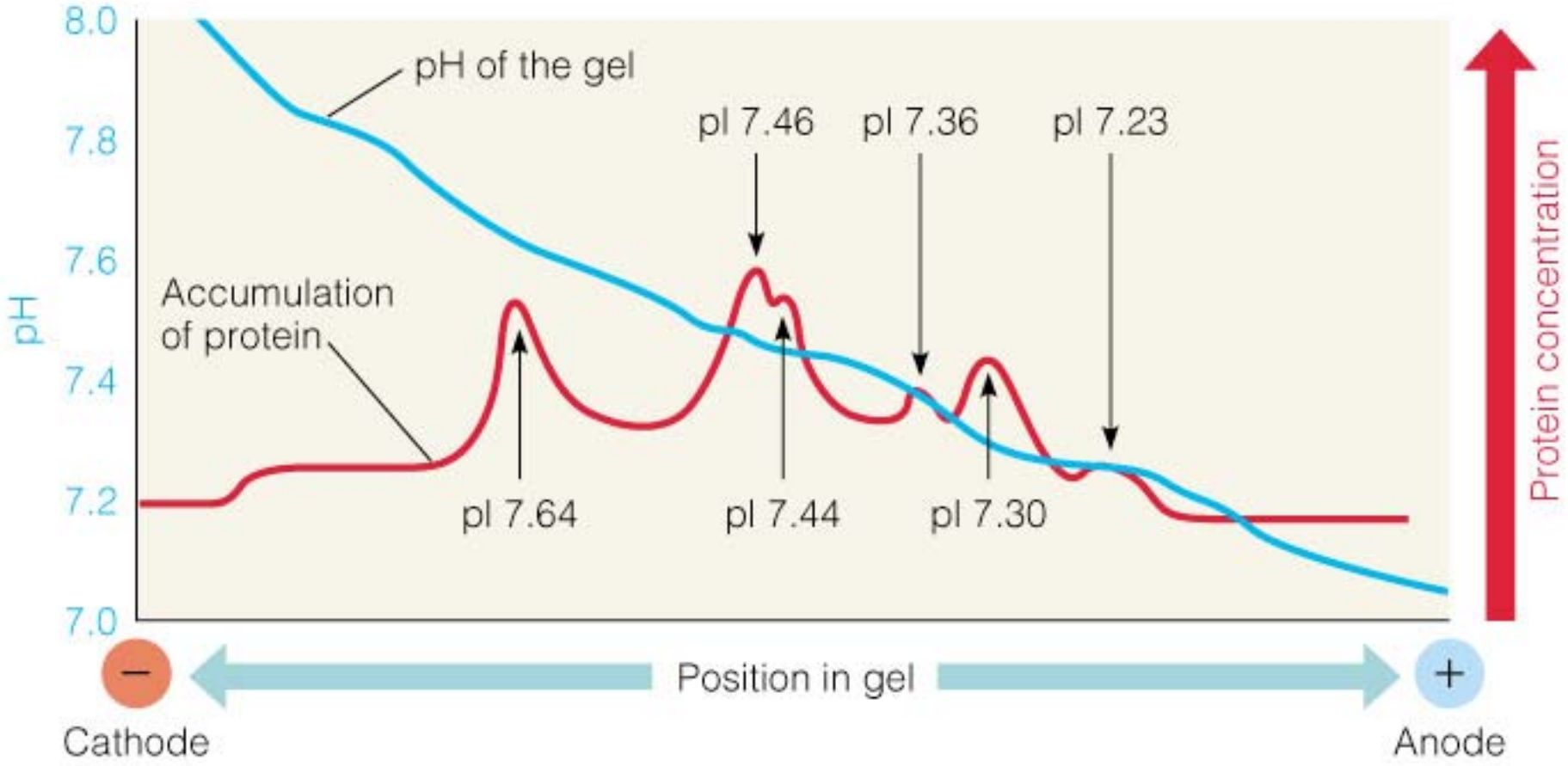
**Isoelectric Point (pI)**, pH at which molecule has net zero charge, determined using computer program for known sequence or empirically (by isoelectric focusing)

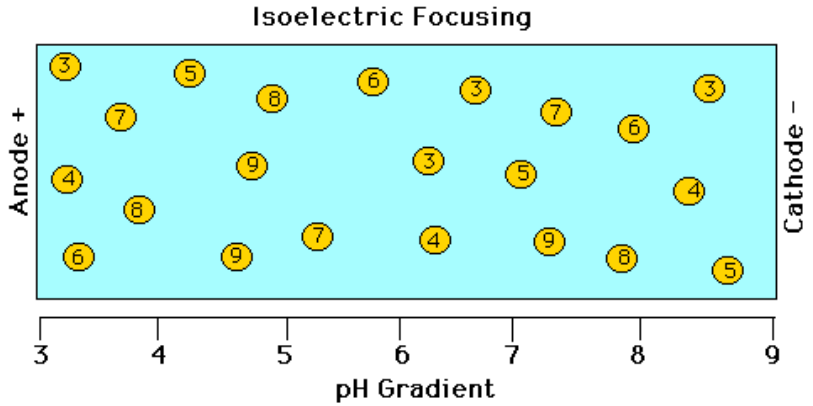
**TABLE 3–6 The Isoelectric Points of Some Proteins**

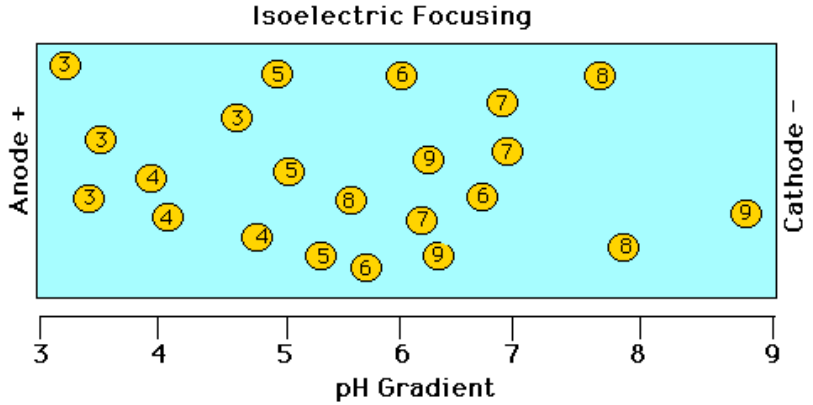
| *Protein* | *pI* |
|---|---|
| Pepsin | <1.0 |
| Egg albumin | 4.6 |
| Serum albumin | 4.9 |
| Urease | 5.0 |
| $\beta$-Lactoglobulin | 5.2 |
| Hemoglobin | 6.8 |
| Myoglobin | 7.0 |
| Chymotrypsinogen | 9.5 |
| Cytochrome *c* | 10.7 |
| Lysozyme | 11.0 |

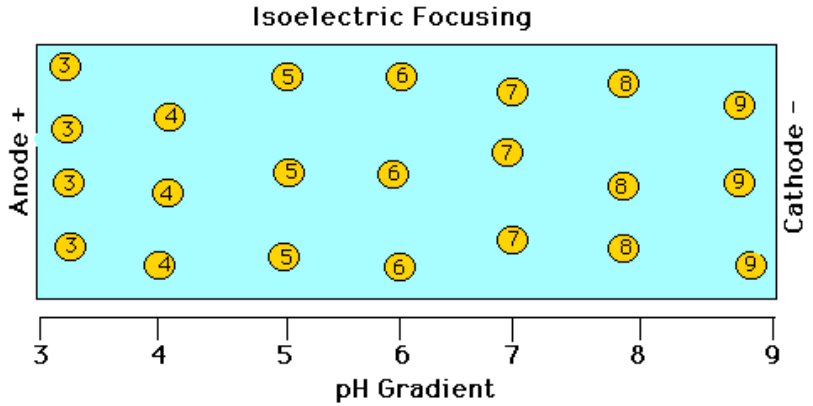
# Isoelectric Focusing

Electrophoresis through polyacrylamide gel in which there is a pH gradient.




Isoelectric Focusing
Anode +
Cathode -
3 4 5 6 7 8 9
pH Gradient


Isoelectric Focusing
Anode +
Cathode -
3 4 5 6 7 8 9
pH Gradient


Isoelectric Focusing
Anode +
Cathode -
3 4 5 6 7 8 9
pH Gradient

# Two-Dimensional Gel Electrophoresis

- Separate proteins based on pI in 1st dimension
- Separate proteins based on molecular weight in 2nd dimension

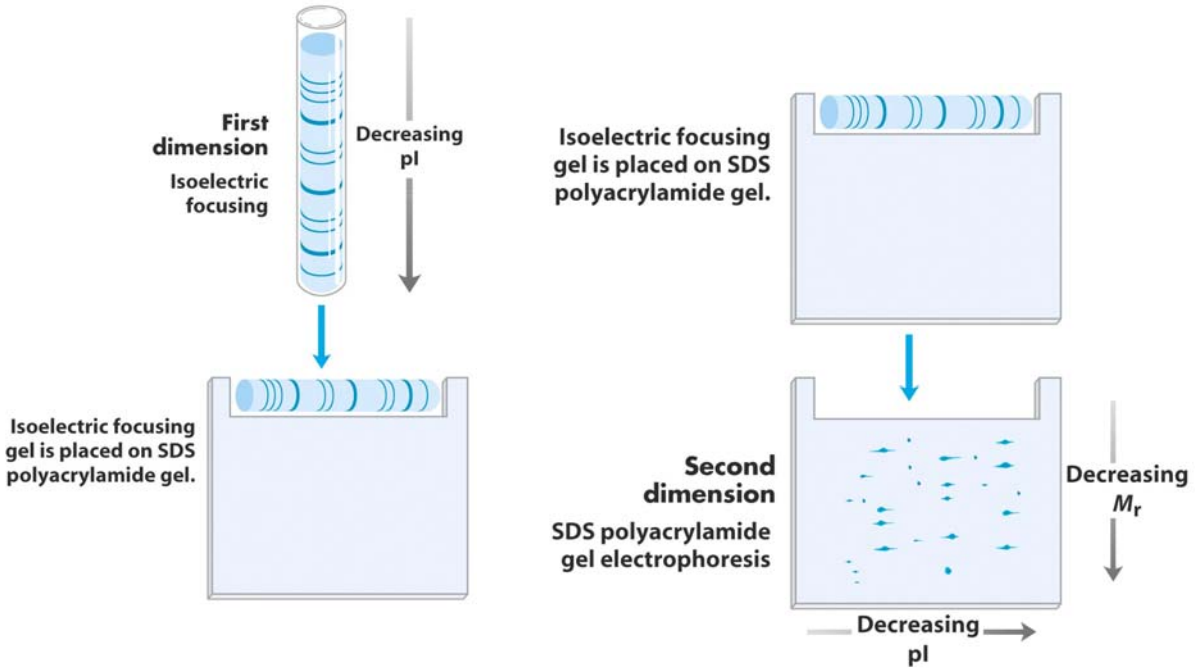
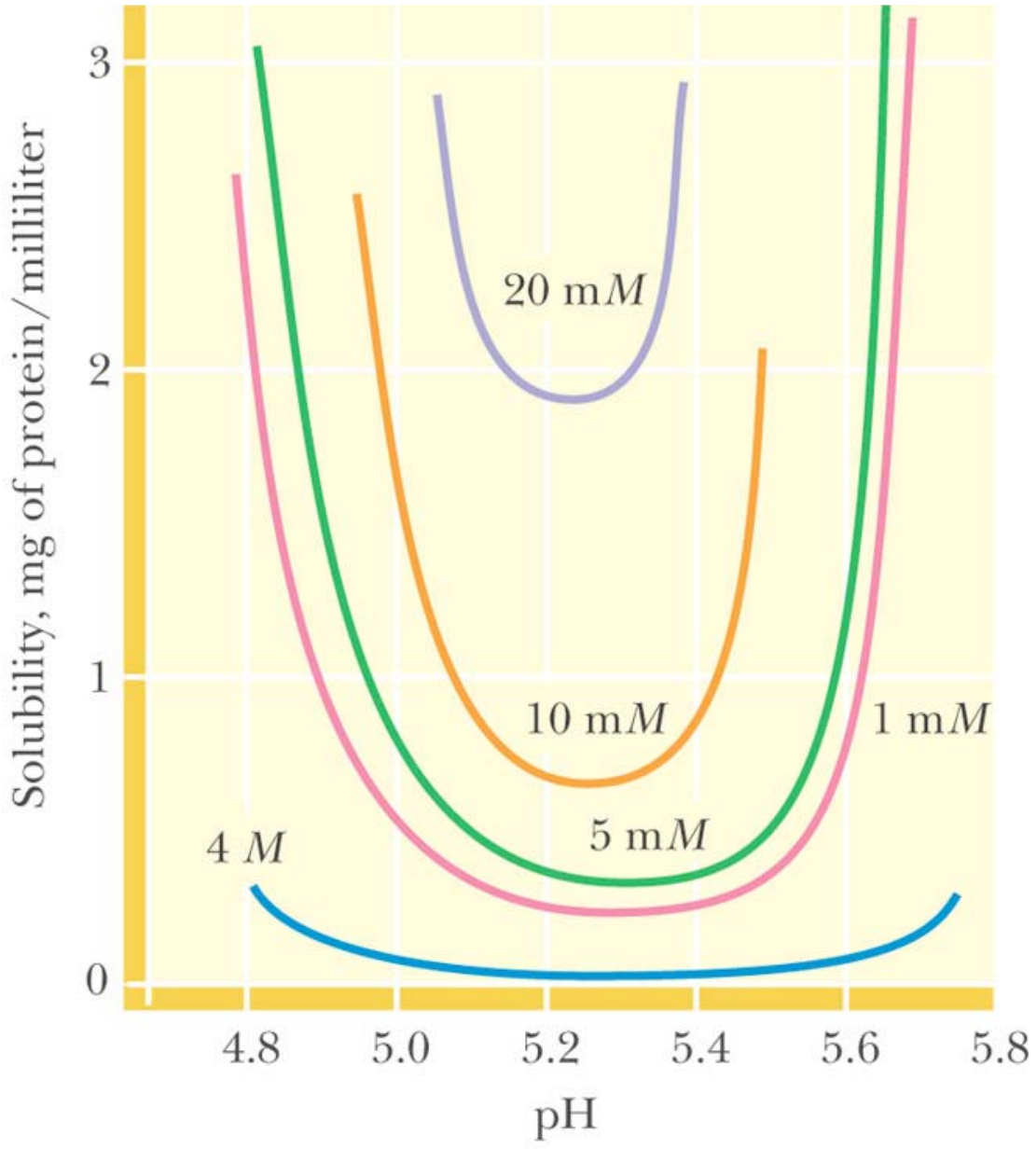
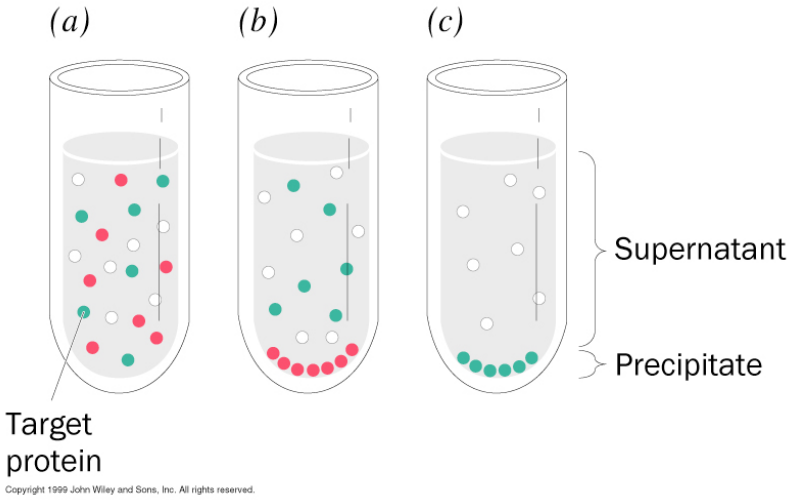


