How To Sequence A Protein ?

Preliminary Steps:

- Break interchain disulfide bonds, if necessary
- Two reagents are commonly used:
  - performic acid
  - Mercaptoethanol
- For multisubunit proteins, the individual protein chains must be separated by disrupting noncovalent interchain interactions
- Individual protein chains are separated by electrophoresis or chromatography on the basis of size or charge.

Determining Amino Acid Sequence

Once each protein is purified the amino acid sequence is determined by:

1) determining the amino acid composition (how many of each amino acid are in the protein)
2) identifying the amino and carboxyl terminal amino acids
3) cleaving the protein into two or more sets of peptides using specific enzymatic or chemical reagents such as trypsin or cyanogen bromide
4) determining the amino acid sequence of each of the peptide fragments
5) determining the entire protein sequence from the sequences of overlapping peptide fragments
6) locating the position of disulfide bridges between cysteines

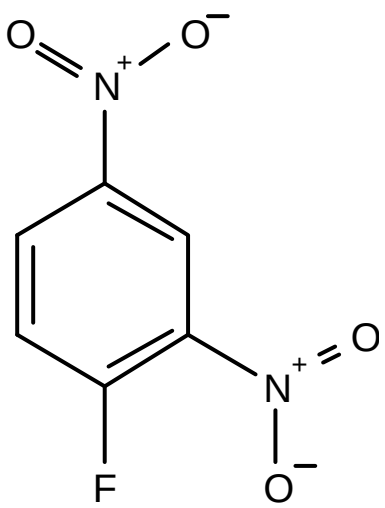
Determining Amino Acid Composition

- Hydrolysis (6M HCl, 2M TFA)
  - Destroys W
  - Partially destroys S, T
  - Loss of amides N$\rightarrow$D, Q $\rightarrow$E (Asx, Glx)
  - Incomplete cleavage hydrophobic residues
- Separating and quantifying individual amino acids by ion exchange HPLC using an amino acid analyzer

Determining the N-Terminal Amino Acid

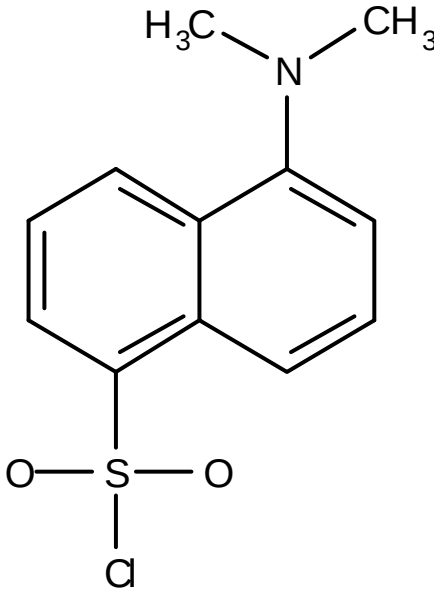
- The N-terminal amino acid is determined using either chemical reagents or enzymes
- Chemical reagents include:
  - Sanger's reagent
  - dansyl chloride
  - Edman Degradation

Determining the N-Terminal Amino Acid



- Sanger's reagent
- Treat with dinitrofluorobenzene to form a dinitrophenyl (DNP) derivative of the amino-terminal amino acid
- Acid hydrolysis
- Extract the DNP-derivative from the acid hydrolysate with organic solvent
- Identify the DNP-derivative by chromatography and comparison with standards

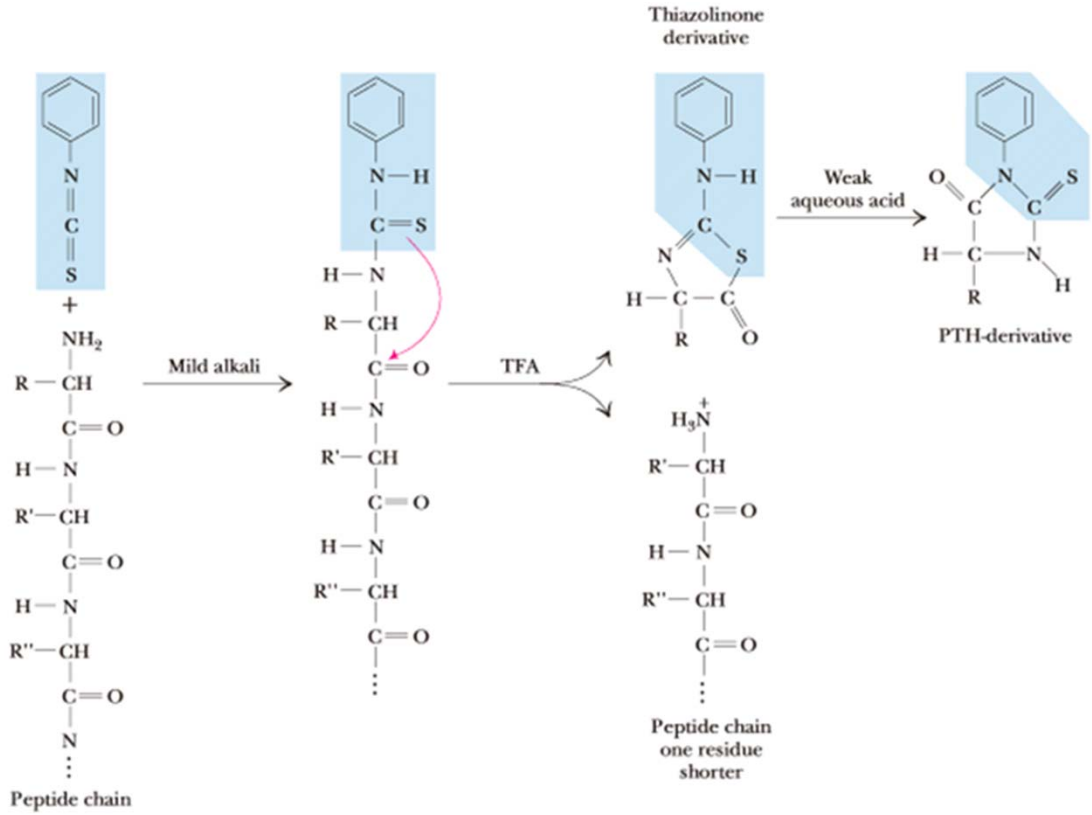
Determining the N-Terminal Amino Acid



- Dansyl chloride (dimethylaminonaphthylenesulfonyl chloride)
- Forms a highly fluorescent derivative of the amino-terminal amino acid
- Identified by chromatography and fluorescence detection after acid hydrolysis
- Highly senstive
- Best choice when the amount of protein is limited

Edman Degradation

- phenylisothiocyanate (phenyl-N=C=S) adds to N-terminus then acid treatment cleaves the N-terminal amino acid as a PTH derivative
- the remaining protein chain is intact and the cycle can be repeated
- under ideal conditions the sequence of 30-60 amino acids can be determined



C-terminal analysis

- Enzymatic analysis (carboxypeptidase)
  - Carboxypeptidase A cleaves any residue except Pro, Arg, and Lys
  - Carboxypeptidase B (hog pancreas) only works on Arg and Lys
  - Carboxypeptidase C, any residue

Peptide Fragments

- After determining the amino acid composition and the N & C-terminal amino acids, at least two different sets of protein fragments are needed for sequencing
- The sequencing methods currently available are only accurate for peptides up to about 20-30 amino acids, 60 under ideal conditions
- Reagents that cleave the protein chain only at a few specific sites forming fragments that are less than 20-30 amino acids in length

Why 2 Sets of Fragments?

