

Bio312 Protein Structure and Function

Class Today

- Self introduction
- Handbook Highlights
- Lecture 1

About Dr. Han

- **Research Keywords:** Virulence regulators; Protein expression and purification; Protein crystallography; Enzyme Kinetics; Drug design and screen
 - Email address: Lanlan.Han@xjtu.edu.cn
 - Office telephone number: (0512)89165331
 - Room number and office hours: SA344, by appointment

Handbook Highlights -- Contents

Methods and Applications of Proteome Analysis. (Lecture 1-6);

- Lecture 1: Mischaracterization of proteins
- Lecture 2: Identification of proteins in complex mixtures
- Lecture 3: Bottom-up and top-down proteomics
- Lecture 4: Quantitative mass spectrometry-based proteomics
- Lecture 5: Protein post-translational modifications
- Lecture 6: Targeted proteomics

Acquisition of proteins and protein structure determination (lecture 7-14);

- Lecture 7: Protein folding and unfolding
- Lecture 8: Overview of techniques and applications
- Lecture 9: Circular dichroism and infrared spectroscopy
- Lecture 10: X-ray crystallography
- Lecture 11: Nuclear magnetic resonance: Structural Determination
- Lecture 12: Overview of Techniques for protein-ligand interactions
- Lecture 13: Modelling: Computational Prediction of Structure & Dynamics
- Lecture 14: Relating structure to function

Lecture Room: SA136

Lectures: Mondays, 14:00-15:50

Tutorials: Thursdays, 14:00-14:50

Handbook Highlights -- Learning Outcomes

1. Know the methods used for the extraction, enrichment and analysis of proteins
2. Discuss how proteomics-based approaches can be used to study fundamental and applied biological problems
3. Describe methods of analysis of post-translational modifications of proteins and implications for cell function
4. Understand how technologies such as nuclear magnetic resonance (NMR), Xray crystallography and other physical methods can be used to determine the detailed fine structure of proteins
5. Explain how the knowledge of protein structure can be used to explain function, in particular the structural basis of receptor signaling and the function of molecular motors.

Handbook Highlights-Assessment

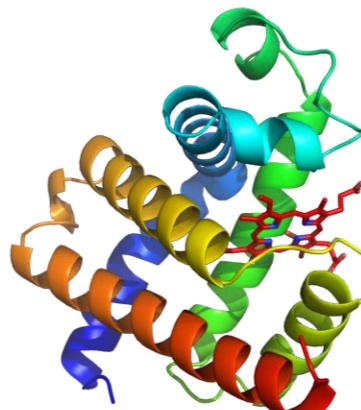
Assessment Task	Learning Outcomes	Weighting	Release Date	Due Date
Coursework Assignment 1	A,B	15%	14/Mar/2025	24/Mar/2025
Coursework Assignment 2	C,D	15%	11/Apr/2025	21/Apr/2025
Coursework Assignment 3	E	10%	08/May/2025	15/May/2025
Final Exam	A-E	60%		Closed-book

What are proteins?

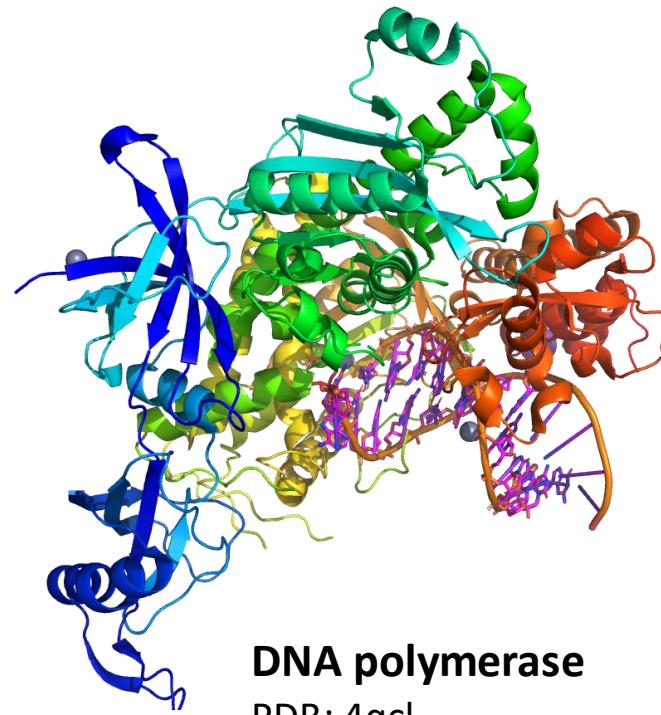
- **Proteins are polymers of amino acids**
- Proteins play a crucial role in biological processes and have many important biological functions
 - **Enzymes** – biological catalysts
 - **Defense proteins**
 - **Antibodies** produced in response to **antigens**
 - **Transport proteins**
 - **Regulatory proteins**
 - **Structural proteins**
 - **Movement proteins**
 - **Nutrient Proteins**

How do Proteins Act?

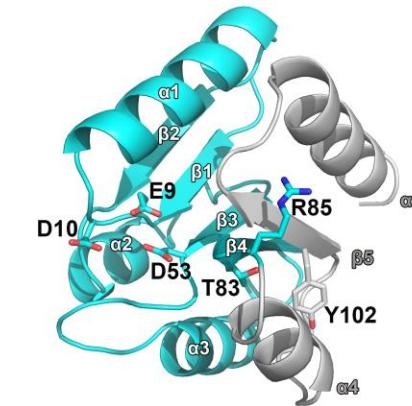
- Proteins have many important biological functions
 - Binding
 - Catalysis
 - Molecular switches
 - Structural component



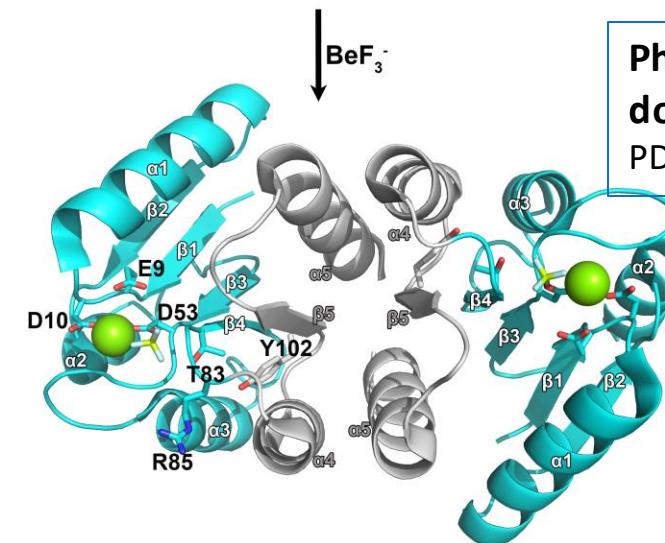
Myoglobin
PDB: 1a6k



DNA polymerase
PDB: 4qcl



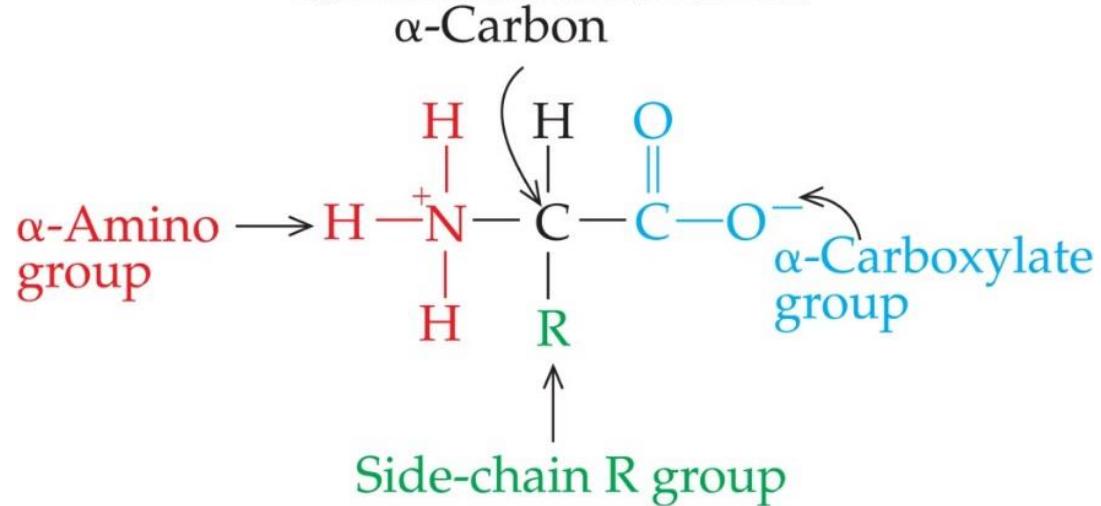
PhoB REC domain monomer
PDB ID: 1B00



PhoB REC domain Dimer
PDB ID: 1ZES

Protein Building Blocks - Amino Acids

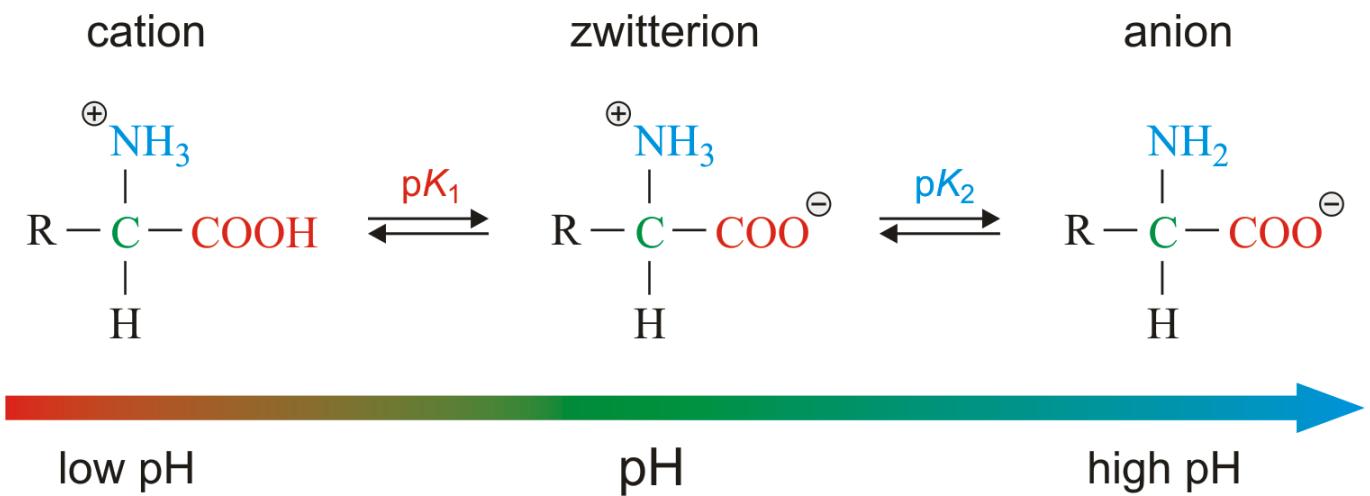
Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



- **α-Amino acids** contain both an amine and an acid
- 20 common amino acids in nature
 - Differ in R group
- At physiologic pH the amino acid has:
 - Carboxyl group in $-COO^-$
 - Amino group in $-NH_3^+$
- Neutral molecule with equal number of + and - charges is a **zwitterion**