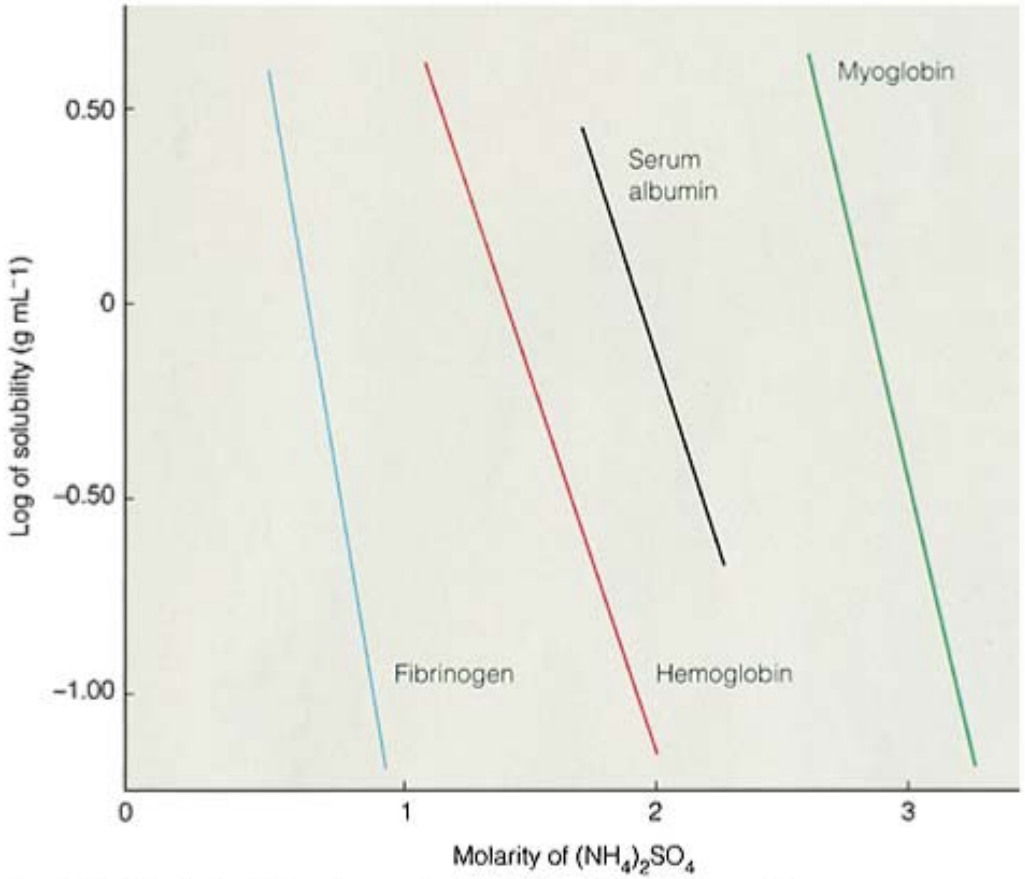

Figure 5.11
The solubility of most globular proteins is markedly influenced by pH and ionic strength. This figure shows the solubility of a typical protein as a function of pH and various salt concentrations.



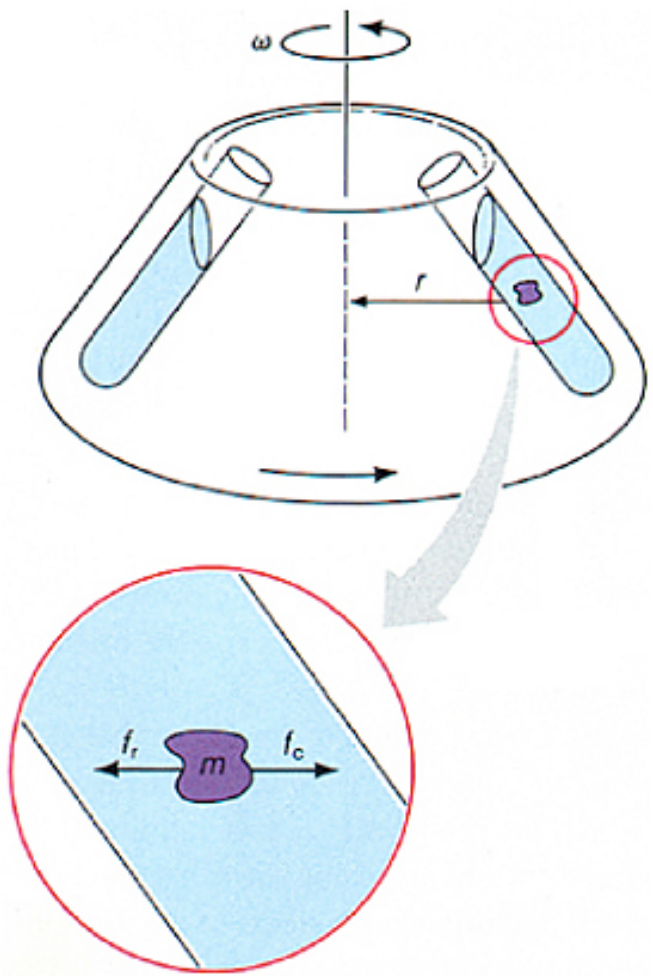
# “Salting Out”: Ammonium Sulfate Precipitation in Protein Fractionation







# Centrifugation



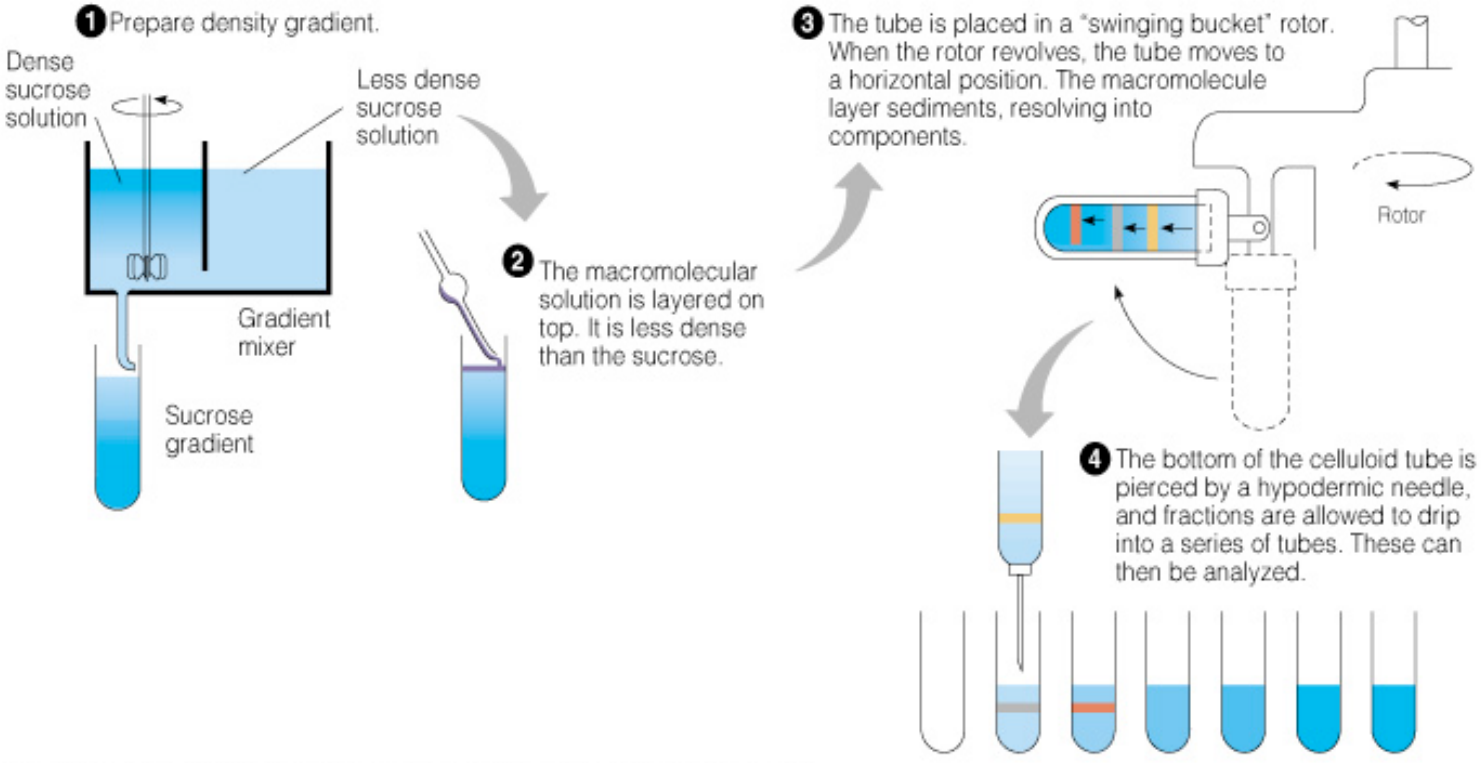



Low-speed, high-speed, or ultracentrifugation: different spin speeds and g forces

## Centrifugation Methods

- Differential (Pelleting) – simple method for pelleting large particles using fixed-angle rotor (pellet at bottom of tube vs. supernatant solution above)
- Zonal ultracentrifugation (e.g. sucrose-gradient) – swinging-bucket rotor
- Equilibrium-density gradient ultracentrifugation (e.g. CsCl) – swinging-bucket or fixed-angle rotor