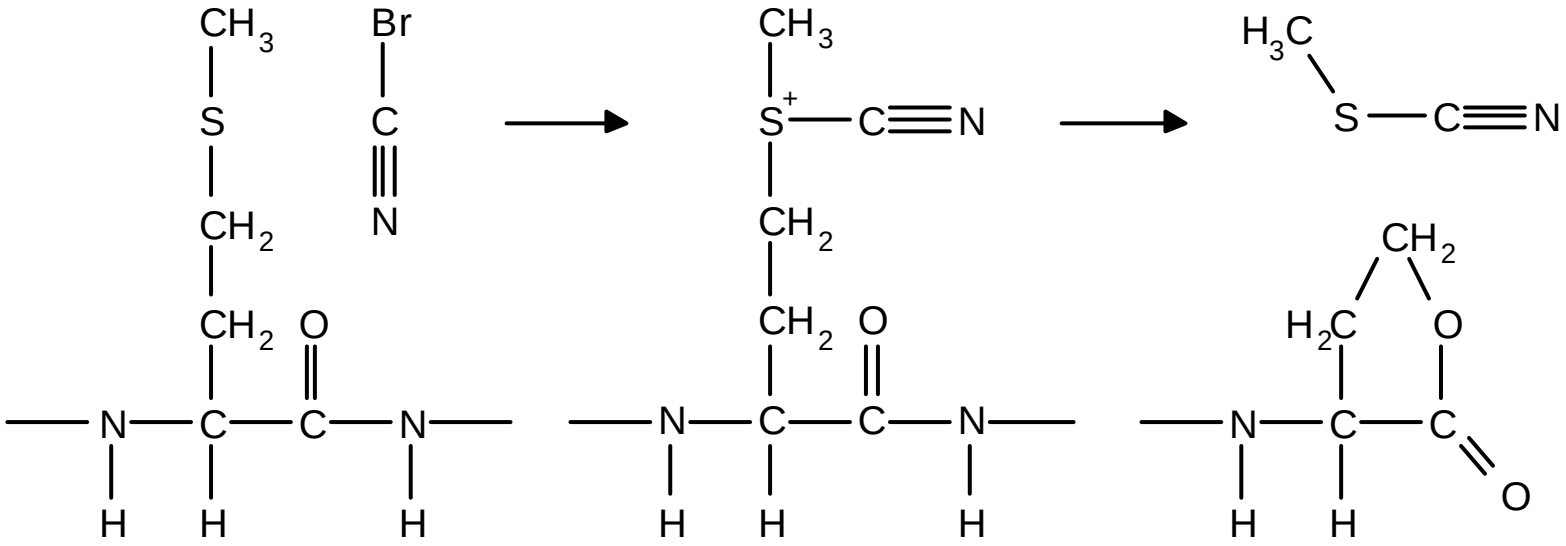
- Why can't the entire protein amino acid sequence be determined from a single set of peptide fragments obtained by cleavage with a single reagent?
- There’s no way to determine how the fragments are connected with just one set
- A second or third set of fragments are used to deduce how the fragments are connected by identification and comparison of overlapping sequnces

Protein Cleavage Reagents

- Chemical or enzymatic reagents can be used to prepare protein fragments
- The most commonly used reagents are:
  - cyanogen bromide (Met→homoserine lactone)
  - various enzymes including
    - Trypsin (R or K)
    - Chymotrypsin (F or Y or W)
    - Clostripain (R>K) (can be incomplete for K)
    - Staphylococcal protease (D or E)
    - various endopeptidases

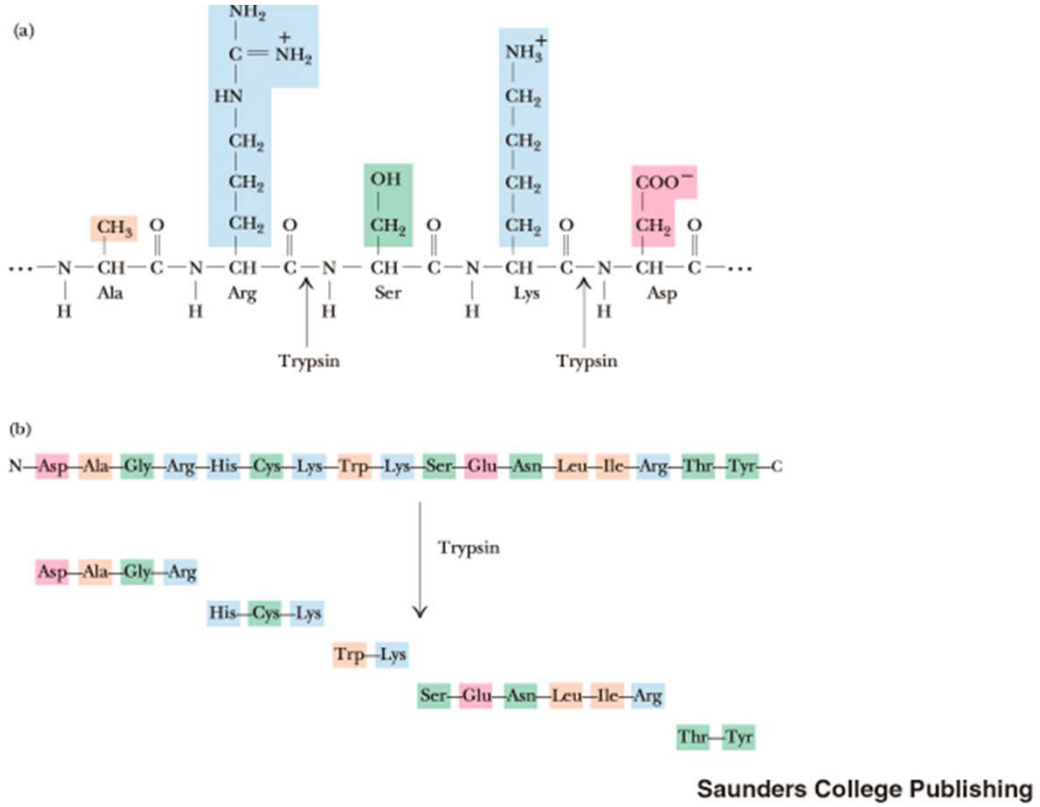
Cyanogen Bromide

- It cleaves protein chains at internal methionines by reaction with the methionine sulfur as illustrated below



Enzymatic fragmentation

- Trypsin (R or K)
- Chymotrypsin (F or Y or W)
- Clostripain (R>K) (can be incomplete for K)
- Staphylococcal protease (D or E)



Endopeptidases

- The following endopeptidases are less specific than the enzymes metioned above
- Pepsin, papain, subtilisin, thermolysin, elastase
  - (papain is the active ingredient in meat tenderizer, soft contact cleansing solutions, some laundry detergents)
- These enzymes are most often used to further reduce the size of large tryptic or chymotryptic fragments

How To Sequence A Protein ?