Zwitterions

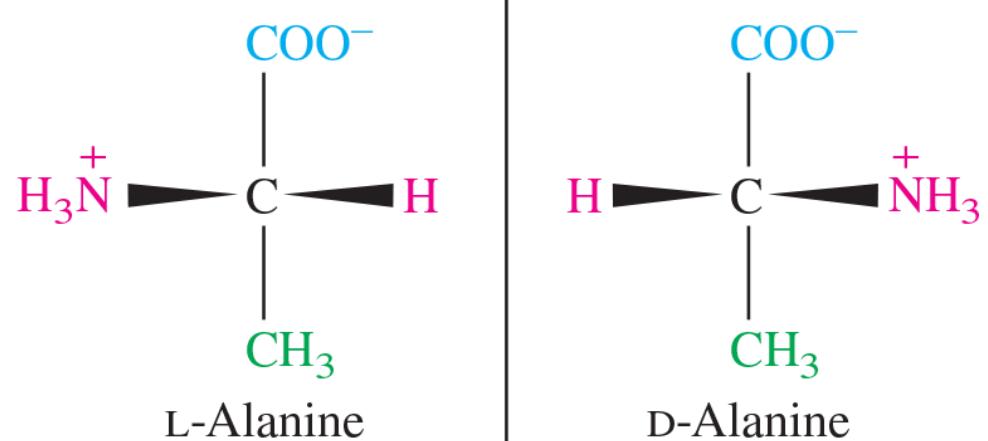
- By changing the pH, you can affect the net charge on the zwitterions
- The pH point at which there is no net charge on the zwitterions is called the *isoelectric point (pI)*



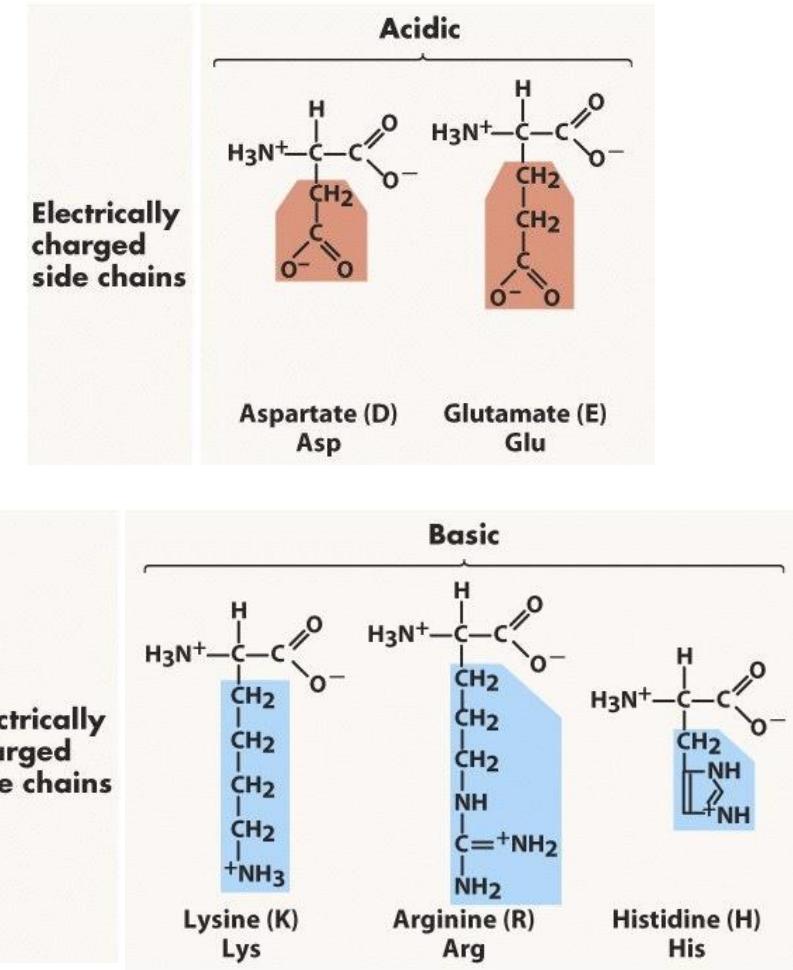
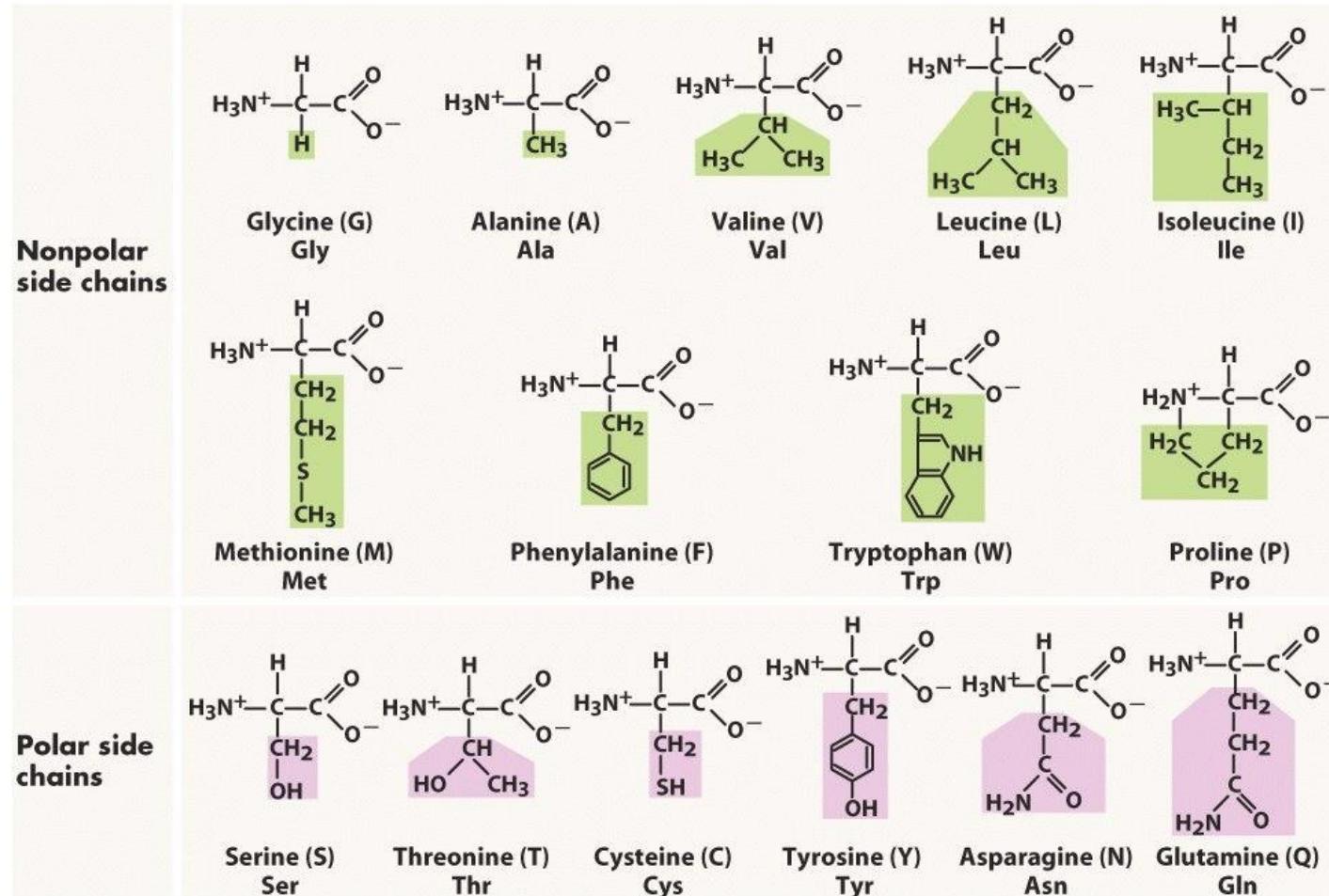
- $\text{pH} < \text{pI}$, positively charged
- $\text{pH} > \text{pI}$, negatively charged

Stereoisomers of Amino Acids

- The α -carbon of amino acids is chiral
 - 19 out of the 20 are stereoisomers
 - *Glycine* is the only common amino acid that is not chiral
- D vs L is decided by the amino group
 - L for left in Fischer projection.
 - D for right.
- Amino acid configuration isolated from proteins is L-



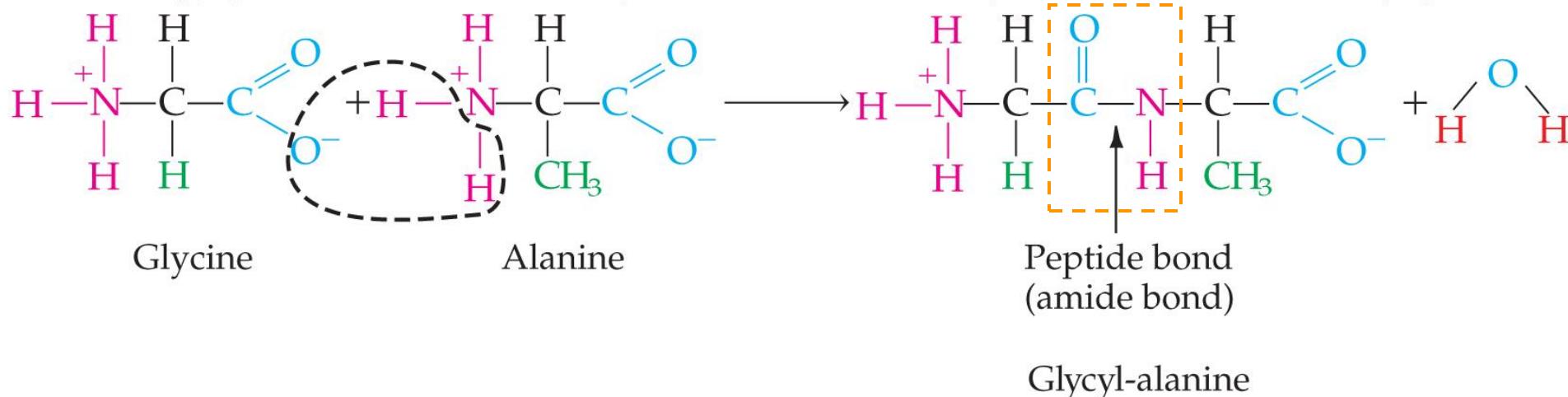
Structures and Abbreviations of the 20 Common Amino Acids