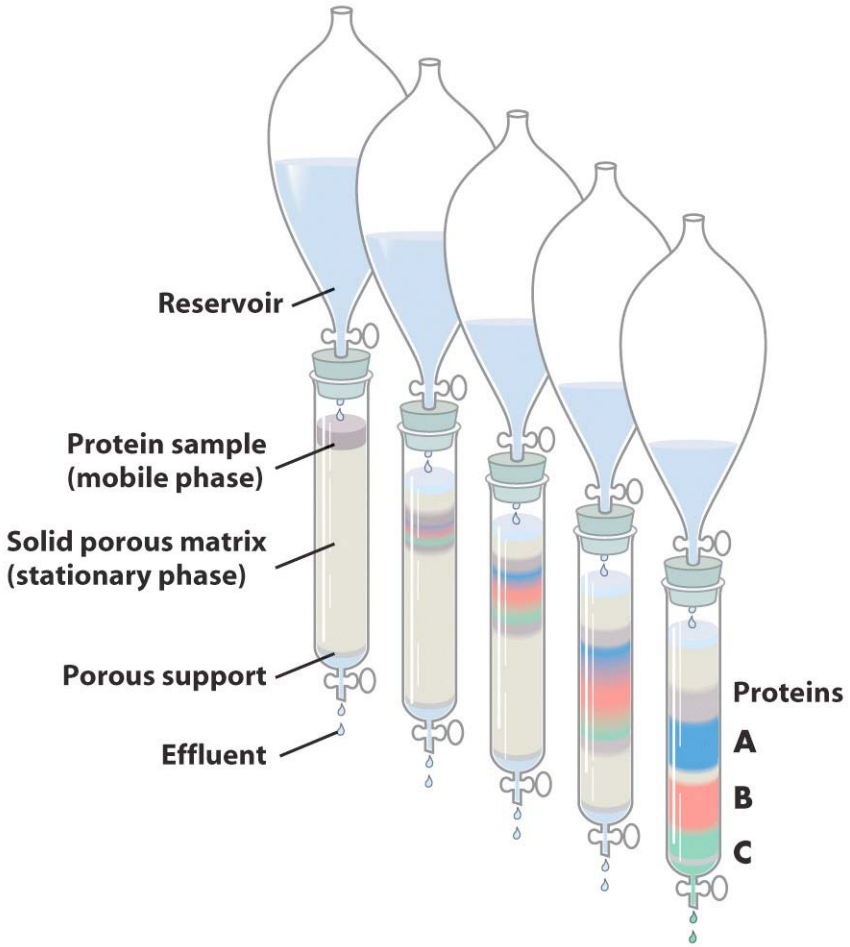
# Zonal Centrifugation: Sucrose-Gradient Preparative Centrifugation

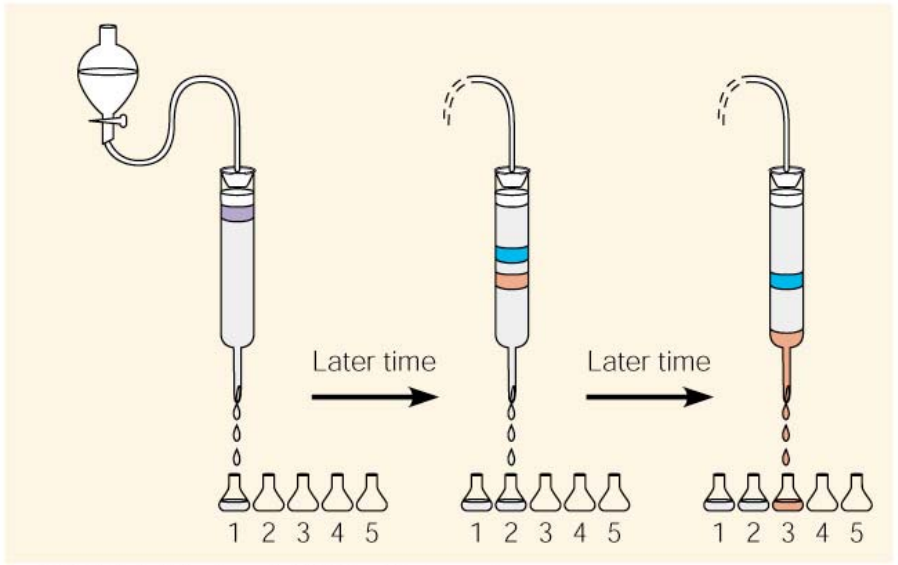





Separates by sedimentation coefficient
(determined by size and shape of solutes)

# Column Chromatography



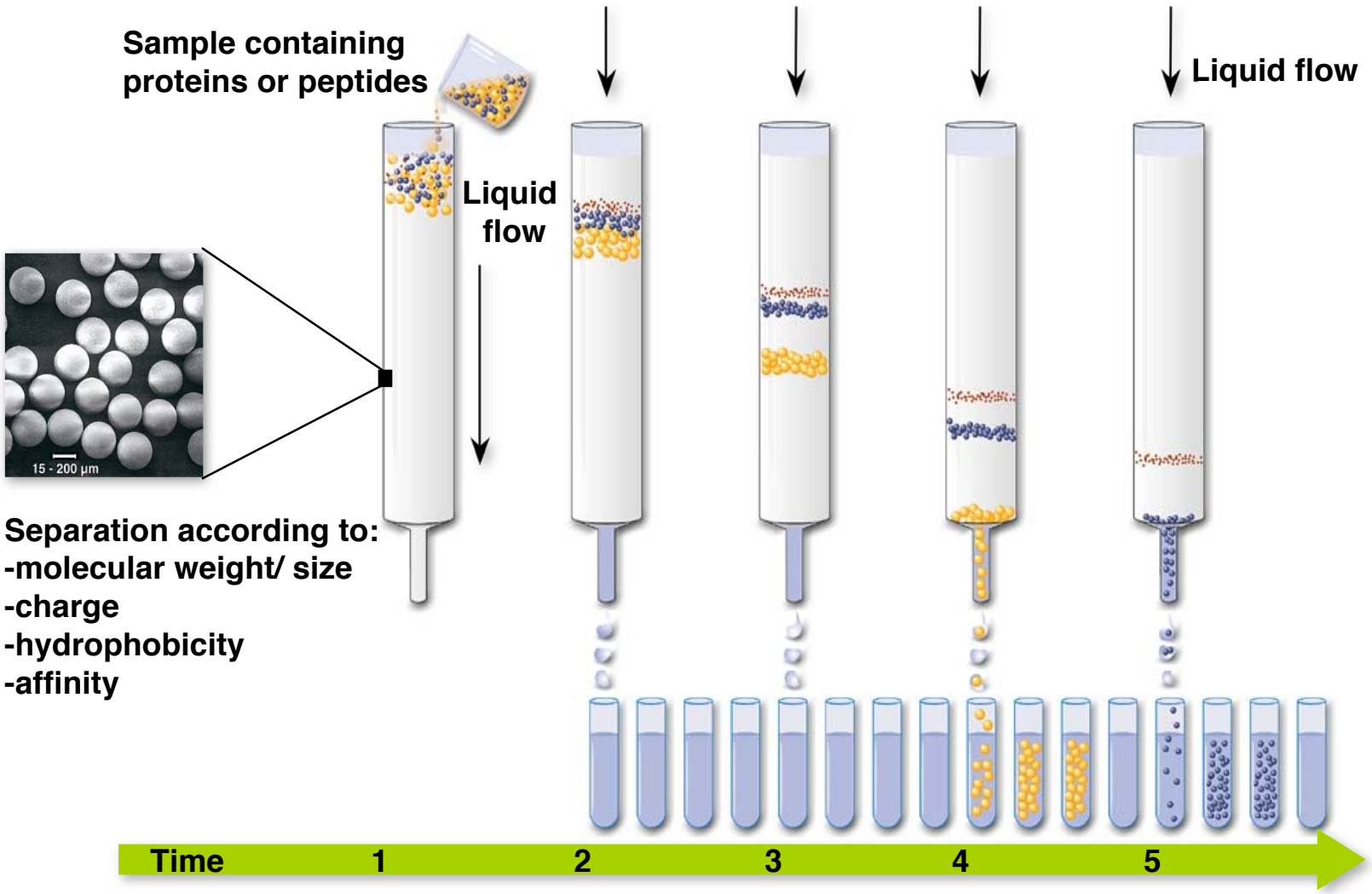





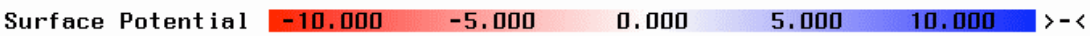

Flow-through

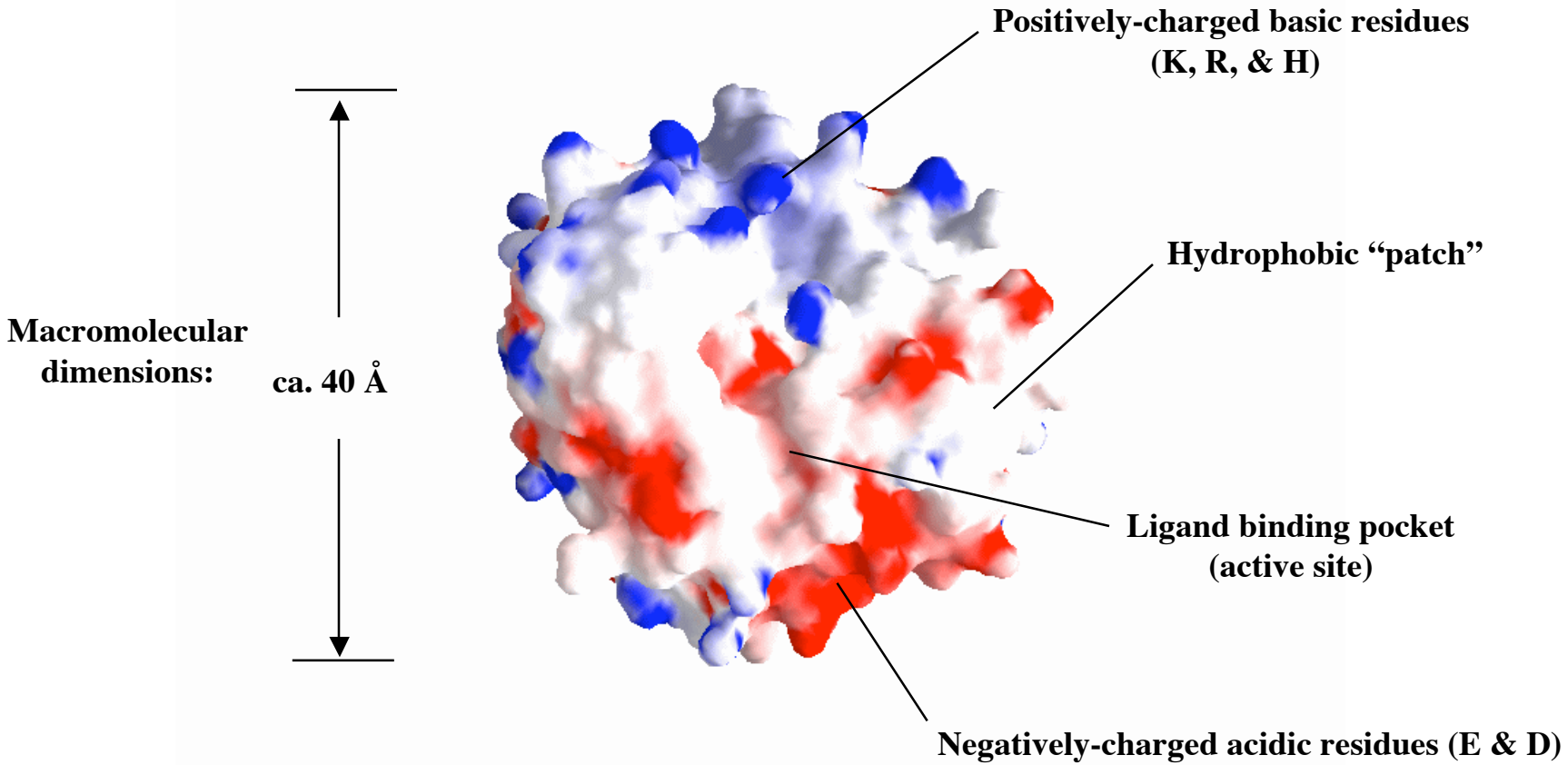
Eluate

# Chromatography



# Proteins are Amphiphilic Macro-Ions



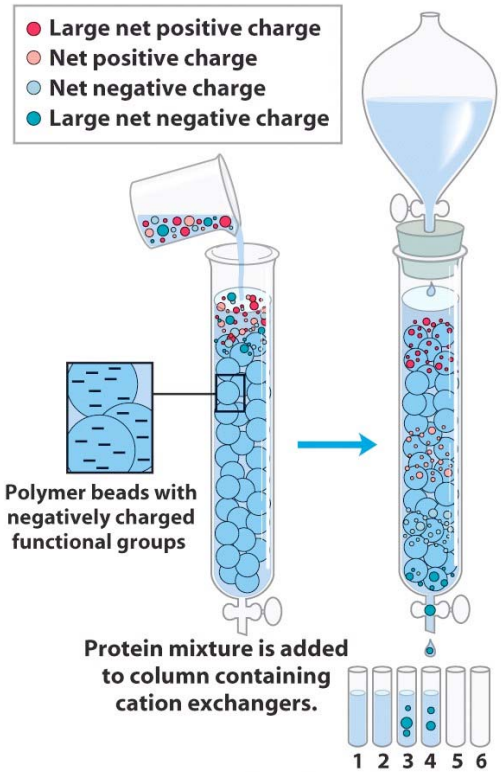




*>>> The charged groups, hydrophobic regions, size, and solvation affect the biophysical properties of the protein and largely determine its purification behavior.*