Preliminary Steps:

- Break interchain disulfide bonds, if necessary
- Two reagents are commonly used:
  - performic acid
  - Mercaptoethanol
- For multisubunit proteins, the individual protein chains must be separated by disrupting noncovalent interchain interactions
- Individual protein chains are separated by electrophoresis or chromatography on the basis of size or charge.

Determining Amino Acid Sequence

Once each protein is purified the amino acid sequence is determined by:

1) determining the amino acid composition (how many of each amino acid are in the protein)
2) identifying the amino and carboxyl terminal amino acids
3) cleaving the protein into two or more sets of peptides using specific enzymatic or chemical reagents such as trypsin or cyanogen bromide
4) determining the amino acid sequence of each of the peptide fragments
5) determining the entire protein sequence from the sequences of overlapping peptide fragments
6) locating the position of disulfide bridges between cysteines

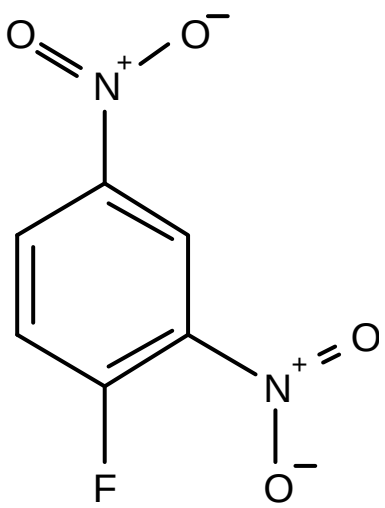
Determining Amino Acid Composition

- Hydrolysis (6M HCl, 2M TFA)
  - Destroys W
  - Partially destroys S, T
  - Loss of amides N$\rightarrow$D, Q $\rightarrow$E (Asx, Glx)
  - Incomplete cleavage hydrophobic residues
- Separating and quantifying individual amino acids by ion exchange HPLC using an amino acid analyzer

Determining the N-Terminal Amino Acid

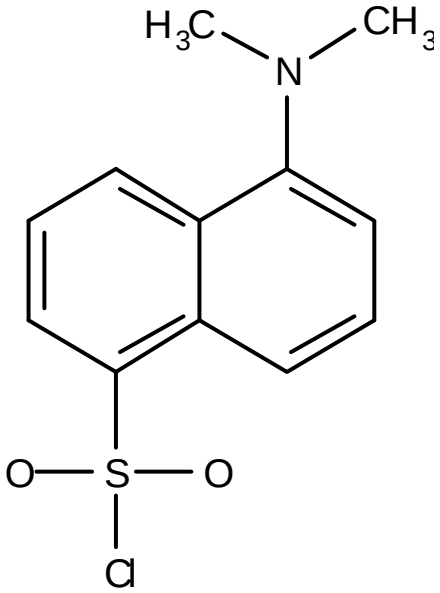
- The N-terminal amino acid is determined using either chemical reagents or enzymes
- Chemical reagents include:
  - Sanger's reagent
  - dansyl chloride
  - Edman Degradation

Determining the N-Terminal Amino Acid



- Sanger's reagent
- Treat with dinitrofluorobenzene to form a dinitrophenyl (DNP) derivative of the amino-terminal amino acid
- Acid hydrolysis
- Extract the DNP-derivative from the acid hydrolysate with organic solvent
- Identify the DNP-derivative by chromatography and comparison with standards

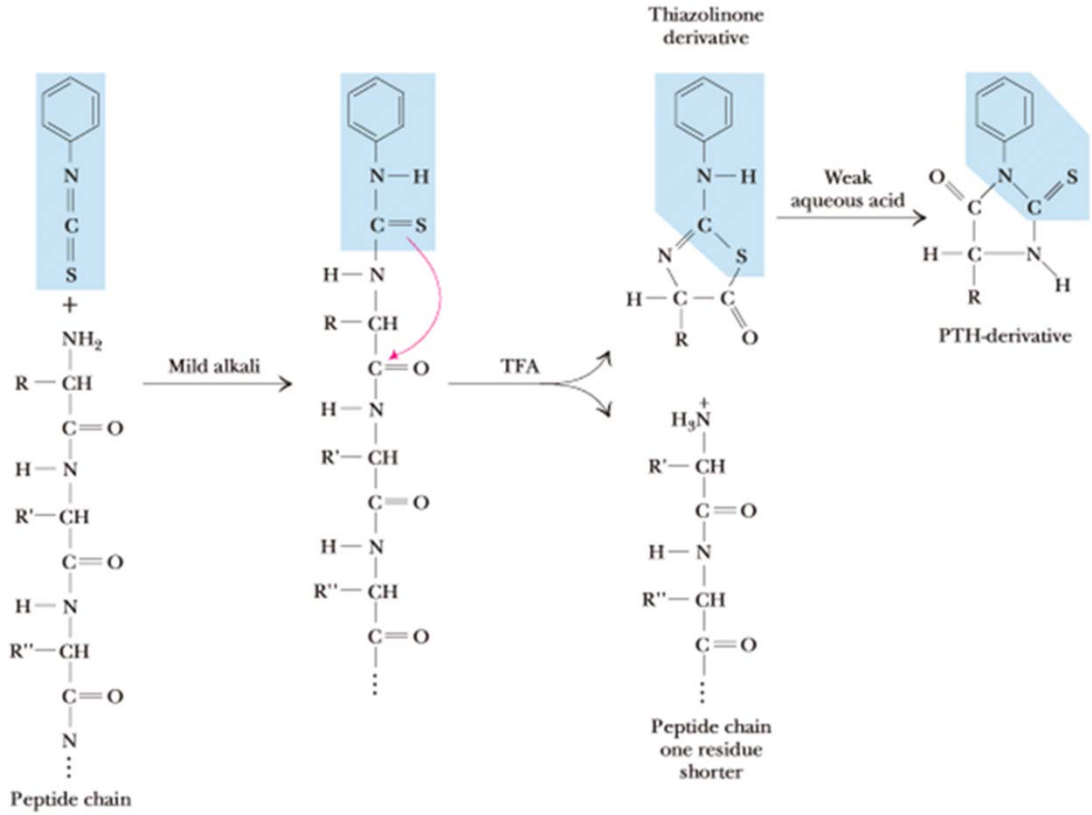
Determining the N-Terminal Amino Acid



- Dansyl chloride (dimethylaminonaphthylenesulfonyl chloride)
- Forms a highly fluorescent derivative of the amino-terminal amino acid
- Identified by chromatography and fluorescence detection after acid hydrolysis
- Highly senstive
- Best choice when the amount of protein is limited

Edman Degradation

- phenylisothiocyanate (phenyl-N=C=S) adds to N-terminus then acid treatment cleaves the N-terminal amino acid as a PTH derivative
- the remaining protein chain is intact and the cycle can be repeated
- under ideal conditions the sequence of 30-60 amino acids can be determined



C-terminal analysis

- Enzymatic analysis (carboxypeptidase)
  - Carboxypeptidase A cleaves any residue except Pro, Arg, and Lys
  - Carboxypeptidase B (hog pancreas) only works on Arg and Lys
  - Carboxypeptidase C, any residue

Peptide Fragments

- After determining the amino acid composition and the N & C-terminal amino acids, at least two different sets of protein fragments are needed for sequencing
- The sequencing methods currently available are only accurate for peptides up to about 20-30 amino acids, 60 under ideal conditions
- Reagents that cleave the protein chain only at a few specific sites forming fragments that are less than 20-30 amino acids in length

Why 2 Sets of Fragments?