The Peptide Bond

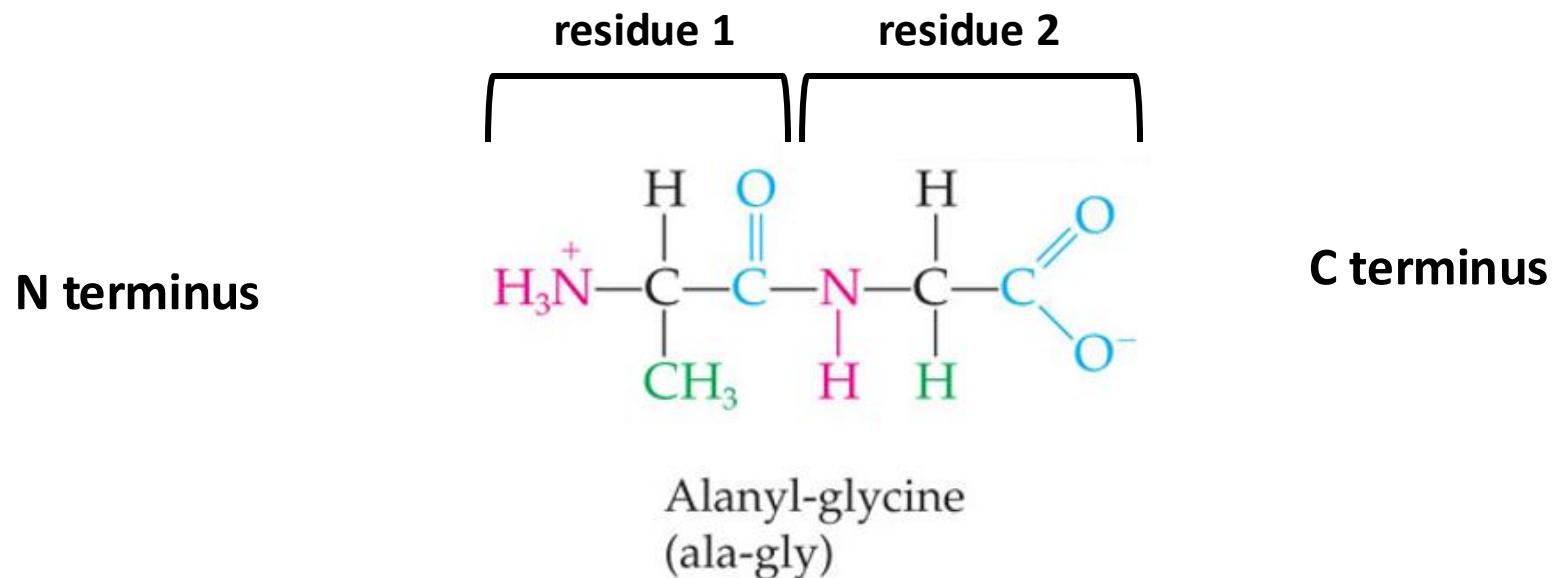
- Proteins are polymers of L- α -amino acids
 - Carboxyl group of one amino acid is linked to the amino group of another amino acid
 - Linkage is an amide bond or **peptide bond**
 - This reaction is a dehydration reaction as water is released

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.

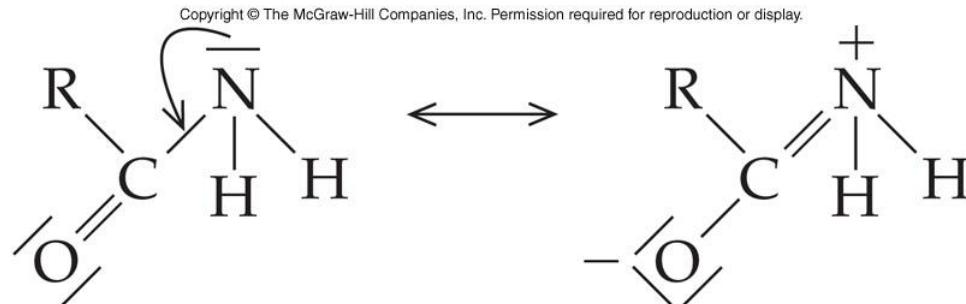


Dipeptides

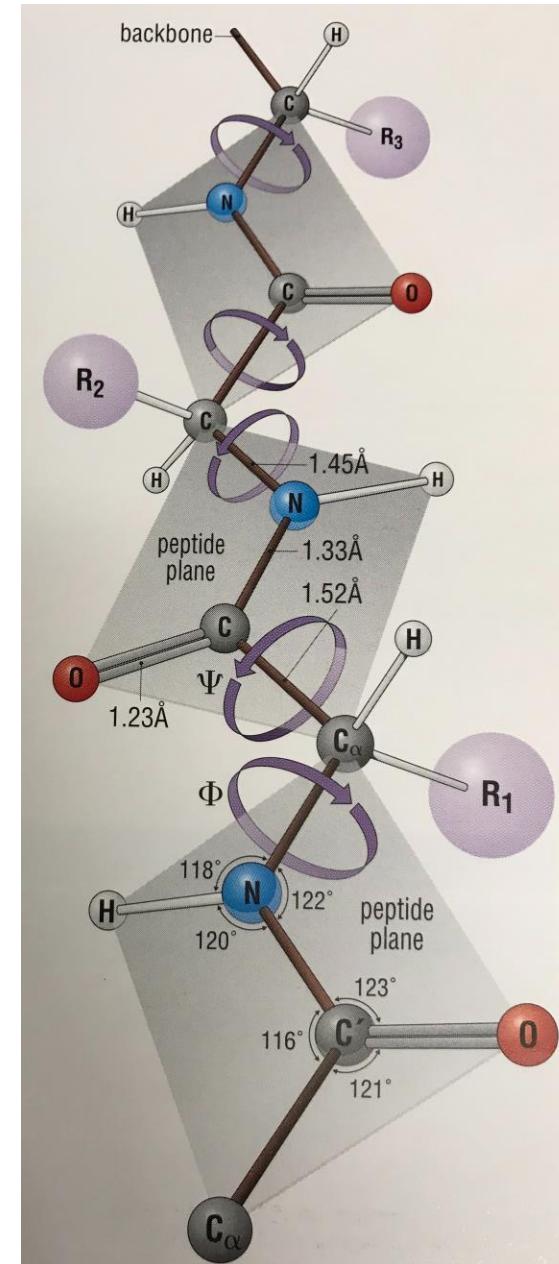
- Condensing or dehydrating two amino acids produces a dipeptide
 - Amino acid structures are written with the N-terminal on the left



Structure of Peptide Bond



- Resonance
 - Increase the polarity of the peptide bond
 - Partial double bond property: the carbonyl O, Carbonyl C, and the amide N) are **coplanar**
- N- C_{α} and C_{α} -C bonds are free rotatable



From Sequence to Structure: Primary Structure

- Central Dogma:
 - Genetic information flows: DNA → RNA → Protein
- **Primary structure** is the amino acid sequence of the polypeptide chain connected by the peptide bonds
 - Is determined by the sequence of a gene (DNA)
 - Between proteins, the more similar the sequence the more similar:
 - Their function
 - Between species; within species
 - Their evolutionary history

From Sequence to Structure: Primary Structure

- Sequence alignment is a way of arranging primary sequences (of DNA, RNA, or proteins) in such a way as to align areas sharing common properties.

CLUSTAL O (1.2.1) multiple sequence alignment (Partial)

RitR	MGKRILLLEKERNLAHFLSLELQKEQYRVDLVEEGQKALSMALQTDYDLILLNVNLGDMM
PhoB	MARRILVV E DEAPIREMVCFVLEQNGFQPVEAEDYDSAVNQLNEPWPDLLLD D WMLPGGS
	* .:***: : *.* : * : : : . : . * : .. * : . : * * * * : *
RitR	AQDFAEKLSRT---KPASVIMILDHWEDLQEELEVQRFAVSYIYKPVLINLVARISAI
PhoB	GIQFIKHLKRESMTRDIPVVMLT-ARGEEEDRVRGLETGADDYITKPFSPKELVARIKAV
	. : * : : * . * : * : * : : : . : : . * . ** * . : : * * * * . * :

The degree of relatedness, similarity between the sequences is predicted computationally or statistically

Many software tool used for general sequences alignment tasks, e.g., [Clustal Omega](#) or [BLAST](#)

Clustal Omega

[https://www.ebi.ac.uk/
Tools/msa/clustalo/](https://www.ebi.ac.uk/Tools/msa/clustalo/)

Clustal Omega

Input form

Web services

Help & Documentation

Bioinformatics Tools FAQ

Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of

PROTEIN

sequences in any supported format:

```
>sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
MVLSPADKTNVKAAGWKVGAHAGEYGAELERMFLSFPTTKTYFPHFDLHSQAVKGHG
KKVADALTNAVAHVDDMPNALSDLHAHKLRVDPVNFKLLSHCLLVTLAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR
>sp|P01942|HBA_MOUSE Hemoglobin subunit alpha OS=Mus musculus GN=Hba PE=1 SV=2
MVLSGEDKSNIKAAWGKIGGHGAEYGAELERMFASFPTTKTYFPHFDVSHGSAQVKGHG
KKVADALASAAGHLDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLASHHPADFTP
AVHASLDKFLASVSTVLTSKYR
```

Or, upload a file: [Choose File](#) no file selected

[Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

STEP 2 - Set your parameters

OUTPUT FORMAT

ClustalW with character counts

The default settings will fulfill the needs of most users.

[More options...](#) (Click here, if you want to view or change the default settings.)

STEP 3 - Submit your job

Be notified by email (Tick this box if you want to be notified by email when the results are available)

Submit

Basic Local Alignment Search Tool (BLAST)

- The BLAST finds regions of local similarity between sequences.
- The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.
- BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families in the databases.
- Website: <https://blast.ncbi.nlm.nih.gov>
- Introduction to BLAST homepage:
https://ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo_BLASTGuide.pdf

Basic Local Alignment Search Tool (BLAST)

BLAST® » blastp suite

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

```
WVVIDTGTKLWPTLDLKAGLLVFSEIDLGPVEKIYSDILLGVSPPLITALIREFSRDA  
ADGLGLADLHAFILHNRSVVRRALAGVEGVSPDPESRSSVERAFAVRTGTEV  
WEELQRHHVFALPCRQFHWAEPSPGDGHMVRIALSRSTEPELKSVQLRTVLE  
TR
```

Query subrange [?](#)

From
To

Or, upload file [Choose File](#) no file selected [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database Non-redundant protein sequences (nr) [?](#)

Organism exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
- blastp (protein-protein BLAST)**
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST Search database nr using Blastp (protein-protein BLAST) Show results in a new window

BLAST® programs search protein subjects using a protein query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [?](#)

```
MTTQPQLKENLTQWEYLALNSELNIADGHARQALSPGQQKIVNELPVLWAES  
QRPVQQIESEAHQAYFTLLGQHGPVPAEGRVLSCYSSSVSMEILARSLSASVD  
RVALVHPTFDNIADLLRGNGLDLVPVEEDALHGADLSAELLSSVGCVFVTTPNN  
PTGRVLAEERLRLRAEQCAEHGTVLALDTSFRGFDAAHYDHYAVLQEAGCR
```

From
To

Or, upload file [Choose File](#) no file selected [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Subject subrange [?](#)

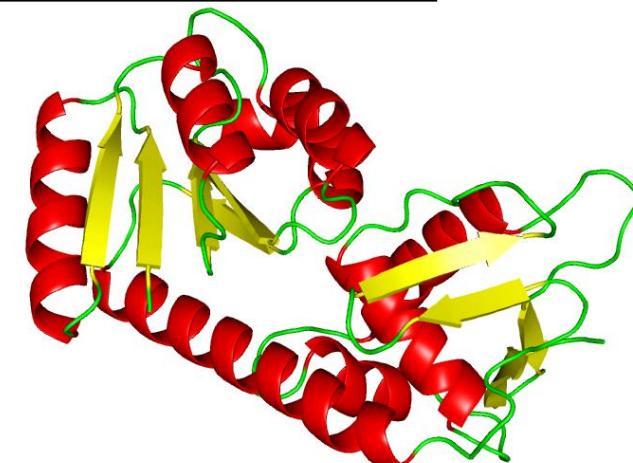
```
no file selected
```

From
To

Or, upload file [Choose File](#) no file selected [?](#)