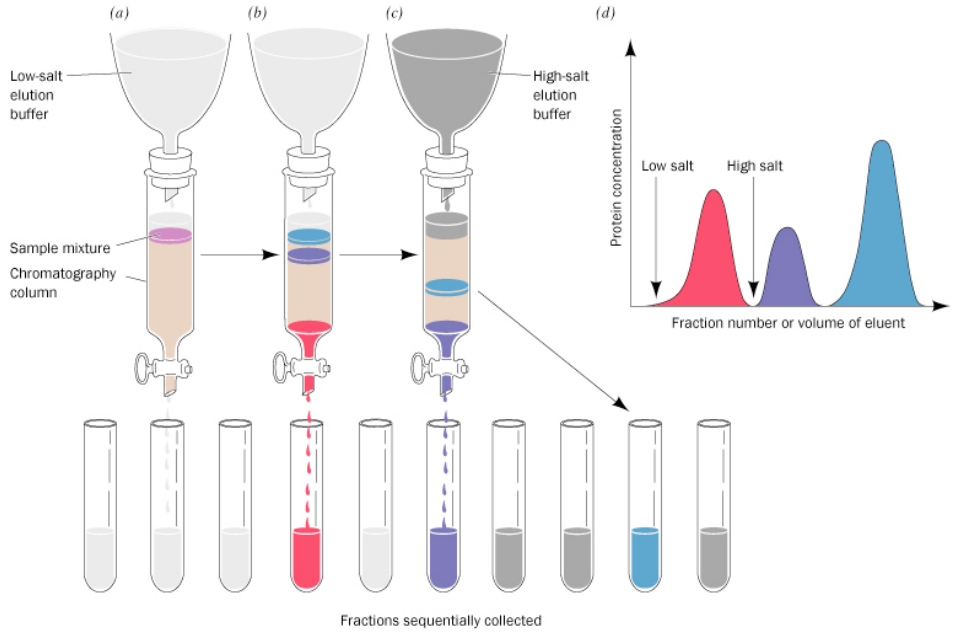
# Different Types of Chromatography

- Gel filtration/size exclusion - separates by size (molecular weight) of proteins
- Ion exchange (cation exchange and anion exchange) - separates by surface charge on proteins
  - Cation exchange: separates based on positive charges of solutes/proteins, matrix is negatively charged
  - Anion exchange: separates based on negative charges of solutes/proteins, matrix is positively charged
- Hydrophobic interaction - separates by hydrophobicity of proteins
- Affinity - separates by some unique binding characteristic of protein of interest for affinity matrix in column

# Ion-Exchange Chromatography




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.



# Ion Exchange (IEX) Chromatography

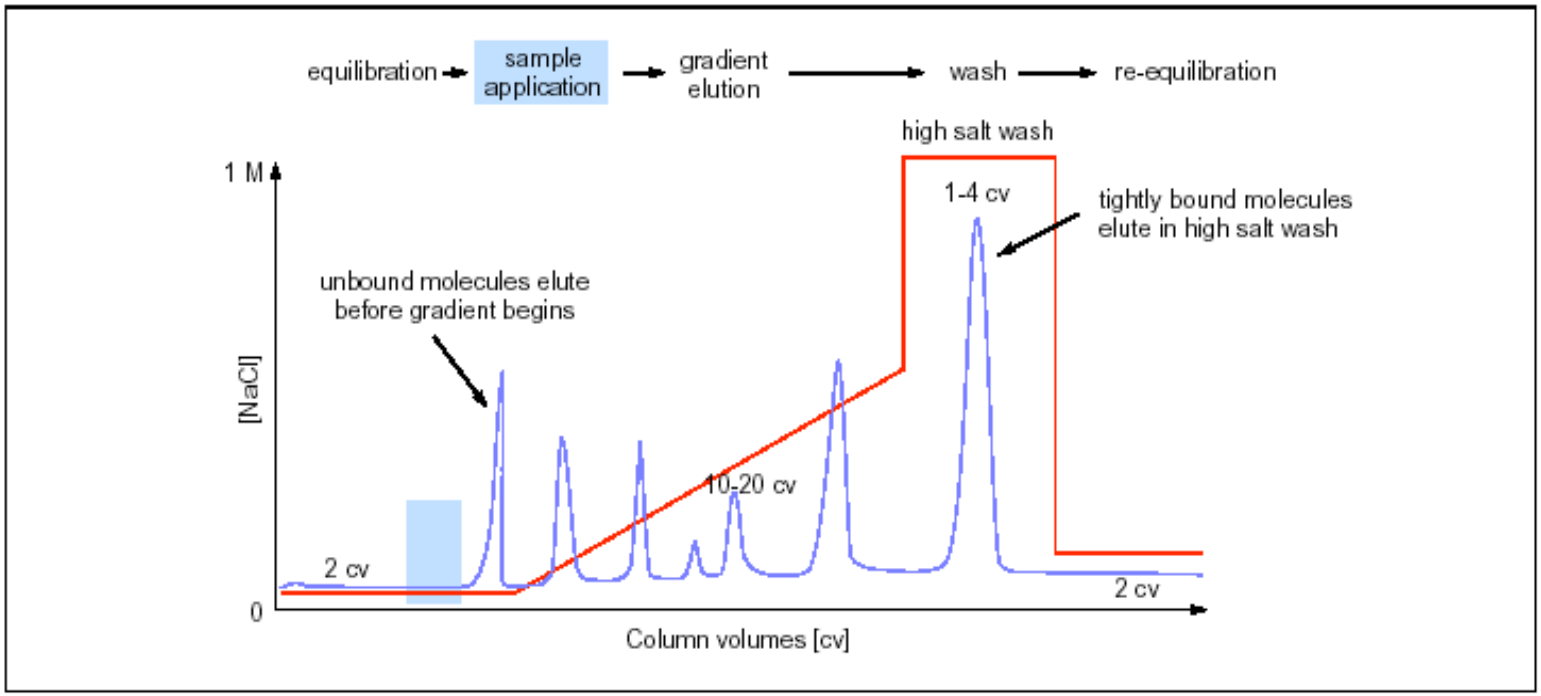
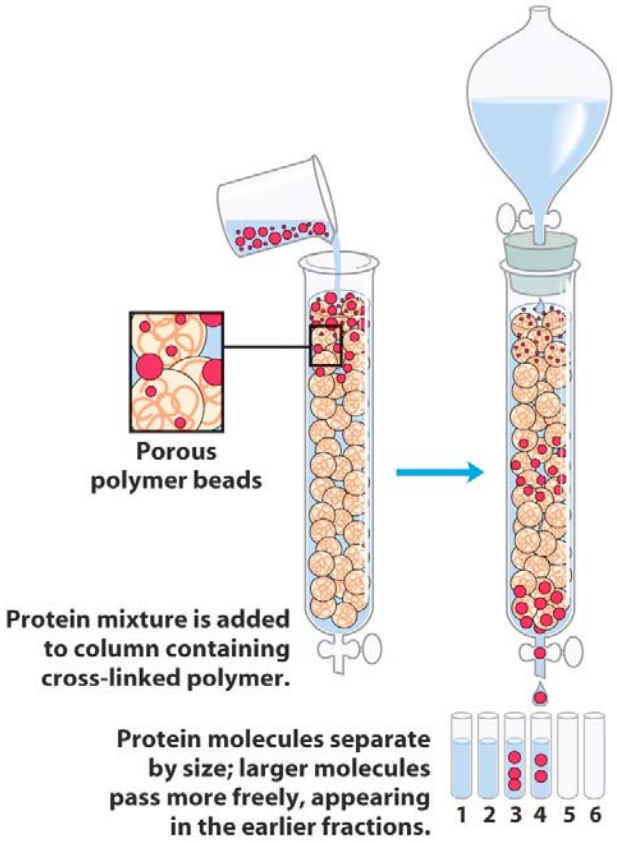
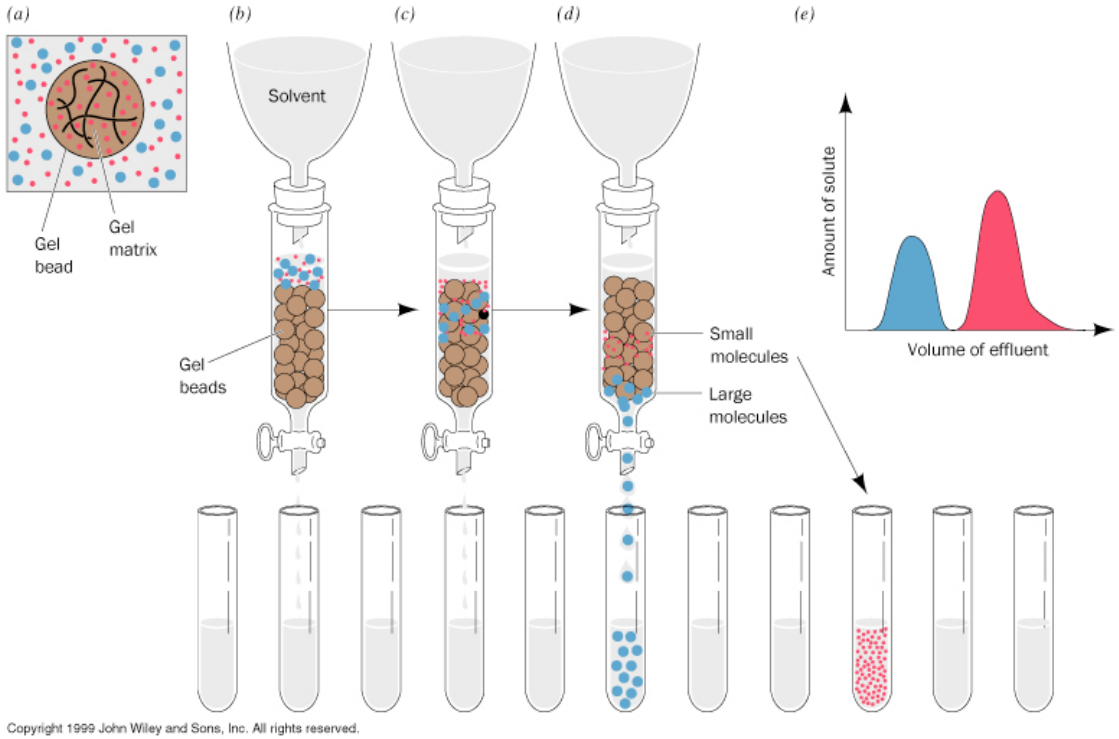



**Fig. 31.** Typical IEX gradient elution.

The net surface charge of proteins varies according to the surrounding pH. When above its isoelectric point (pI) a protein will bind to an anion exchanger, when below its pI a protein will behind to a cation exchanger. Typically IEX is used to bind the target molecule, but it can also be used to bind impurities if required. IEX can be repeated at different pH values to separate several proteins which have distinctly different charge properties, as shown in Figure 32. This can be used to advantage during a multi-step purification, as shown in the example on page 24.

# Gel Filtration Chromatography



# Gel Filtration (GF) Chromatography

GF separates proteins with differences in molecular size. The technique is ideal for the final polishing steps in a purification when sample volumes have been reduced (sample volume significantly influences speed and resolution in gel filtration). Samples are eluted isocratically (single buffer, no gradient Figure 41). Buffer conditions are varied to suit the sample type or the requirements for further purification, analysis or storage step, since buffer composition does not directly affect resolution. Proteins are collected in purified form in the chosen buffer.