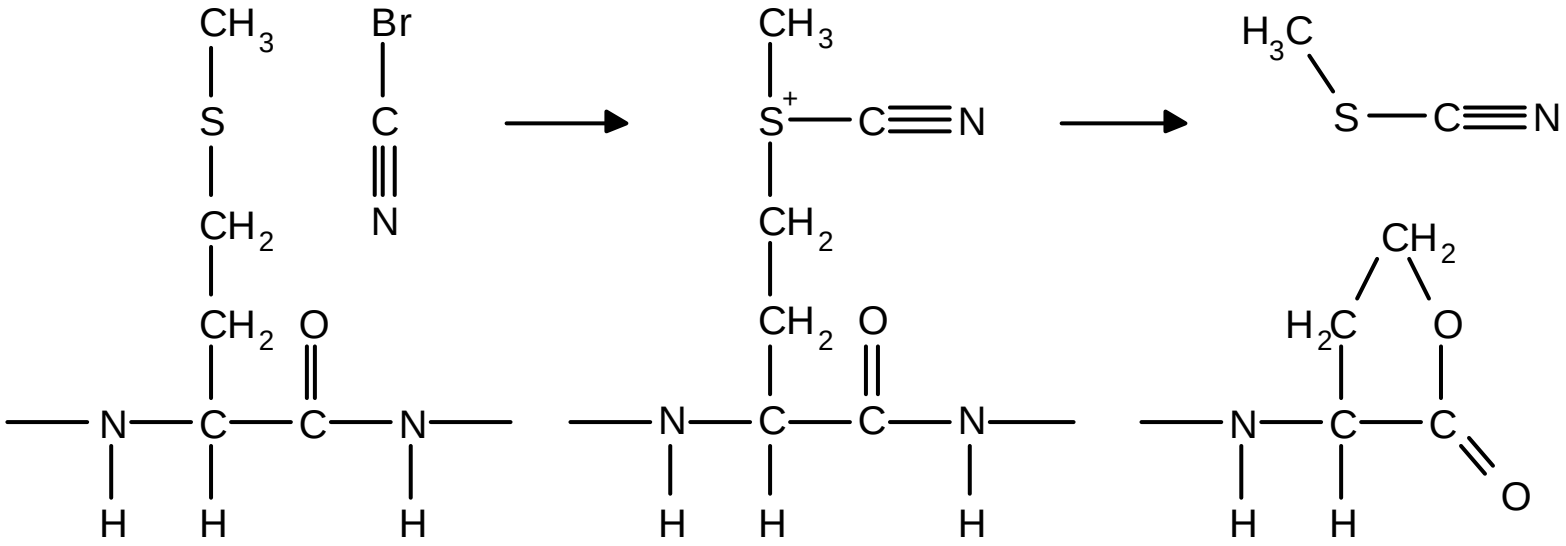
- Why can't the entire protein amino acid sequence be determined from a single set of peptide fragments obtained by cleavage with a single reagent?
- There’s no way to determine how the fragments are connected with just one set
- A second or third set of fragments are used to deduce how the fragments are connected by identification and comparison of overlapping sequnces

Protein Cleavage Reagents

- Chemical or enzymatic reagents can be used to prepare protein fragments
- The most commonly used reagents are:
  - cyanogen bromide (Met→homoserine lactone)
  - various enzymes including
    - Trypsin (R or K)
    - Chymotrypsin (F or Y or W)
    - Clostripain (R>K) (can be incomplete for K)
    - Staphylococcal protease (D or E)
    - various endopeptidases

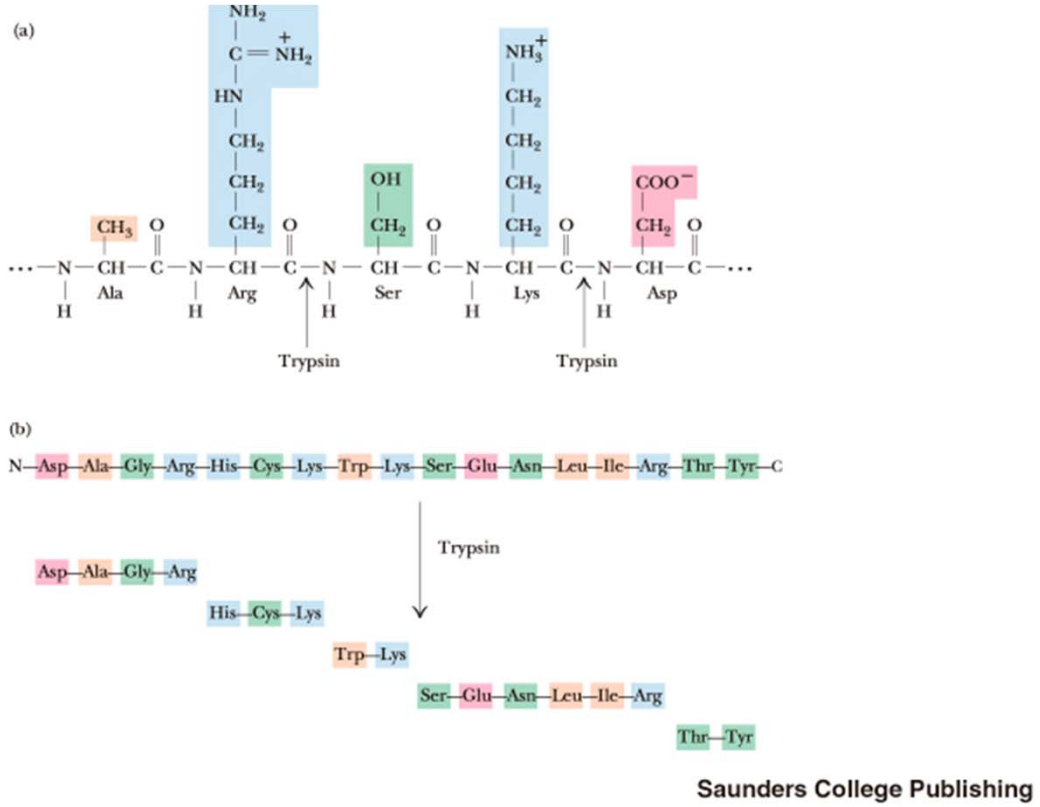
Cyanogen Bromide

- It cleaves protein chains at internal methionines by reaction with the methionine sulfur as illustrated below



Enzymatic fragmentation

- Trypsin (R or K)
- Chymotrypsin (F or Y or W)
- Clostripain (R>K) (can be incomplete for K)
- Staphylococcal protease (D or E)



Endopeptidases

- The following endopeptidases are less specific than the enzymes metioned above
- Pepsin, papain, subtilisin, thermolysin, elastase
  - (papain is the active ingredient in meat tenderizer, soft contact cleansing solutions, some laundry detergents)
- These enzymes are most often used to further reduce the size of large tryptic or chymotryptic fragments