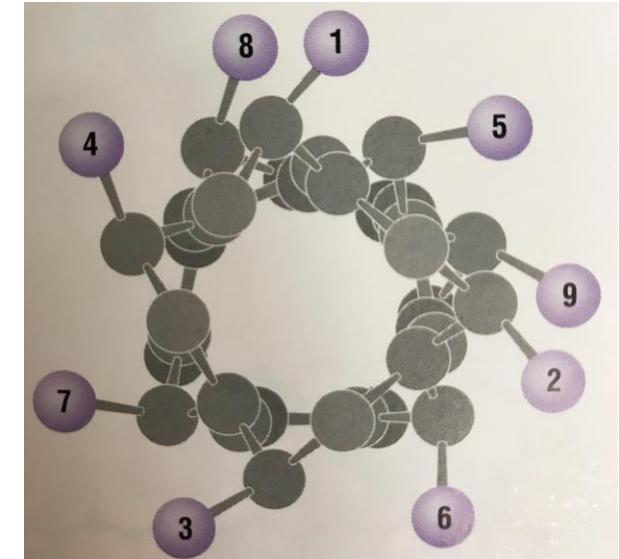
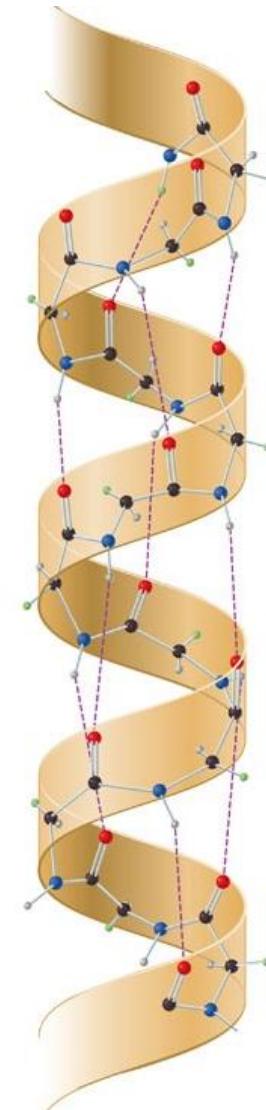
From Sequence to Structure: Secondary Structure

- When the primary sequence of the polypeptide folds into regularly repeating structures, **secondary structure** is formed
- Secondary structure results from hydrogen bonding between the amide hydrogens and carbonyl oxygens of the peptide bonds
- Common secondary structures:
 - α -helix
 - β strand
- Not all regions have a clearly defined secondary structure, some are random or nonregular, e.g., turns, loops.



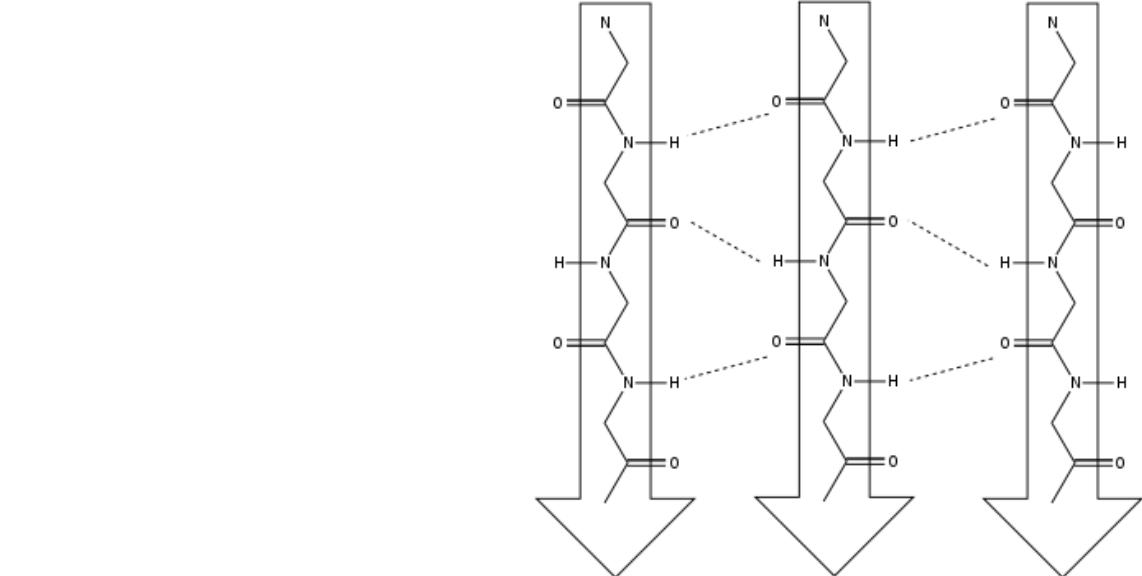
α -Helix

- Amino acids coil “upward”
 - Forms a right-handed helix
 - 3.6 amino acids per turn
 - Repeat distance or pitch is 5.4 angstroms(Å)
- H bonds between $C=O_n$ ----- $H-N_{n+4}$
 - Within the structure, each amide and carbonyl forms a hydrogen bond.
 - All the hydrogen bonds are parallel along the helical axis
- Amino acid side chains stick “out” of the helix.

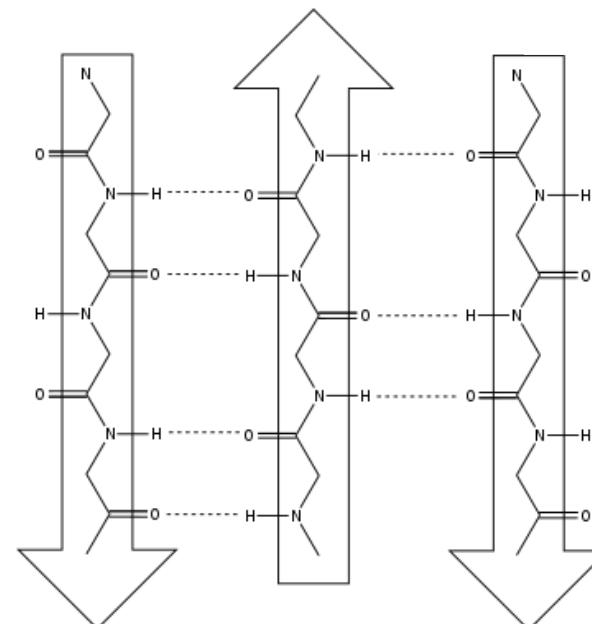


β Sheet

- All of the *carbonyl O and amide H* are involved in the H bonds with the chain nearly completely extended
- Two possible orientations
 - **Parallel** if the N-termini of β strands are head-to-head
 - **Antiparallel** if the N-terminus of one chain is aligned with the C-terminus of the other
 - Common in proteins because it's more stable.
- β sheets are never flat. They always right-handed twist.



Parallel β -sheet



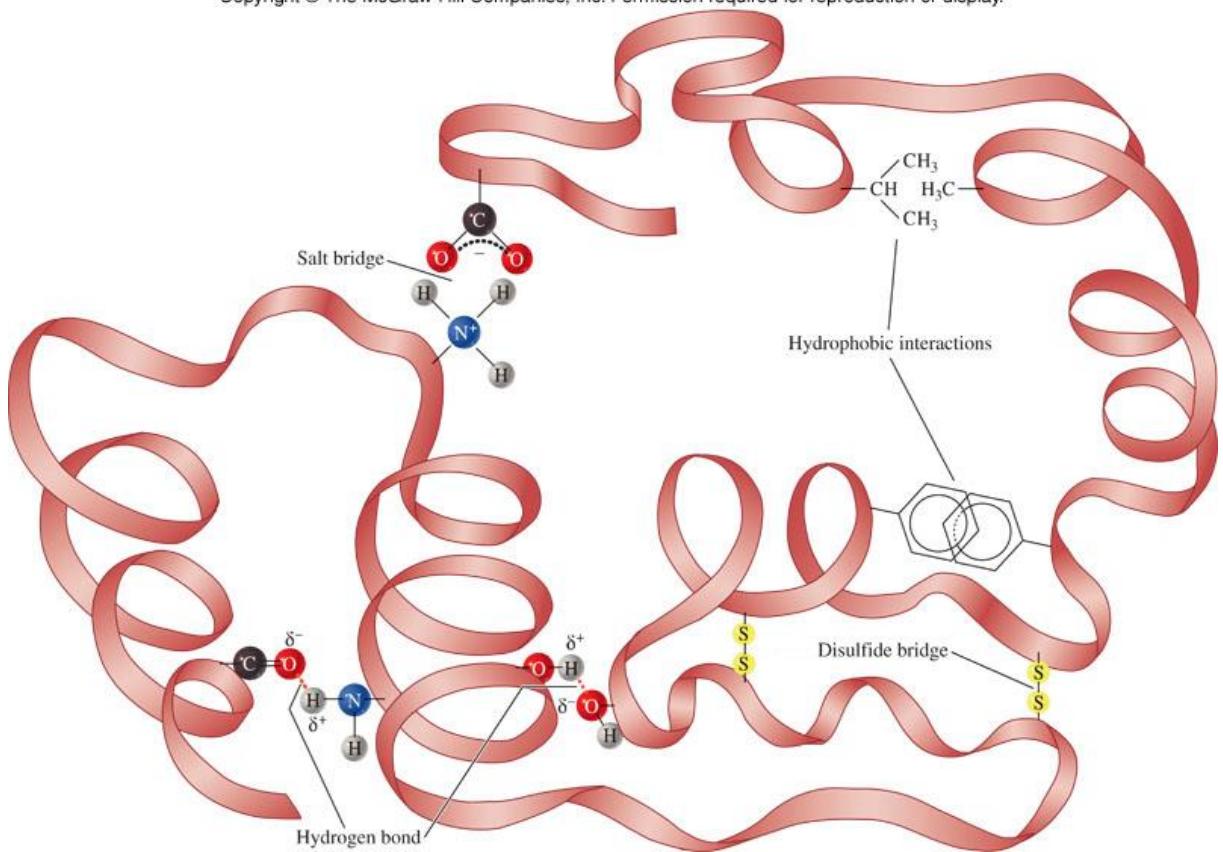
Antiparallel β -sheet

From Sequence to Structure: Tertiary Structure

- In a folded protein, the secondary structure elements fold into a compact form is called the **tertiary structure**
- Occurs due to interactions ***between amino acid side chains.***
 - **Disulfide bridges** between two cysteine residues
 - Non-covalent interactions
 - **Salt bridges** between ionic side chains --COO^- and --NH_3^+
 - **Hydrogen bonds** between H donor and acceptor
 - **Hydrophobic interactions (van der Waals interaction)**: two nonpolar groups are attracted by a mutual repulsion of water

Interactions Involved in Tertiary Structure

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



- Weaker interactions in proteins are dynamic. They form and break constantly, giving protein a globular appearance of an “oil droplet” in a solution.
 - Hydrophobic residues are buried inside
 - Often these movements are vital for ligand binding and catalytic function.