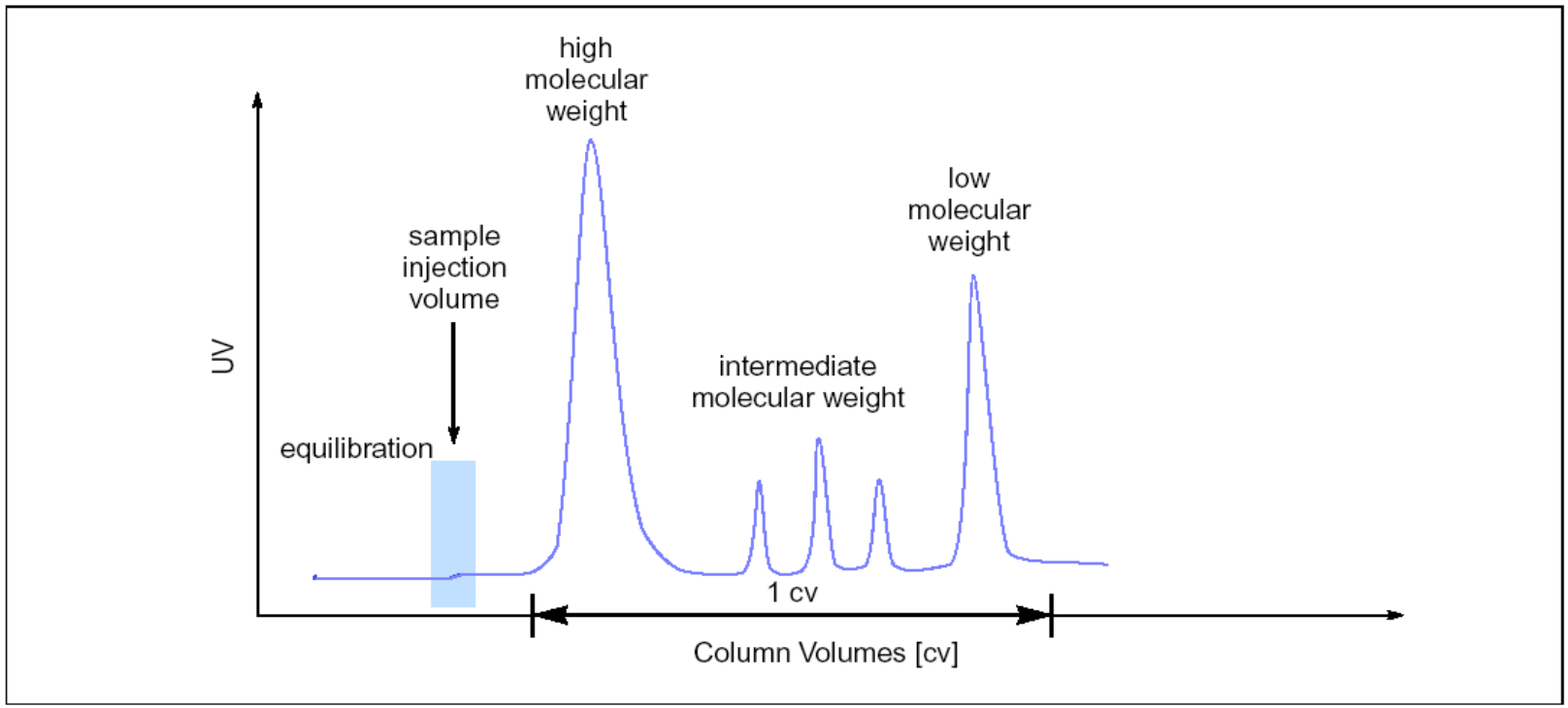



**Fig. 41.** Typical GF elution.

# The principle of gel filtration -- excluded volume

**[Note: gel filtration chromatography is also sometimes called "size exclusion chromatography"]**

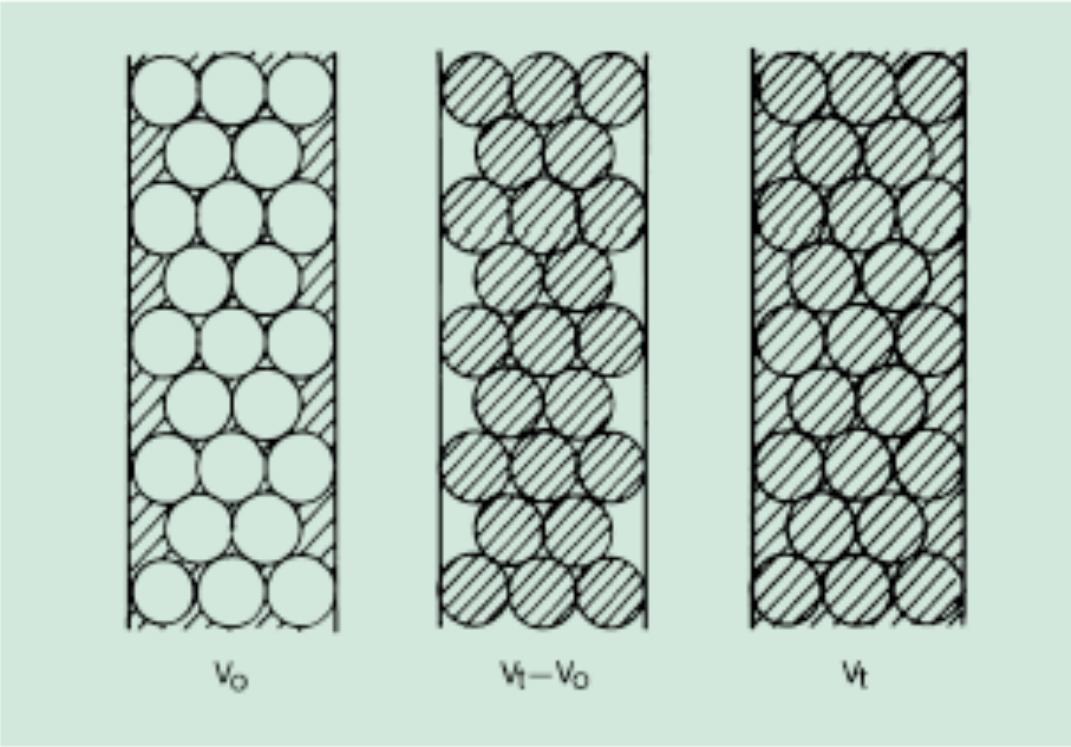
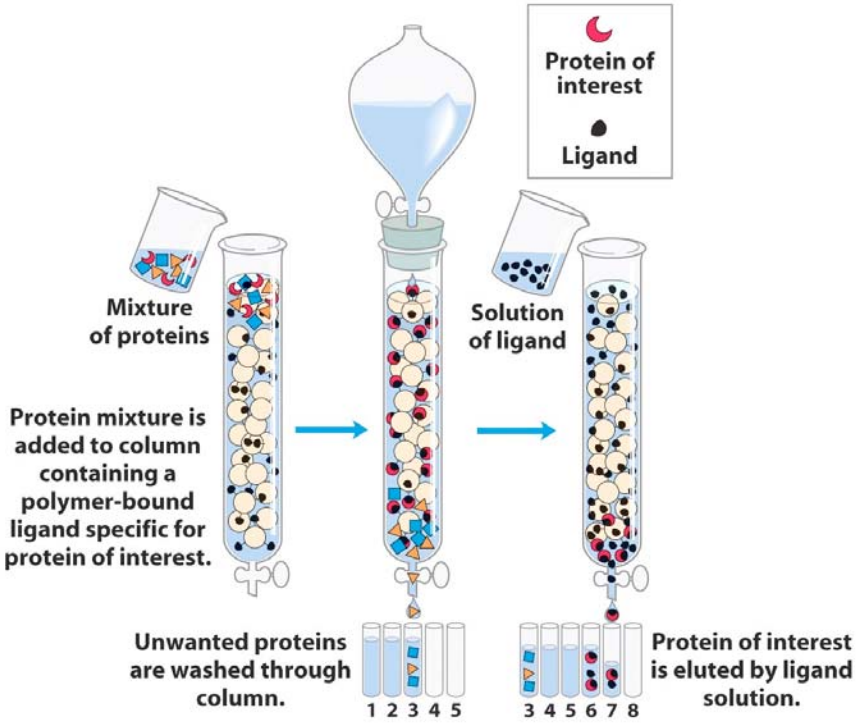
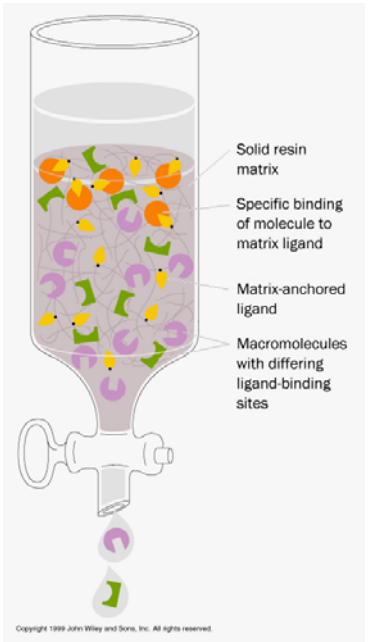


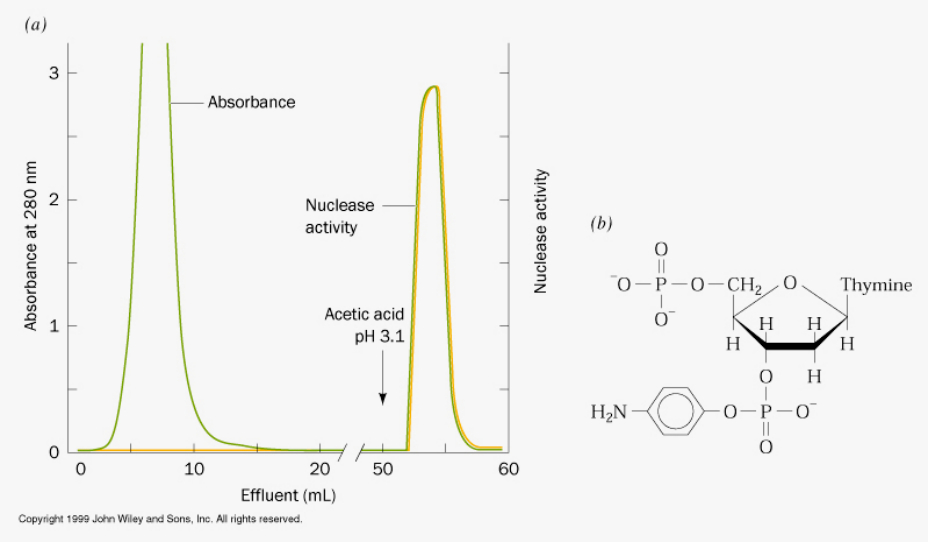

Fig. 3. Diagrammatic representation of $V_t$ and $V_o$. Note that $V_t$-$V_o$ will include the volume of the solid material which forms the matrix. (Fischer, L., Laboratory Techniques in Biochemistry and Molecular Biology. Vol. 1 part II. An introduction to Gel Chromatography. North Holland Publishing Company, Amsterdam. Reproduced by kind permission of the Authors and the Publisher).

**$V_o$ = "void volume"**
**$V_t$ = "bed volume"**
**$V_e$ = "elution volume"**
**$V_i = V_t - V_o$**

# Affinity Chromatography



# Protein Sequencing

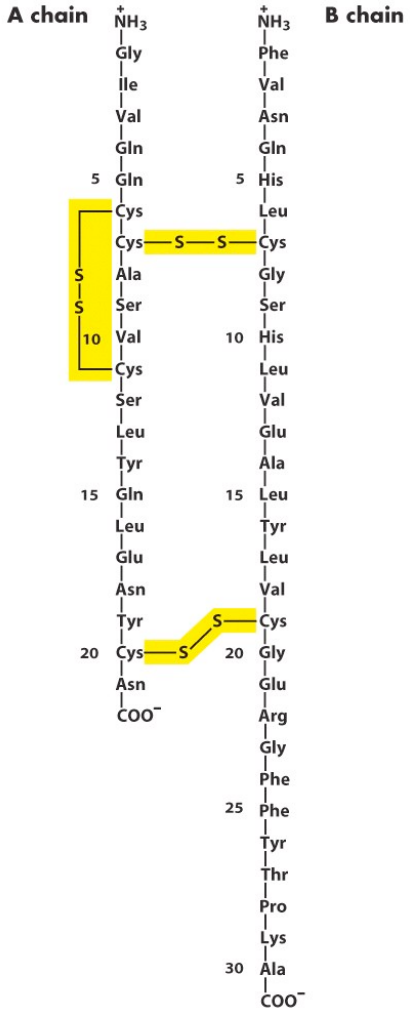
*Frederick Sanger was the first.*

*In 1953, he sequenced the two chains of insulin.*

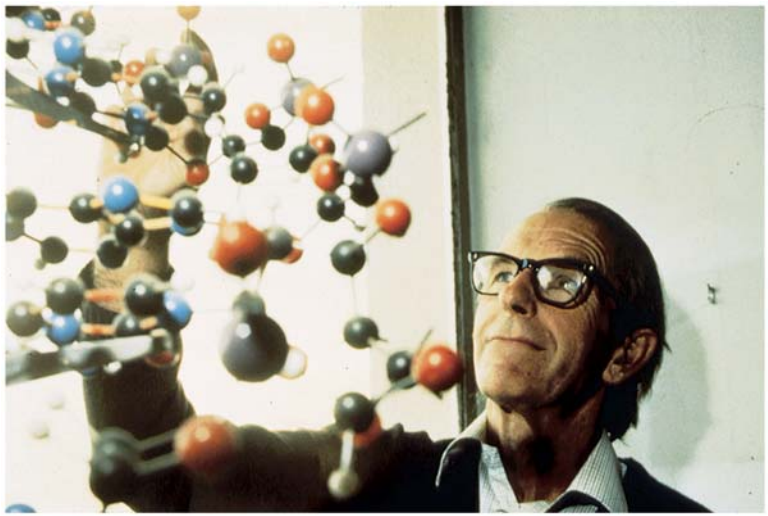
- Sanger's results established that all of the molecules of a given protein have the same, unique sequence and that the polypeptide chain is unbranched (apart from disulfide crosslinks at some cysteines).