SDS-PAGE

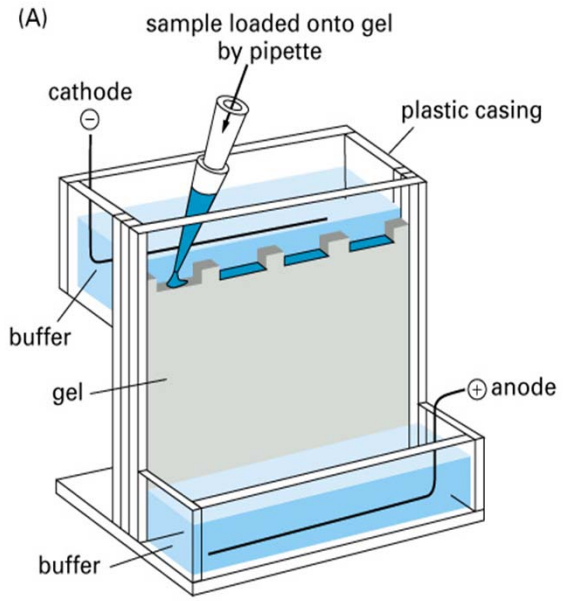


Figure 8–14 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

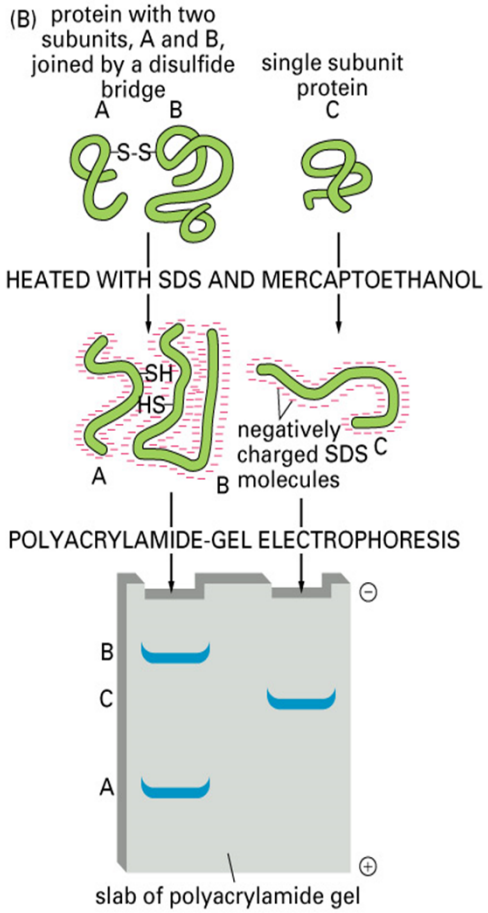
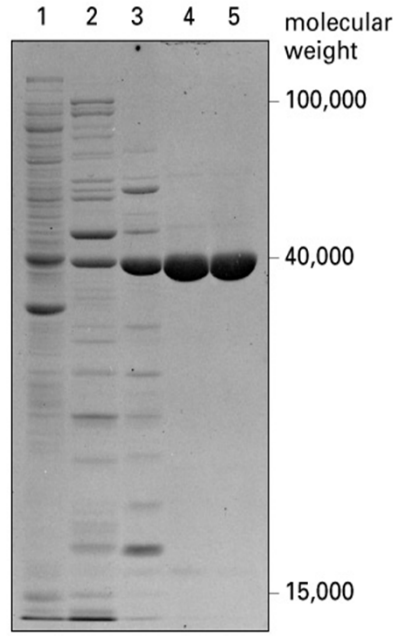


Figure 8–14 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

Coomassie Blue/
Silver Staining



How are Peptide Fragments Separated?

- Usually by column chromatography, often HPLC
- Separations are most often based on differences in polarity (reverse phase) or electric charge (ion exchange)

Edman Degradation

- Edman degradation is most often used to sequence the peptides
- It removes one amino acid from the N-terminal end of the peptide during each cycle of the procedure
- The removal of the N-terminal amino acid is accomplished using the reagent, phenylisothiocyanate

Mass Spectroscopy

- Used for sequencing peptides
- Peptides are fragmented in the mass spectrometer
- The fragments are identified by their mass/charge ratio

Reconstructing the Sequence

- Use two or more fragmentation agents in separate fragmentation experiments
- Sequence all the peptides produced (usually by Edman degradation)
- Compare and align overlapping peptide sequences to learn the sequence of the original polypeptide chain

Compare cleavage by trypsin and staphylococcal protease on a typical peptide:

- Trypsin cleavage: (R or K)

  A-E-F-S-G-I-T-P-K L-V-G-K

- Staphylococcal protease: (D or E)

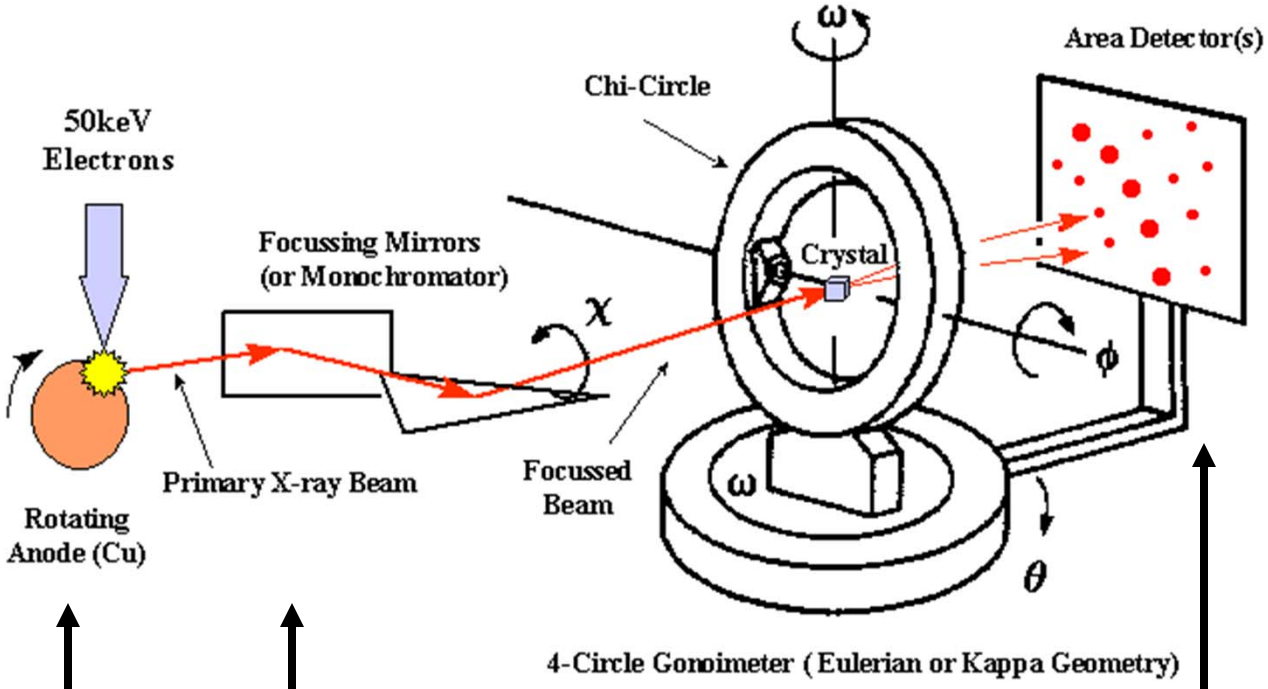
  F-S-G-I-T-P-K L-V-G-K-A-E

- The peptide will be:

  L-V-G-K-A-E- F-S-G-I-T-P-K

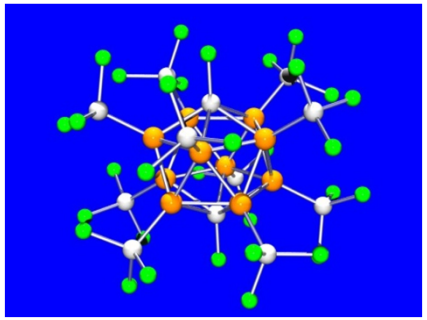
X-ray Crystallography

X-ray crystallography is an experimental technique that exploits the fact that **X-rays are diffracted by crystals**. Based on the diffraction pattern obtained from the periodic assembly of molecules or atoms in the crystal, the electron density can be reconstructed and thus the arrangement of the atoms can be determined.