Bound water molecules of a folded protein are an important part of the structure



- May be found on the surface
- Dissociate and exit only to satisfy the H bonding or salt bridges requirements of residues that will ultimately contact ligands
- Bond to end caps of secondary structure
- Form water network to assist in acid/base chemistry

The Quaternary Structure of Proteins

- When more than one polypeptide come together to make a functional protein.
- **Quaternary structure:** the arrangement of subunits or peptides that form a larger protein
 - **Subunit** is a polypeptide chain having primary, secondary, and tertiary structural features
 - Subunits can be identical polypeptides, nearly identical polypeptides or very different polypeptides.
 - One chain: monomer
 - Two chains: dimer (same chains: homodimer; different chains: heterodimer)
 - Trimer, etc.

Types of Interactions in Quaternary Structure

- Quaternary structure is maintained by the same forces which are active in maintaining tertiary structure
 - **Disulfide bridges** between two cysteine residues
 - Non-covalent interactions
 - **Salt bridges** between ionic side chains --COO^- and --NH_3^+
 - **Hydrogen bonds** between H donor and acceptor
 - **Hydrophobic interactions (van der Waals interaction)**: two nonpolar groups are attracted by a mutual repulsion of water

Proteins Structures

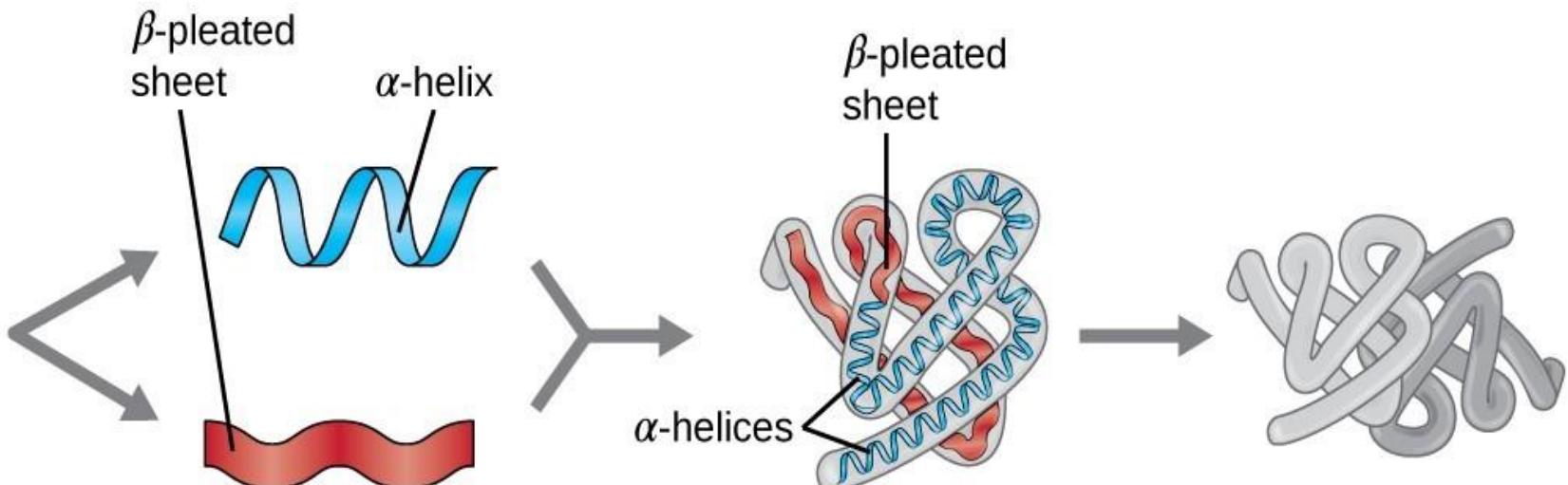


Primary Protein Structure

Sequence of a chain of amino acids

Secondary Protein Structure

Local folding of the polypeptide chain into helices or sheets



Tertiary Protein Structure

three-dimensional folding pattern of a protein due to side chain interactions

Quaternary Protein Structure

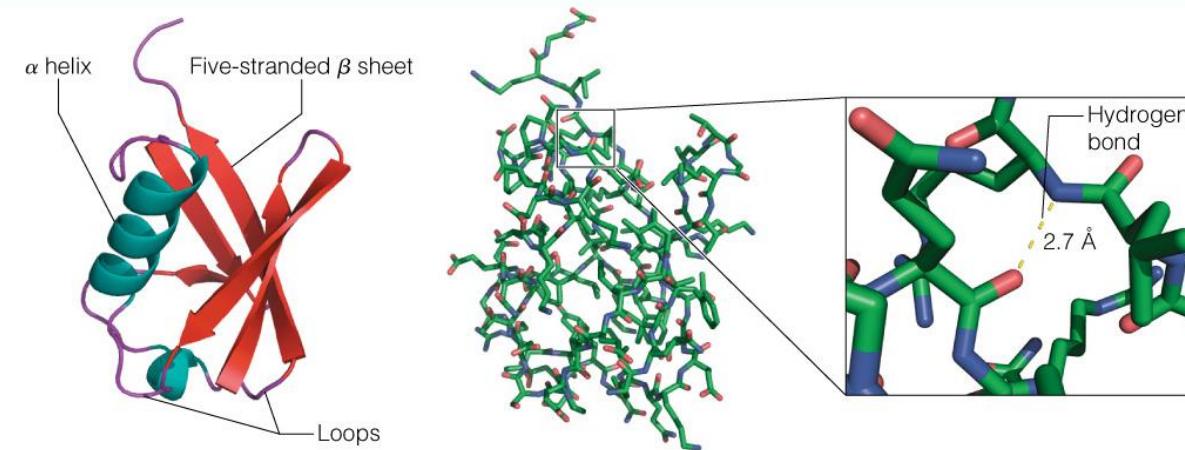
protein consisting of more than one amino acid chain

Protein Modification

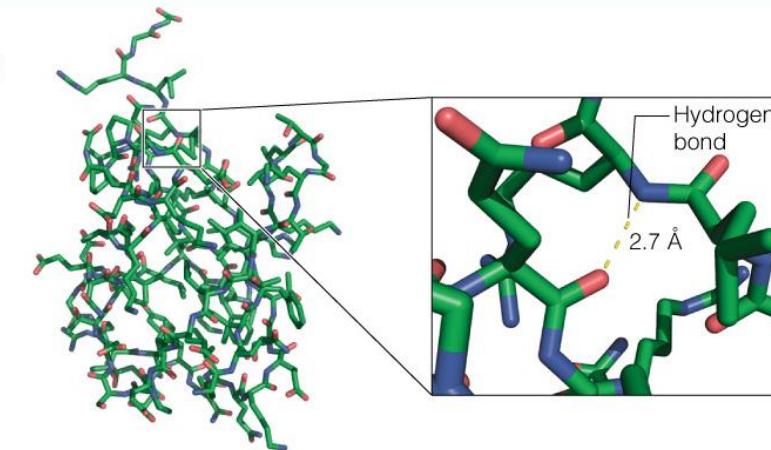
- Most proteins start folding during the translation process!
- Many proteins require post-translational modifications:
 - Some proteins need to be cleaved to be active: zymogens, insulin
 - Some proteins require cofactors and/or form complexes
 - Some require oxidation of cysteines to form disulfide bonds.
 - Some require carbohydrate or lipid modifications

Proteins: What do they “look like?”

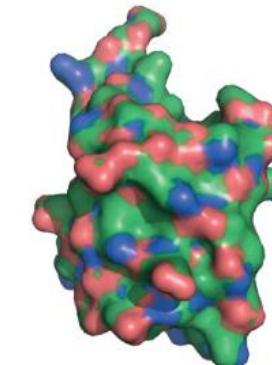
- An individual protein is too small to see
 - X-ray crystallography
 - NMR
 - Electron Microscopy
- Proteins can be represented by different models:
 - Cartoon “ribbon” structure
 - Highlights secondary structure
 - Typically ignores side chains
 - Stick model
 - Highlights side chains
 - Hard to see secondary structure.
 - Surface model
 - Highlights overall “shape” or tertiary structure.
 - Can be color coded by charge, polarity



(a) A cartoon model of the protein backbone. An α helix (cyan), is packed against a five-stranded β sheet (red) composed of parallel and antiparallel strands. Loops are shown in magenta.



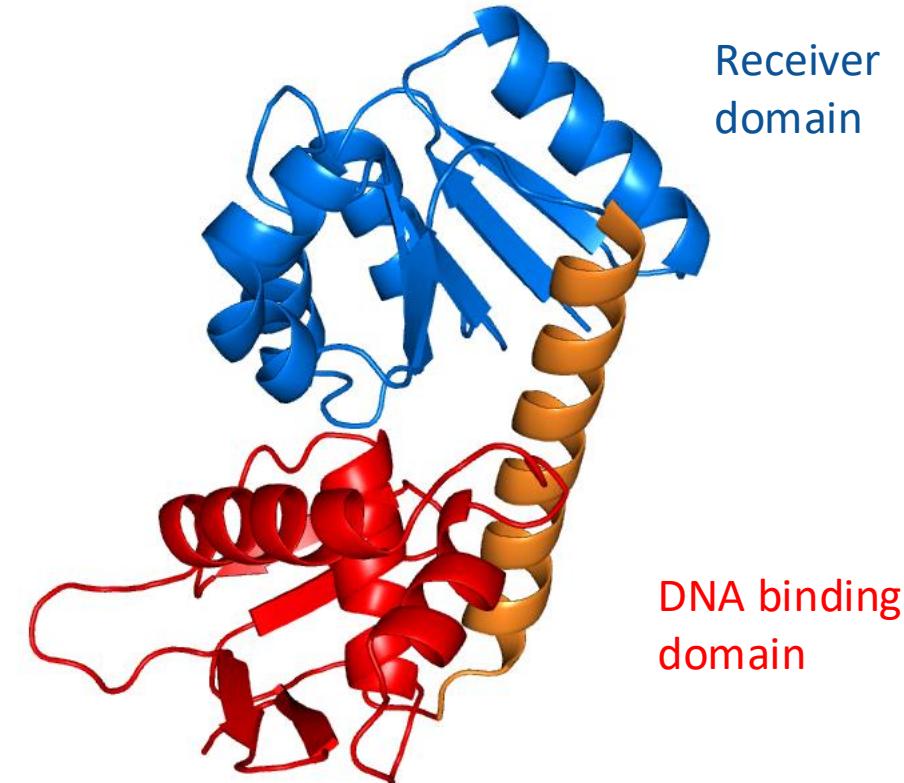
(b) A stick model showing the locations of all atoms (excluding H atoms). C atoms are green, N atoms are blue, and O atoms are red. The inset shows a hydrogen bond of 2.7 Å between main-chain atoms.



Atom coloring is the same as in panel (b).