- Proteins can be sequenced in two ways:
  - *real* amino acid sequencing (the “classic” approach)
  - sequencing the corresponding DNA of the gene (or cDNA copy of the mRNA), then inferring the protein sequence using the genetic code.

# Primary Structure of Bovine Insulin




First protein to be fully sequenced; Fred Sanger, 1953. For this, he won his first Nobel Prize (his second was for the Sanger dideoxy method of DNA sequencing).



# Direct Determination of a Protein Sequence
## *An Eight Step Strategy*

- 1. If more than one polypeptide chain, separate.
- 2. Cleave (reduce) disulfide bridges
- 3. Determine composition of each chain
- 4. Determine N- and C-terminal residues
- 5. Cleave each chain into smaller fragments and determine the sequence of each chain
- 6. Repeat step 5, using a different cleavage procedure to generate a different set of fragments
- 7. Reconstruct the sequence of the protein from the sequences of overlapping fragments
- 8. Determine the positions of the disulfide crosslinks

## Comment re previous slide:

*Obviously, it is easier to just sequence the gene for a protein rather than laboriously sequence the protein itself by direct methods. This is why most (>99%) of protein sequences generated today are from indirect, inferential nucleic acid sequencing data. However, for special cases, e.g., characterizing a new protein isolated from a natural source, direct sequencing is still required. Even in this case, though, often only partial sequencing is required -- enough to generate a unique peptide "fingerprint" -- then the rest of the sequence can be obtained by looking it up in a database of protein sequences generated from DNA sequencing data from the organism from which the protein was isolated.*

Example of inferring protein sequence & function from DNA sequence: the human genome data.

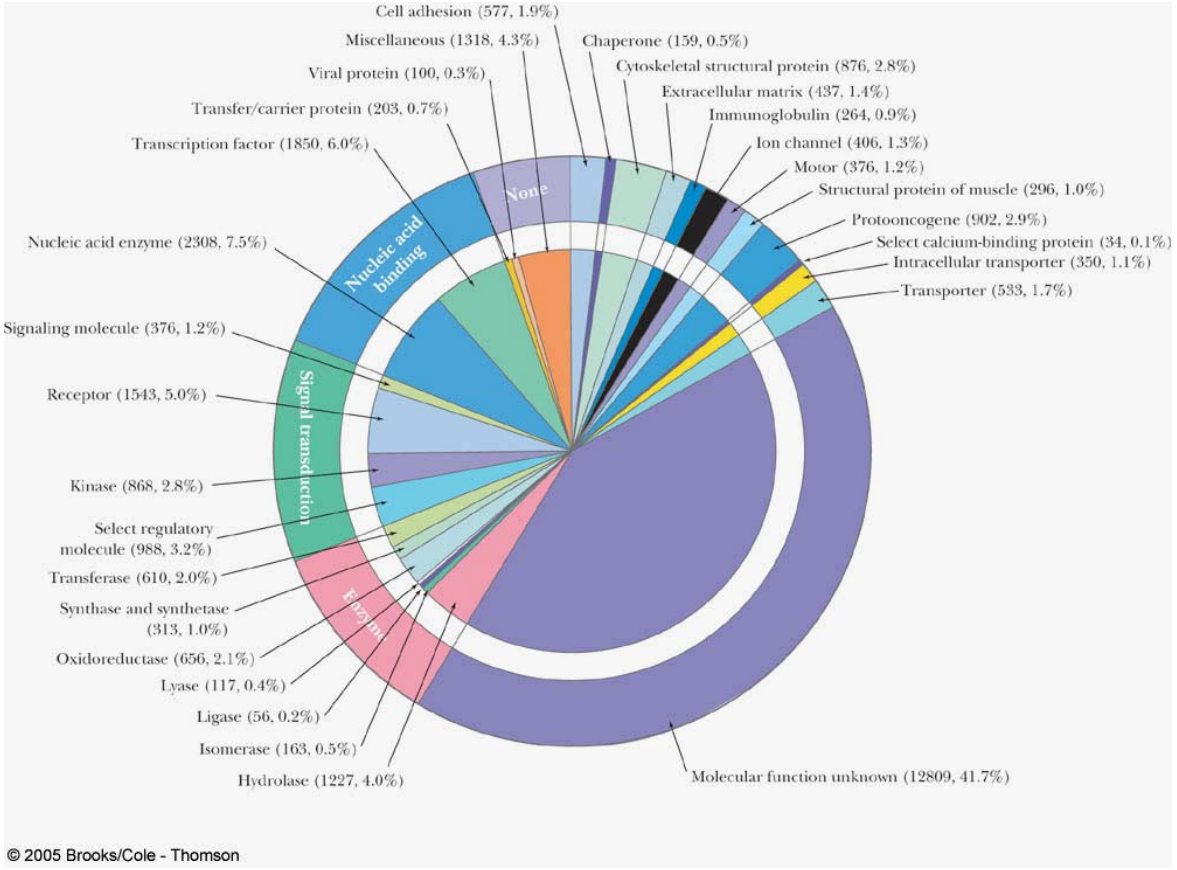



Figure 5.32 Proteins of the human genome grouped according to their molecular function. The numbers and percentages within each functional category are enclosed in parentheses. Note that the function of more than 40% of the proteins encoded by the human genome remains unknown. Considering those of known function, enzymes (including kinases and nucleic acid enzymes) account for about 20% of the total number of proteins; nucleic acid-binding proteins of various kinds, about 14%, among which almost half are gene-regulatory proteins (transcription factors). Transport proteins collectively constitute about 5% of the total; and structural proteins, another 5%. *(Adapted from figure 15 in Venter, J.C., et al., 2001. The sequence of the human genome.* Science ***291:****1304-1351.)*

# Cleavage of Polypeptides for Analysis

- Strong acid (*e.g.* 6M HCl) - not sequence specific
- Sequence-specific proteolytic enzymes (proteases)
- Sequence-specific chemical cleavage (e.g. cyanogen bromide cleavage at methionine residues)

# Protease Specificities

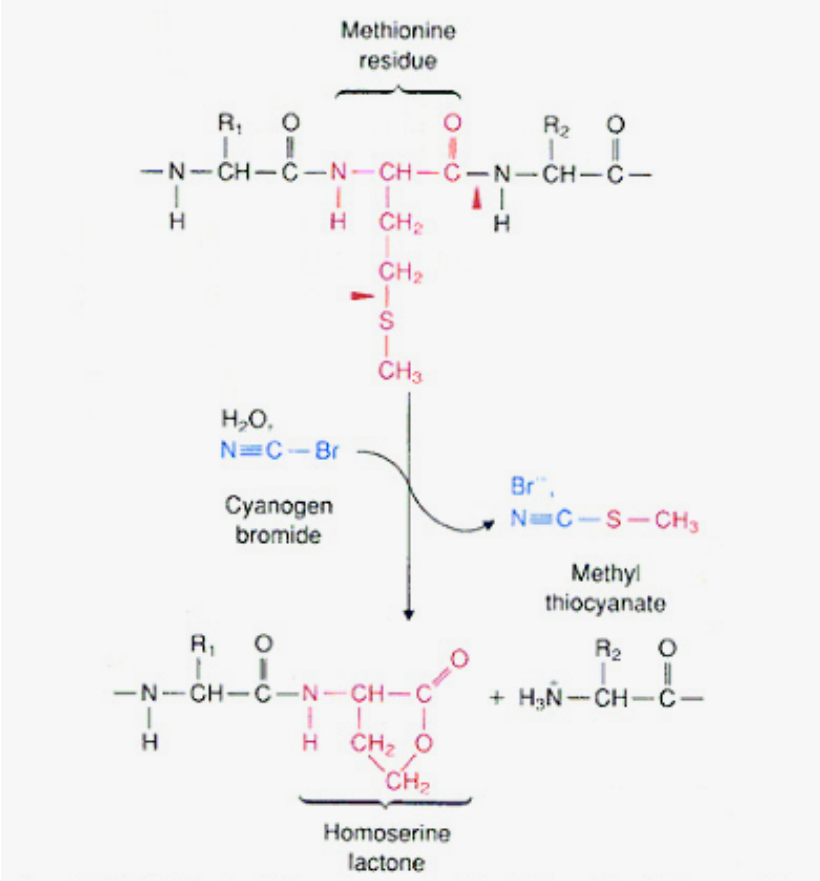
**TABLE 3–7** The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

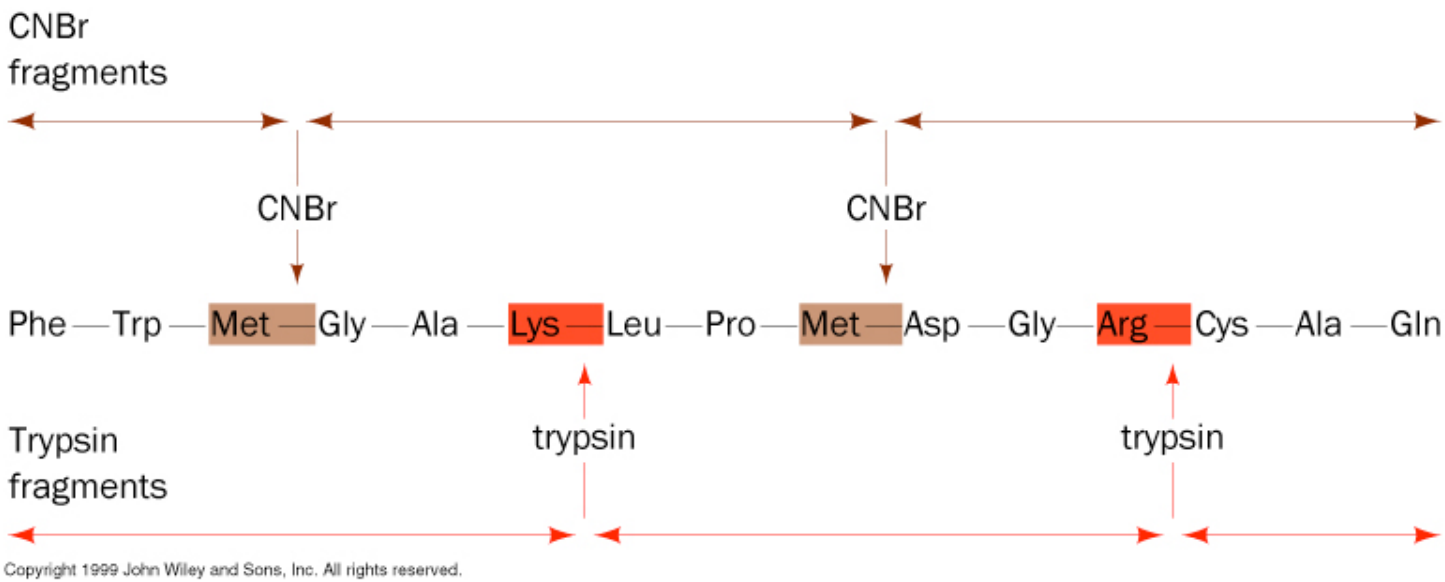
| *Reagent (biological source)** | *Cleavage points*[†] |
|---|---|
| Trypsin (bovine pancreas) | Lys, Arg (C) |
| *Submaxillarus* protease (mouse submaxillary gland) | Arg (C) |
| Chymotrypsin (bovine pancreas) | Phe, Trp, Tyr (C) |
| *Staphylococcus aureus* V8 protease (bacterium *S. aureus*) | Asp, Glu (C) |
| Asp-*N*-protease (bacterium *Pseudomonas fragi*) | Asp, Glu (N) |
| Pepsin (porcine stomach) | 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.