Data Processing

Structure Solution and Refinement



X-ray source

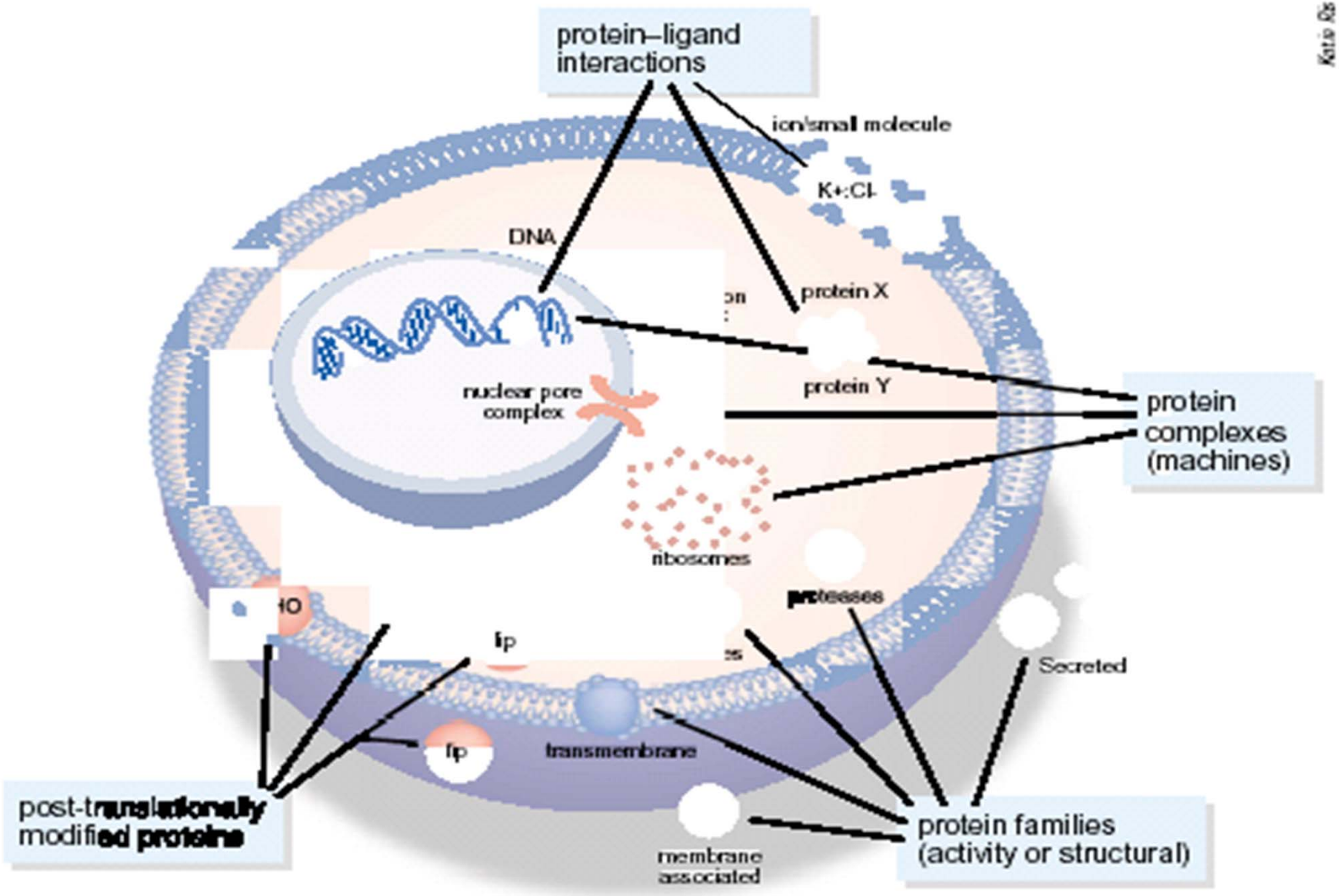
X-ray preparation

Diffraction

Detection

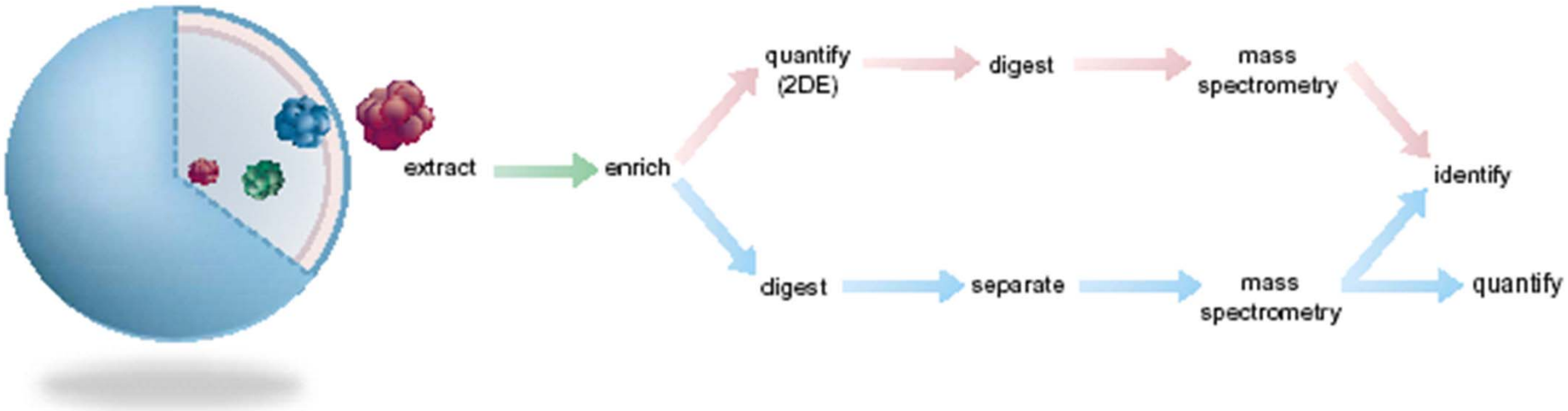
Proteomics

- A new scientific discipline dedicated to understanding the complex relationship of disease and protein expression
  - Uses protein microarrays to test variation in protein expression between healthy and disease states



uences did not always have to be mass spectrometric measurements can distinguish closely related

Process for protein isolation



Protein levels - Determining the abundance of proteins in a sample.