Domains and Motifs

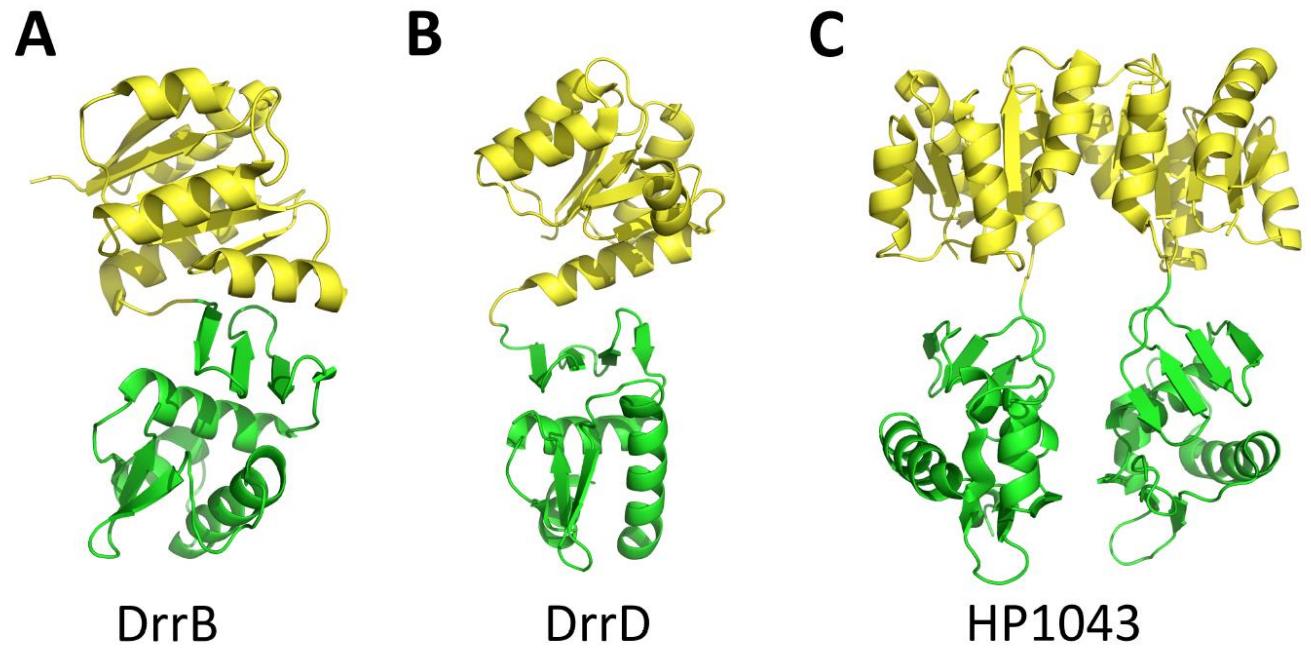
- **Domains:** part of the sequence that appear as conserved modules in proteins that are not related, in global terms.
 - Usually with a distinct three-dimensional fold, carrying a unique function and appearing in different proteins



RitR mutant C128D
PDB: 5VFA

Proteins can be grouped into families based on the domains they contain

- Structural alignment: a method for discovering significant structural motifs.
- But often it is not the case that structural families share a common function.



Structural comparison of various classes of full-length OmpR/PhoB subfamily structures. Rec domain (REC domain) colored in yellow, DNA binding domain colored in green.

Domains and Motifs

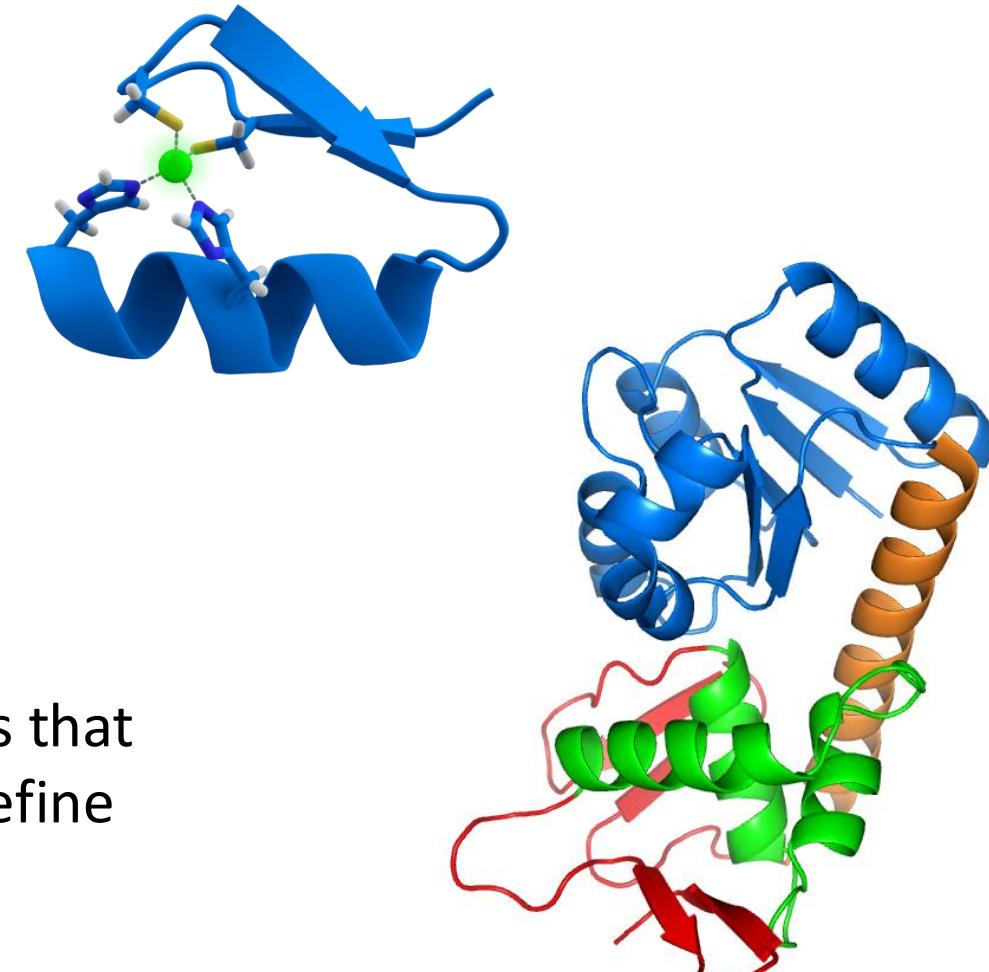
- **Motifs** are commonly used in two ways.

1. Short conserved sequences, which appear in a variety of other molecules.

e.g., **Zinc finger motif** in many DNA binding proteins
CXX(XX)CXXXXXXXXXXXXHXXXH.

2. A set of contiguous secondary structure elements that either have a particular functional significance or define a portion of an independently folded domain.

e.g., **Helix-turn-helix motif** in DNA binding domain



RitR mutant C128D
PDB: 5VFA

Stability of Proteins

- Proteins must fold to a globular conformation to carry out the most important tasks in living organisms
- The unfolding of proteins is called ***denaturation***.
 - Temperature
 - pH
 - Organic Solvents/Detergents
 - Heavy Metals
 - Mechanical Stress
- Peptide bonds between amino acids can be broken through ***hydrolysis***.
 - This changes the primary structure!
 - Strong acid or strong base
 - Certain enzymes

Methods for Stabilization of Proteins

- Correct pH (buffer)
- Maintain temperatures (usually low)
- Minimize processing times
- Minimize agitation
- Minimize denaturing chemicals
- Add reducing agents (Oxidation can cause inactivation typically intracellular proteins)
- Add protease inhibitors

Examples of Stabilizers

A.These reduce free water levels by hydrogen bonding with H₂O

- Glycerol
- Sugar
- Polyethene Glycol

B. BSA (Bovine serum albumin)

- added to proteins which are at LOW concentration

Storage of Proteins

- Similar conditions apply as with Stability
- Freezing (and thaw) is typically OK
- Concentration of contaminants