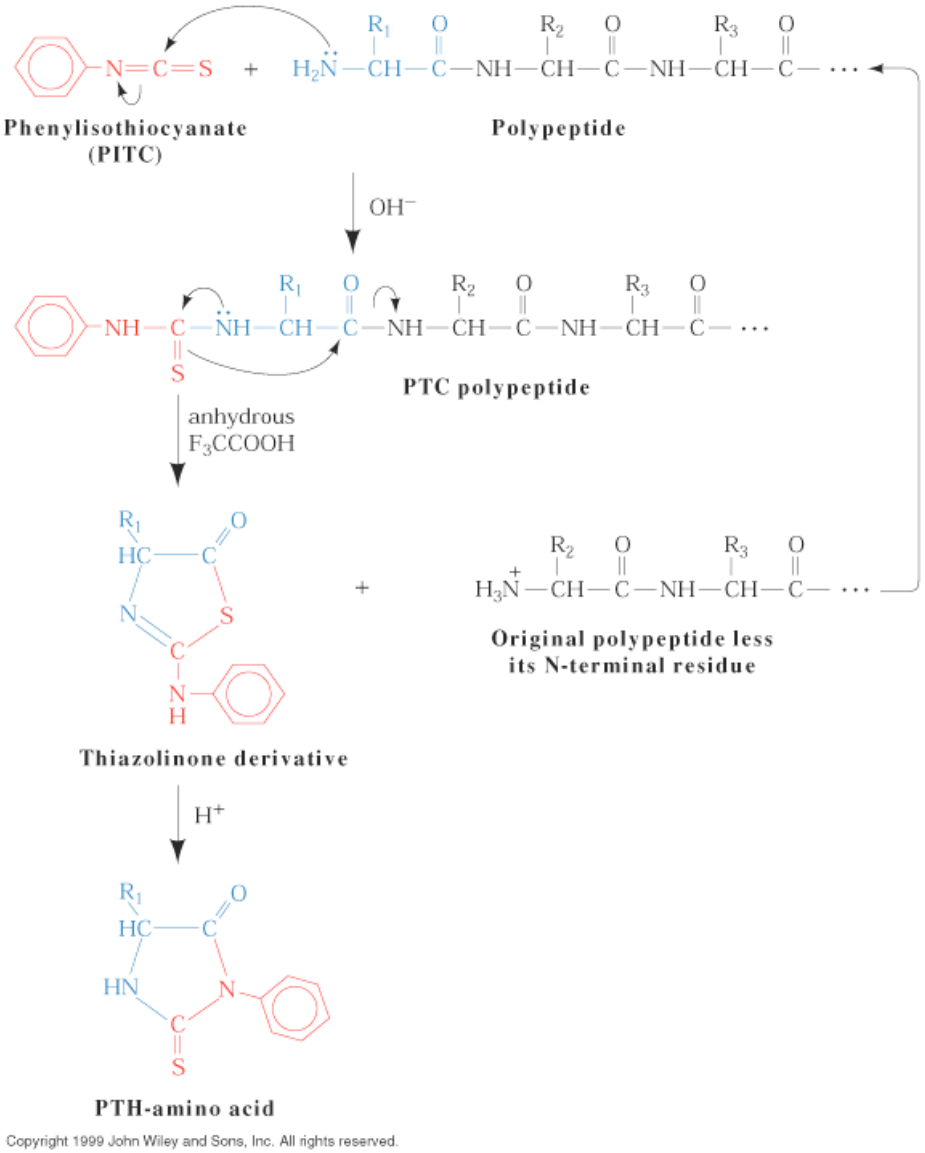
[†]Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

# Cyanogen Bromide Cleavage at Methionine Residues





# Protein Sequencing: Edman Degradation

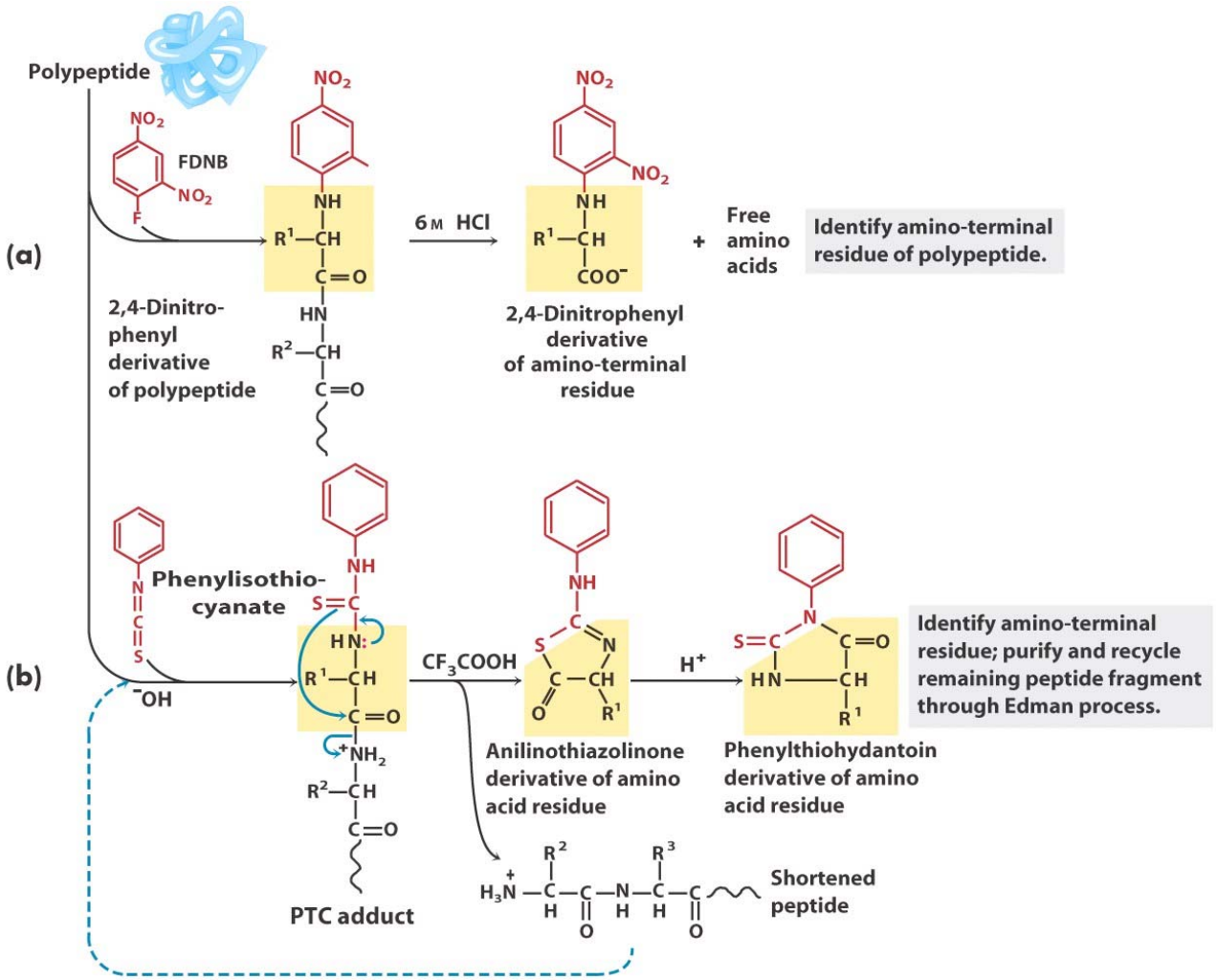




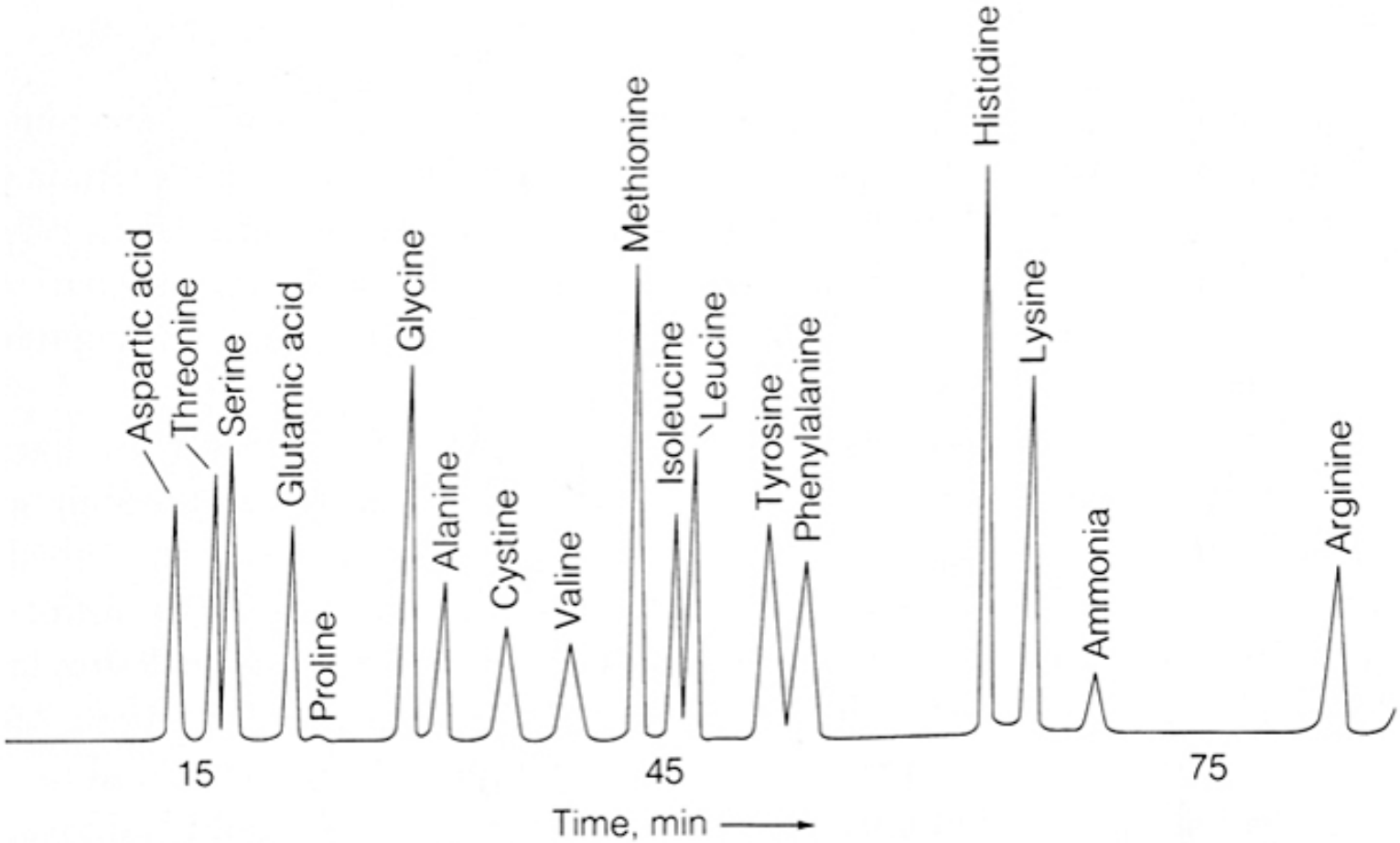
PTC = phenylthiocarbamyl

$F_3CCOOH$ = trifluoroacetic acid

PTH = phenylthiohydantion

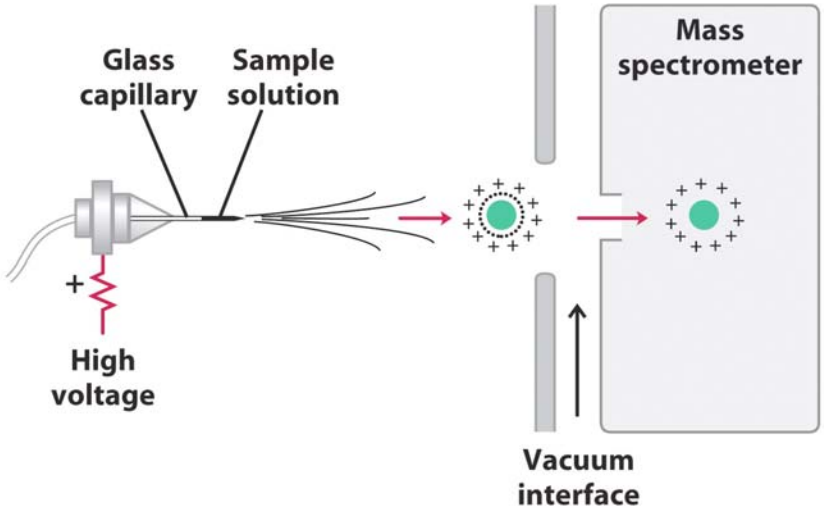

Polypeptide
FDNB
(a)
2,4-Dinitro-phenyl derivative of polypeptide
6 M HCl
2,4-Dinitrophenyl derivative of amino-terminal residue
+ Free amino acids
Identify amino-terminal residue of polypeptide.
(b)
Phenylisothio-cyanate
CF3COOH
H+
Anilinothiazolinone derivative of amino acid residue
Phenylthiohydantoin derivative of amino acid residue
Identify amino-terminal residue; purify and recycle remaining peptide fragment through Edman process.
PTC adduct
Shortened peptide