2D gel electrophoresis, mass spectrometry, protein microarrays

Western Blot

- Determine the presence and level of a protein in a cell lysate.

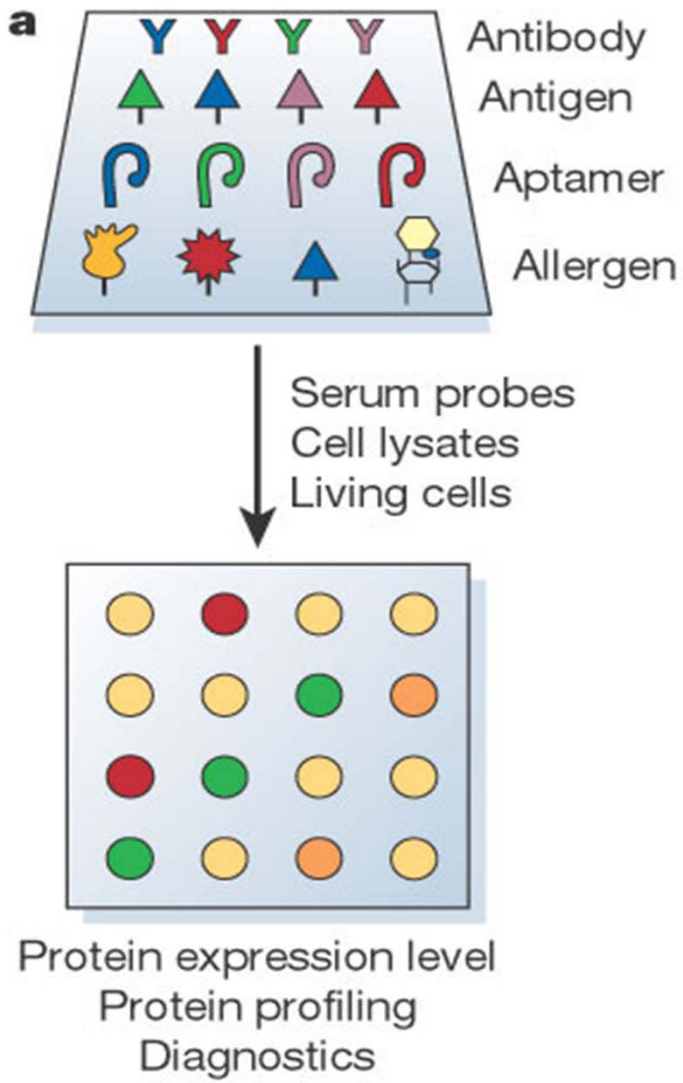
Monitoring protein levels - large scale

- 2D gel electrophoresis
  - Old technology - not as useful for lowly expressed proteins.
- Mass spectrometry
  - Many new techniques for protein detection and quantitation being developed.
- Protein microarrays
- Many developing technologies

Protein microarrays

- Similar to DNA microarrays
  - Plate, Probe, Attachment
- Advantage
  - Poor correlation between mRNA and protein expression
- Analysis of thousands of proteins at one time.
- Many different types
  - Antibody arrayed - detect many proteins
  - Proteins arrayed - detect interacting proteins
  - Proteins arrayed - detect interacting small molecules
  - Etc.

Analytical Microarrays

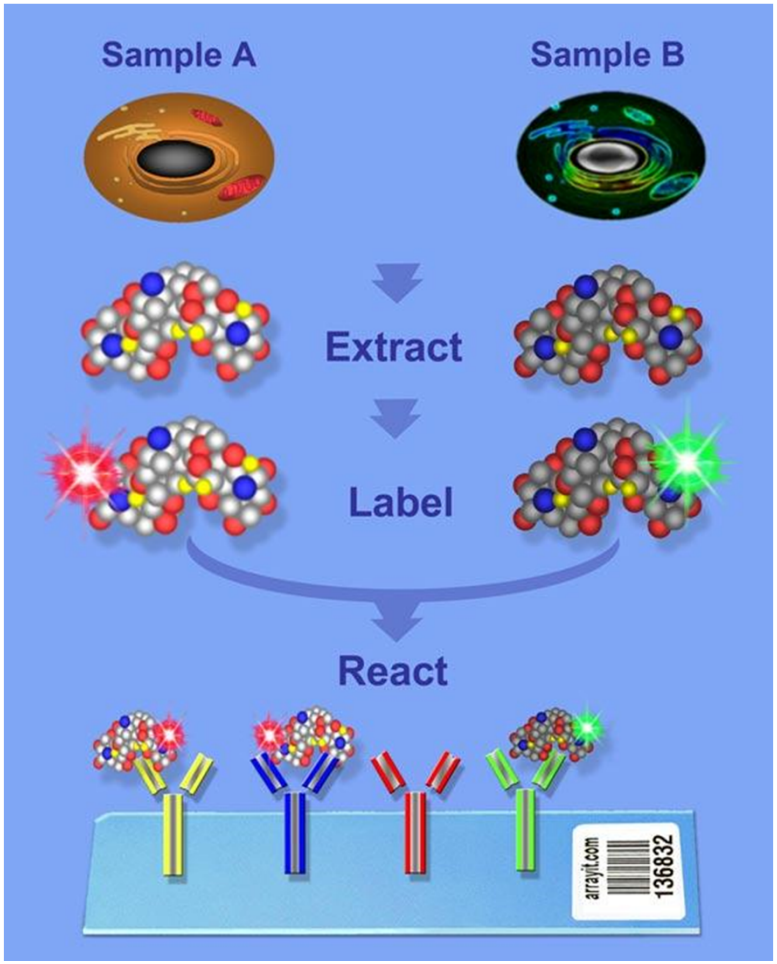


- Profiles Mixture of Proteins
  - Measure Binding Affinity
  - Specificity
  - Protein Expression Levels
- Most Common
- 3 main probe ligand types
  - Antibodies
  - Antigens
  - Carbohydrates of small molecules
  - DNA or RNA Aptamers

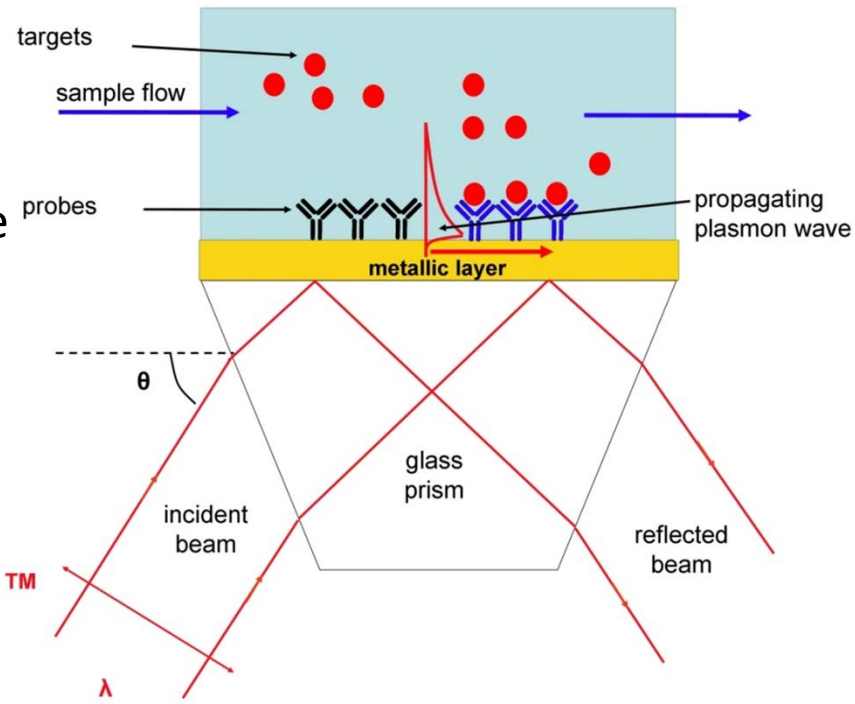
Similar to the procedure in DNA microarray experiments, protein samples from two biological states to be compared are separately labelled with red or green fluorescent dyes, mixed, and incubated with the chips. Spots in red or green colour identify an excess of proteins from one state over the other