Lyophilization

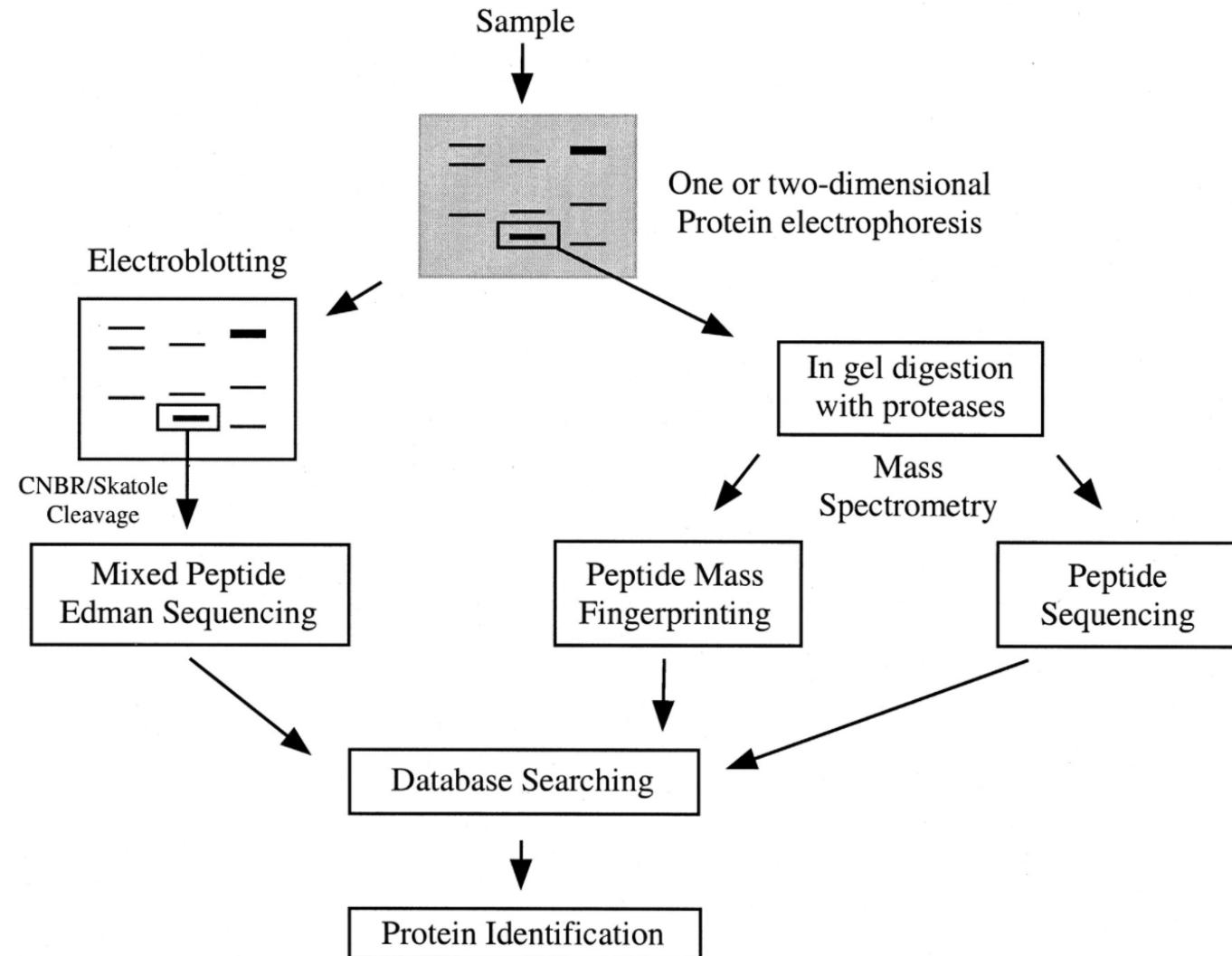


Lyophilization Freeze Dryer System

Drying of protein

- Freeze protein
- Increase temperature
- Apply vacuum
- Remove water vapor

Strategies for Protein Identification: Primary Structure



Protein Identification: Primary Structure

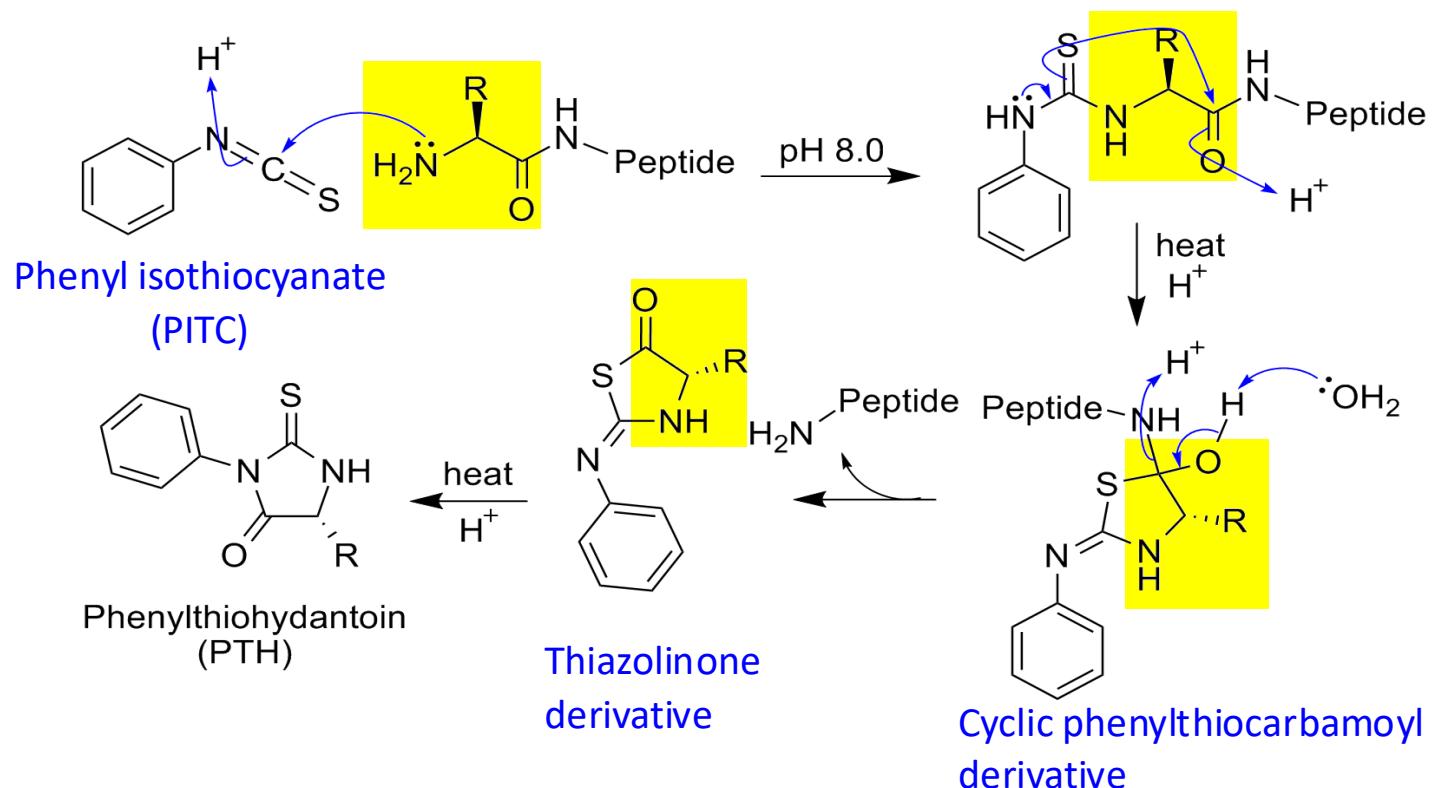
Edman Sequencing

- Obtain N-terminal amino acid sequences
- A variable alternative to MS
- Determine the true start amino acid of a protein

Mass Spectrometry

- Obtain peptide masses or amino acid sequences, and thus identify the protein by searching databases
- Determine the type and location of protein modifications

Edman Sequencing



Works with 30 aa fragments
(in practical)

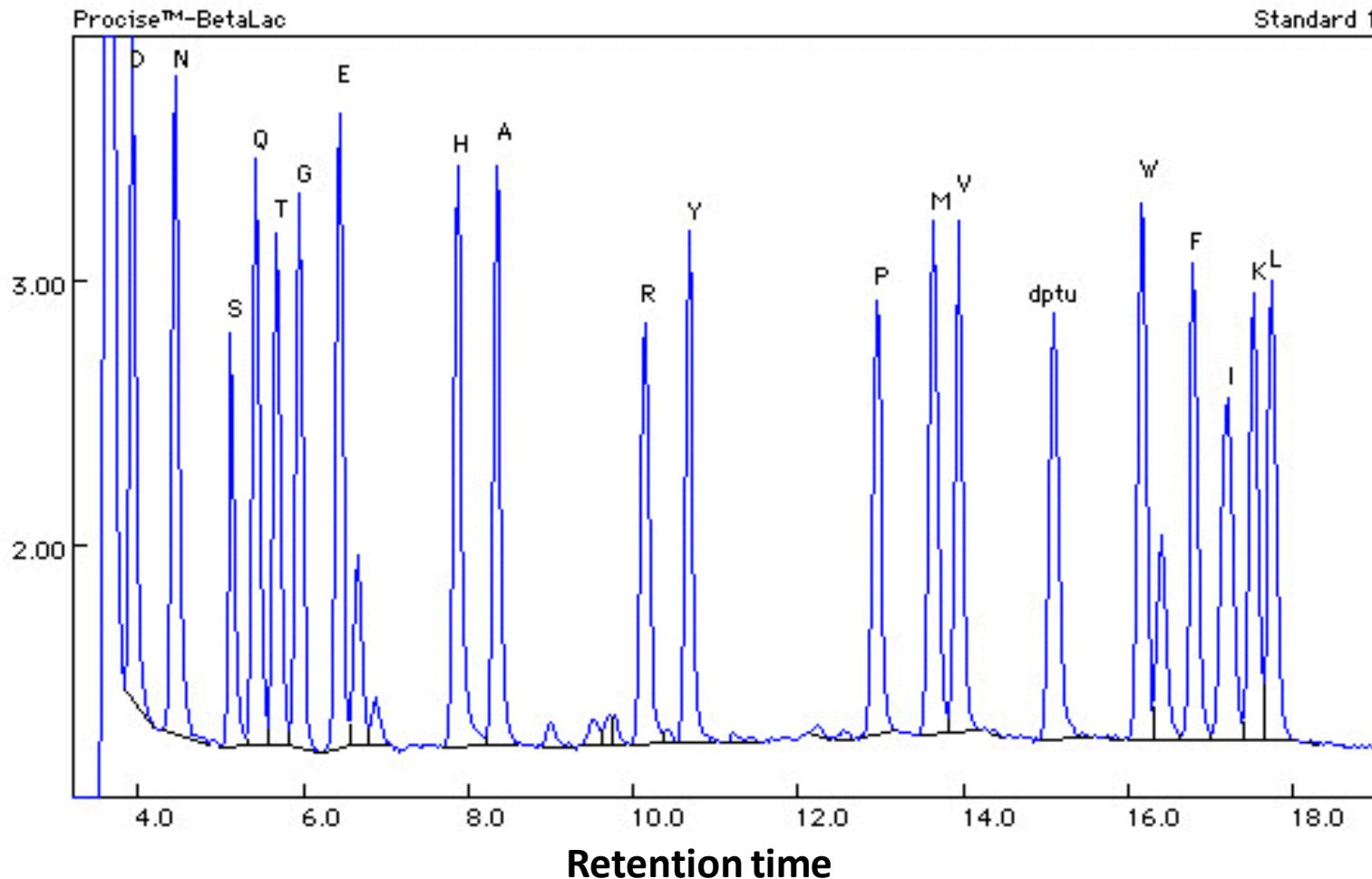
Cut protein with cyanogen bromide (CNBr)
• cuts on carboxyl side of Met

Phenylisothiocyanate (PITC)
• binds to and releases N-terminal residue

Chromatography against known standards

Reagent	Cleavage Properties
<i>Chemical agents</i>	
70% formic acid	Asp-↓-Pro
Cyanogen bromide in 70% formic acid	Met-↓
2-nitro-5-thiocyanobenzoate, pH 9	↓-Cys
Hydroxylamine, pH 9	Asn-↓-Gly
Iodobenzoic acid in 50% acetic acid	Trp-↓
<i>Endoproteases</i>	
Trypsin	Arg/Lys-↓
Lys-C	Lys-↓
Arg-C	Arg-↓
Glu-C (bicarbonate)	Glu-↓
Glu-C (phosphate)	Asp/Glu-↓
Asp-N	↓-Asp
Chymotrypsin	Phe/Tyr/Trp/Ile/Met-↓ (also Ile/Val-↓)