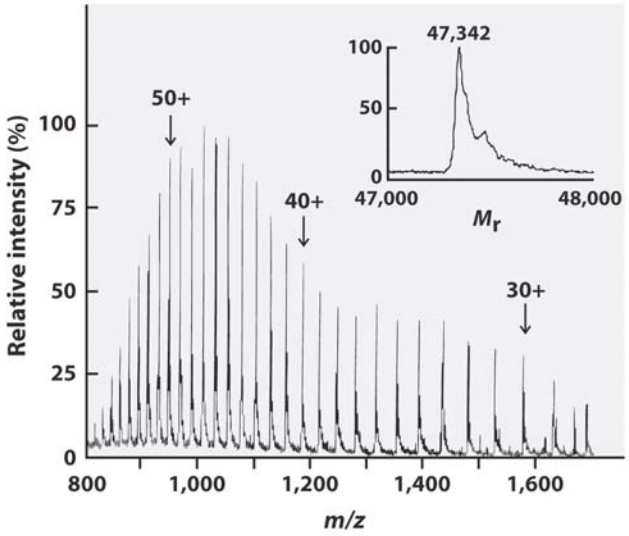
# Separation of Amino Acids by HPLC

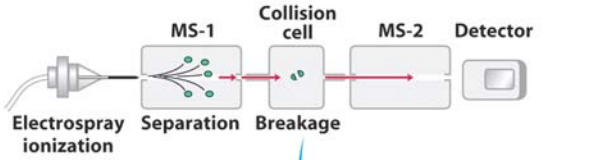


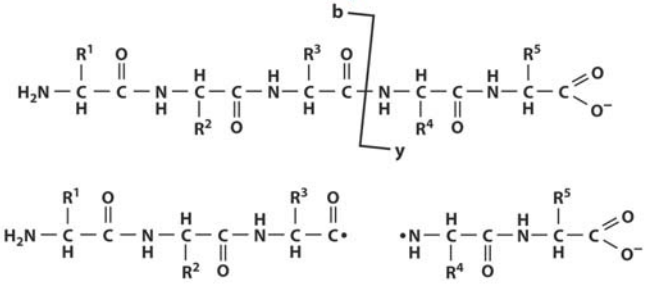
*Another, more modern way, to sequence proteins is via mass spectrometry. The mass spectrometry method also meshes well with modern proteomics databases where one can have "look up" tables of peptides based on their mass. Thus mass spectrometry can facilitate peptide fingerprinting and subsequent gene product identification. Mass spectrometry can also be done on very small amounts of material (even impure material), which is another advantage.*

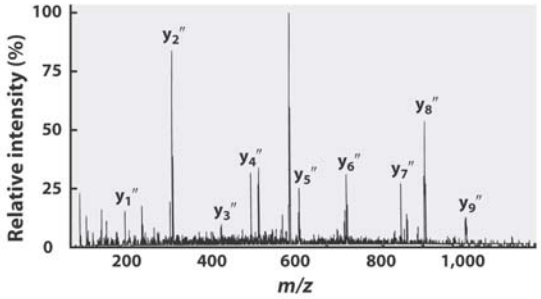
# Protein/Peptide Identification by Mass Spectrometry



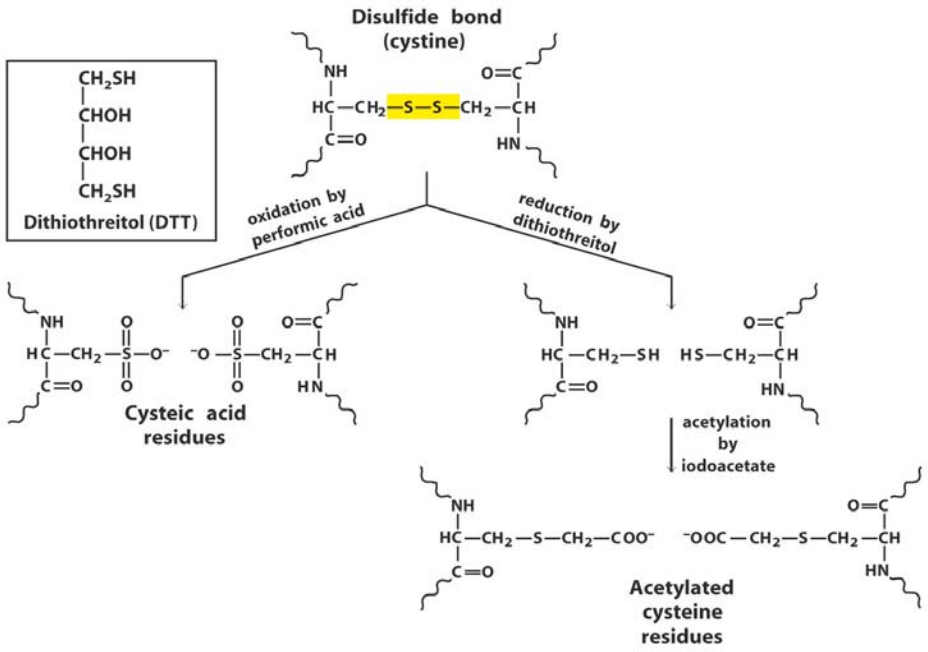



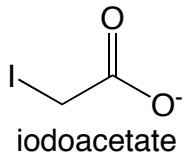



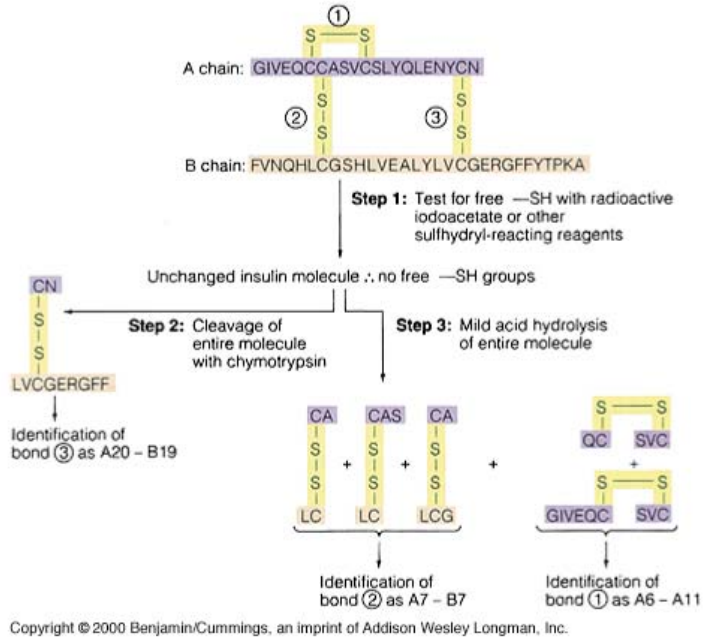



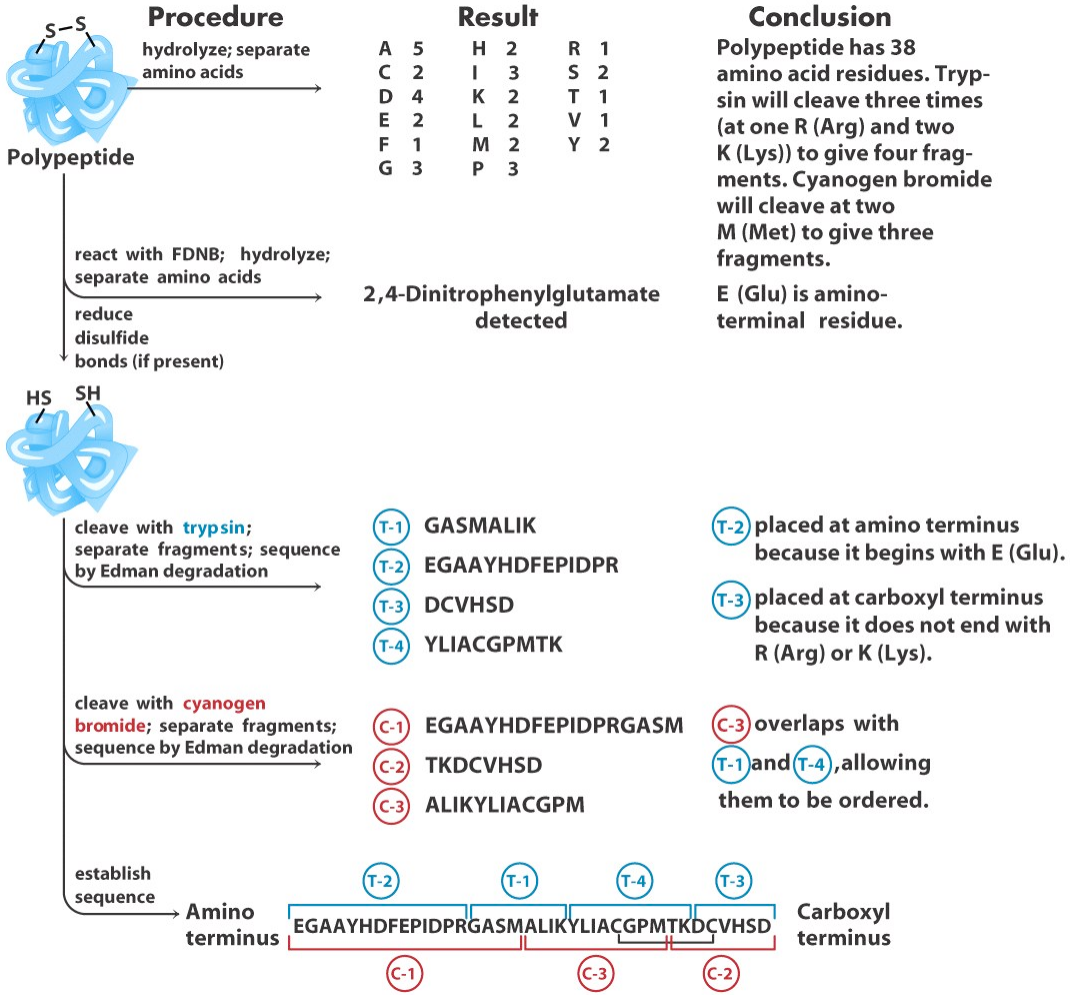
# Locating the Disulfide Bonds in Insulin



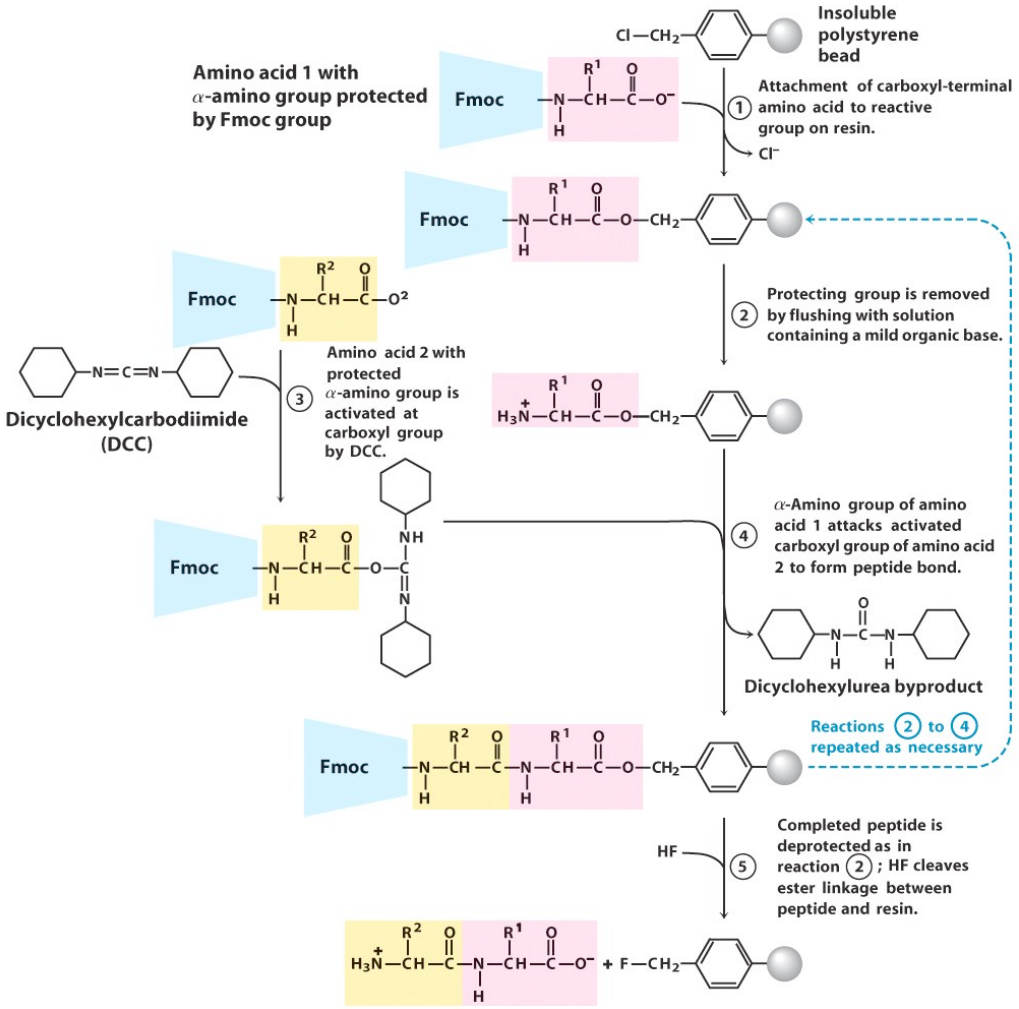



# Determining Primary Structure of an Entire Protein



# Reactions in Solid-Phase Peptide Synthesis



# Molecular Biology & Biochemistry 694:407 & 115:511

## *Methods of Protein Analysis*

**Sept. 16th, 2005, Lecture**

R. Bruce Merrifield

R. Bruce Merrifield

*Special thanks for many slides in this lecture again goes to Dr. Gabriel Fenteany, Dept. of Chemistry, University of Illinois at Chicago (www.chem.uic.edu/fenteany/teaching/452).*