Sample Preparation

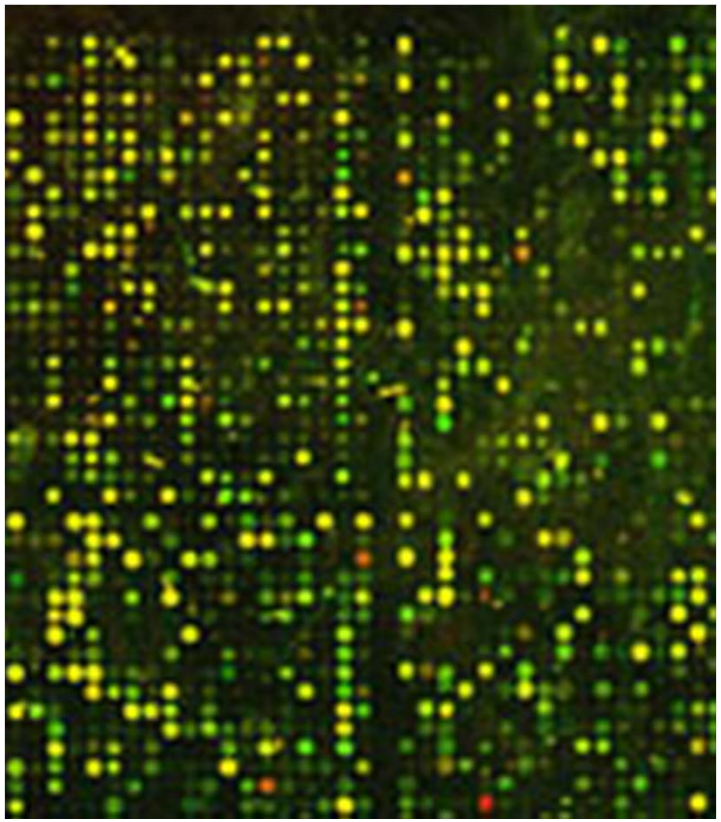
- Sample extracted from cells or tissues
- Bio-Rad assay
- Labeled
  - Fluorescent Dye
    - Cy3/Cy5 via Lysines
  - Photochemical
  - Radioisotope
  - May interfere



- Unlabeled
  - Antibody Sandwich
    - 2nd antibody with label incubated on top of sample
  - Surface Plasmon resonance
    - Measure electromagnetic waves
    - Angle changes in the order of 0.1° with 1 nm film adsorption
    - Needs special equipment
  - Don't affect protein structure



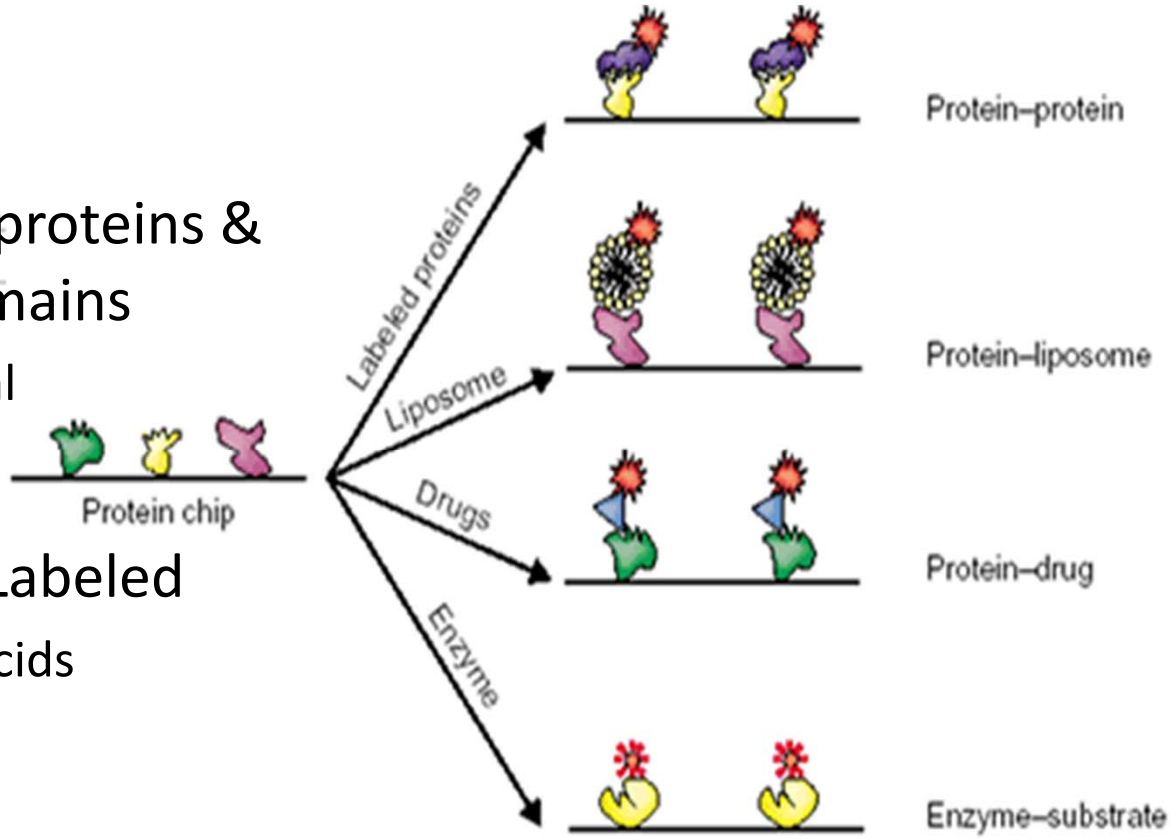
Detection & Quantification



- Scanner
  - Detects dye
  - Adjusts for background
- Reference spots
  - Labeled known concentrations
- Computational Analysis

Functional Microarrays

- Plates
  - Full length proteins & protein domains
    - Functional
- Samples
  - Purified & Labeled
    - Nucleic Acids
    - Proteins
    - Lipids
    - Small Molecules



Reverse Phase Microarrays

- Plates
  - Cell Lysate
- Sample
  - Antibodies of interest
    - Primary
      - Attach to spots
  - Secondary
    - Attach to primary
    - Labeled
- Detect Altered Proteins
  - Post-translation modification problems
  - Disease