HPLC Chromatogram of Known Standard Amino Acids

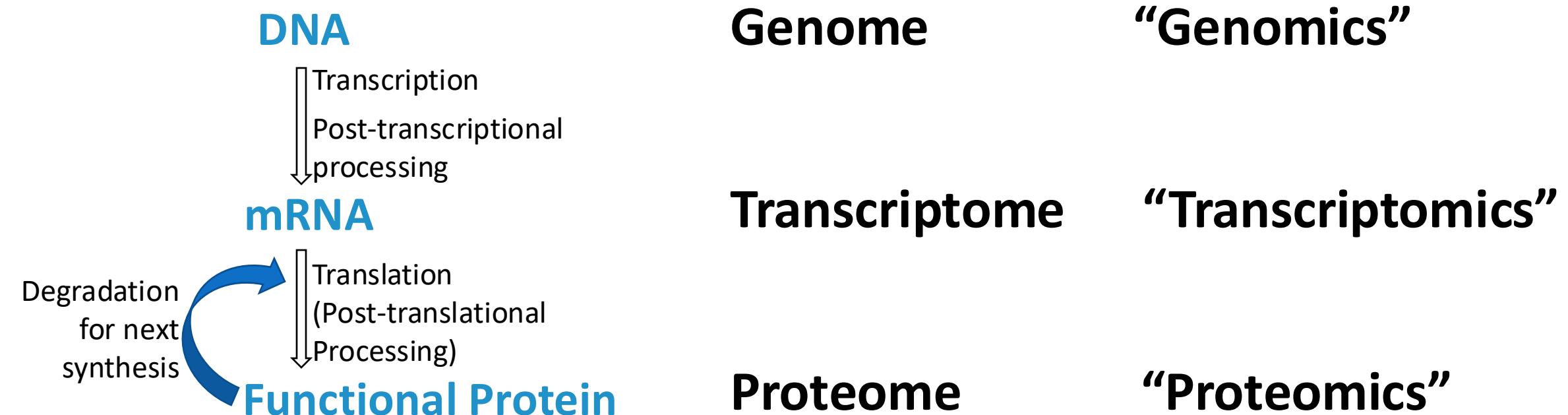


PROTEINS IN COMPLEX MIXTURES

- A MASS SPECTROMETRY APPROACH

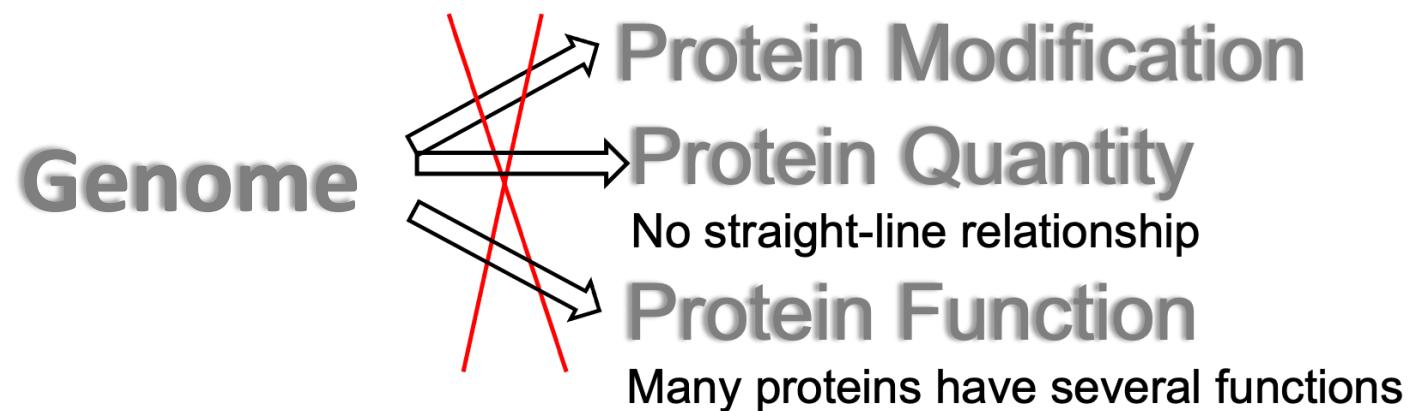
Introduction of Proteomics

- Determine the functions of genes and their products, allowing them to be linked into pathways and networks, and ultimately providing a detailed understanding of how biological systems work.

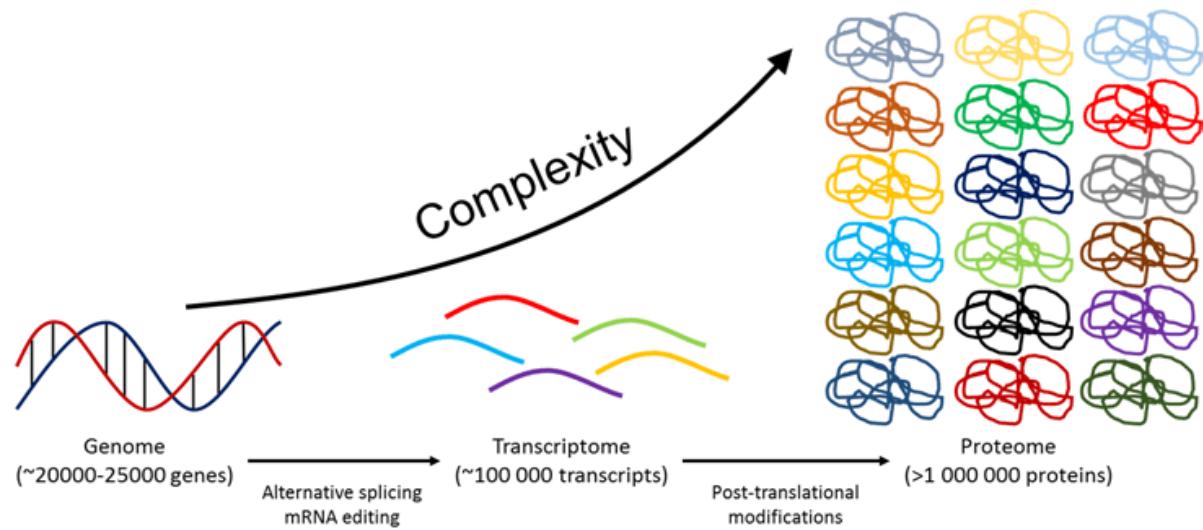


From Genomics to Proteomics

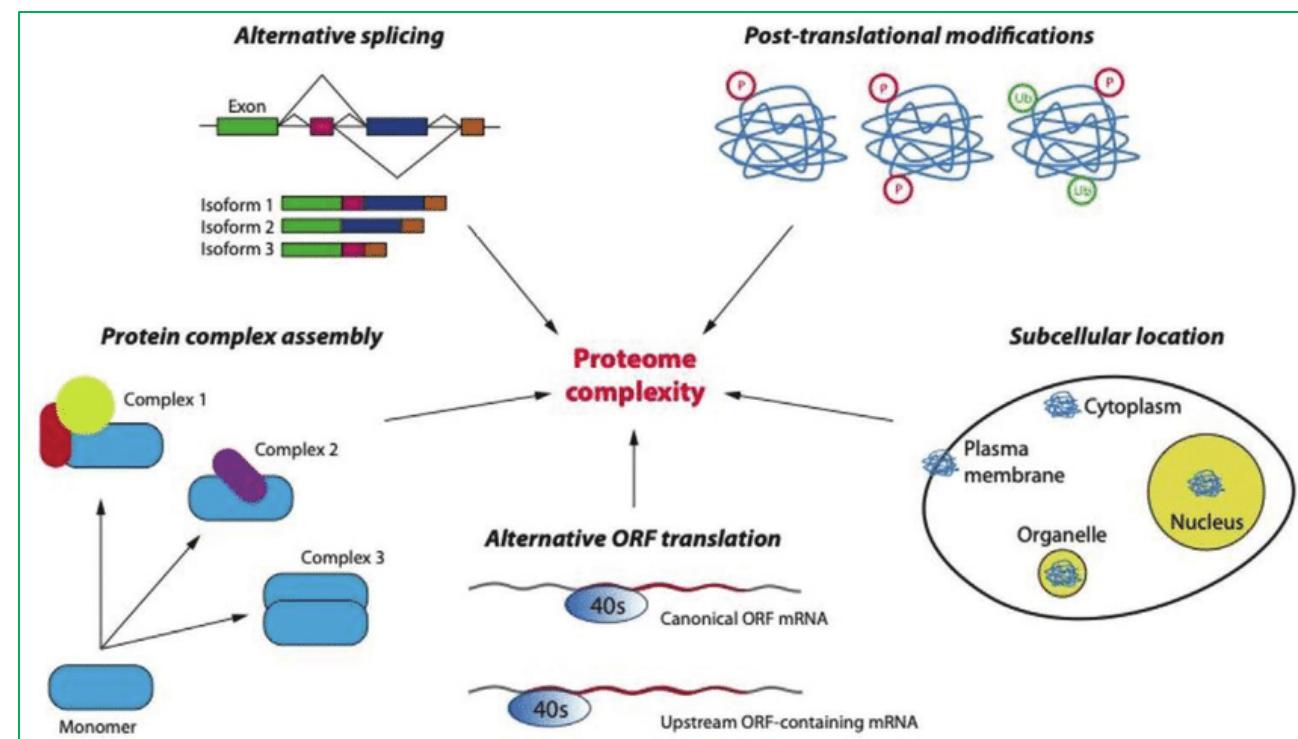
- Proteomics is a rapid growing area of molecular biology that is aimed to characterize the entire proteins of a cell line, tissue, or organism.



Complexity of Proteomics



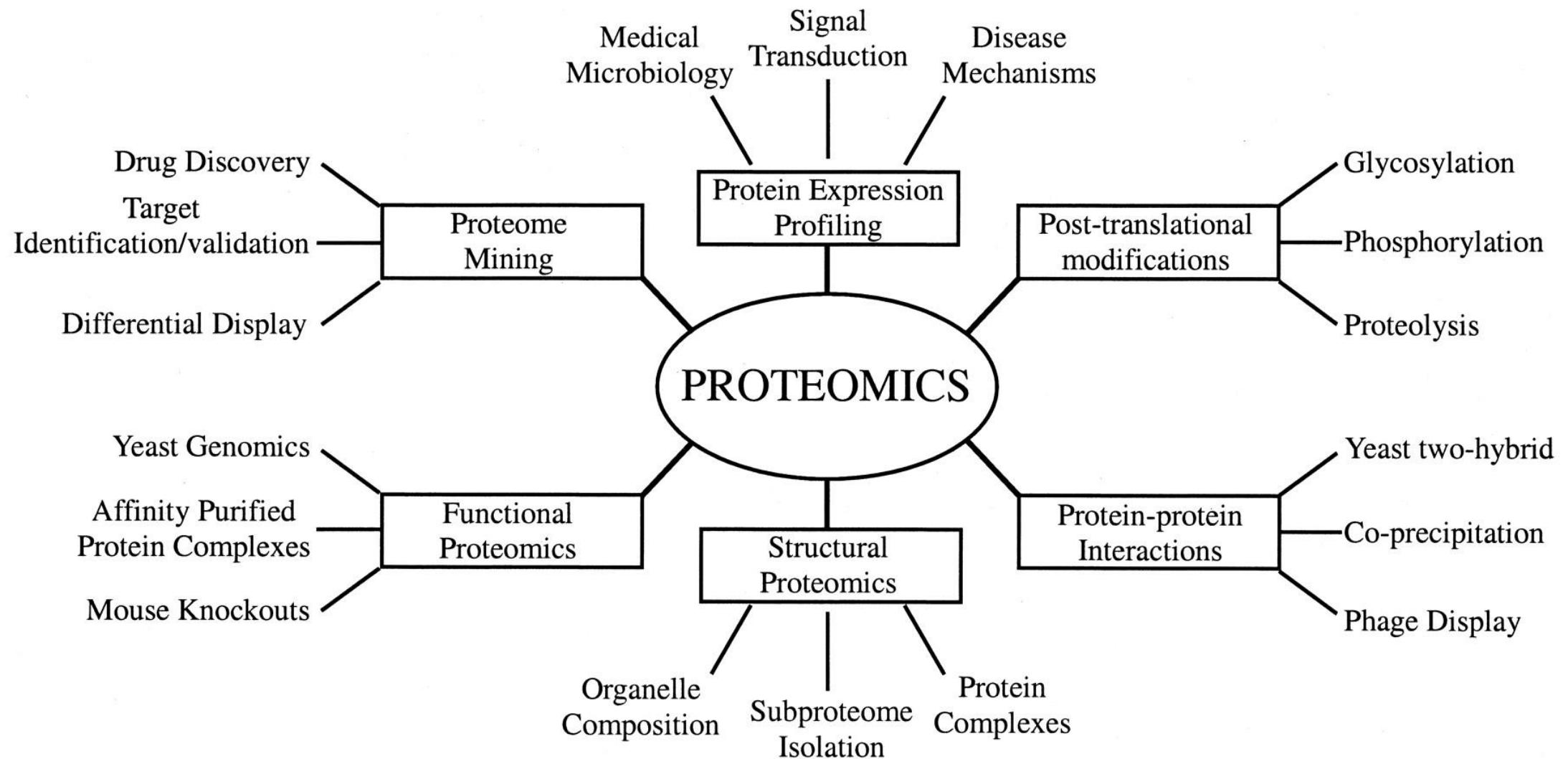
1. Numbers of proteins
2. Diversity of cells or tissues
3. Dynamic changes in protein levels



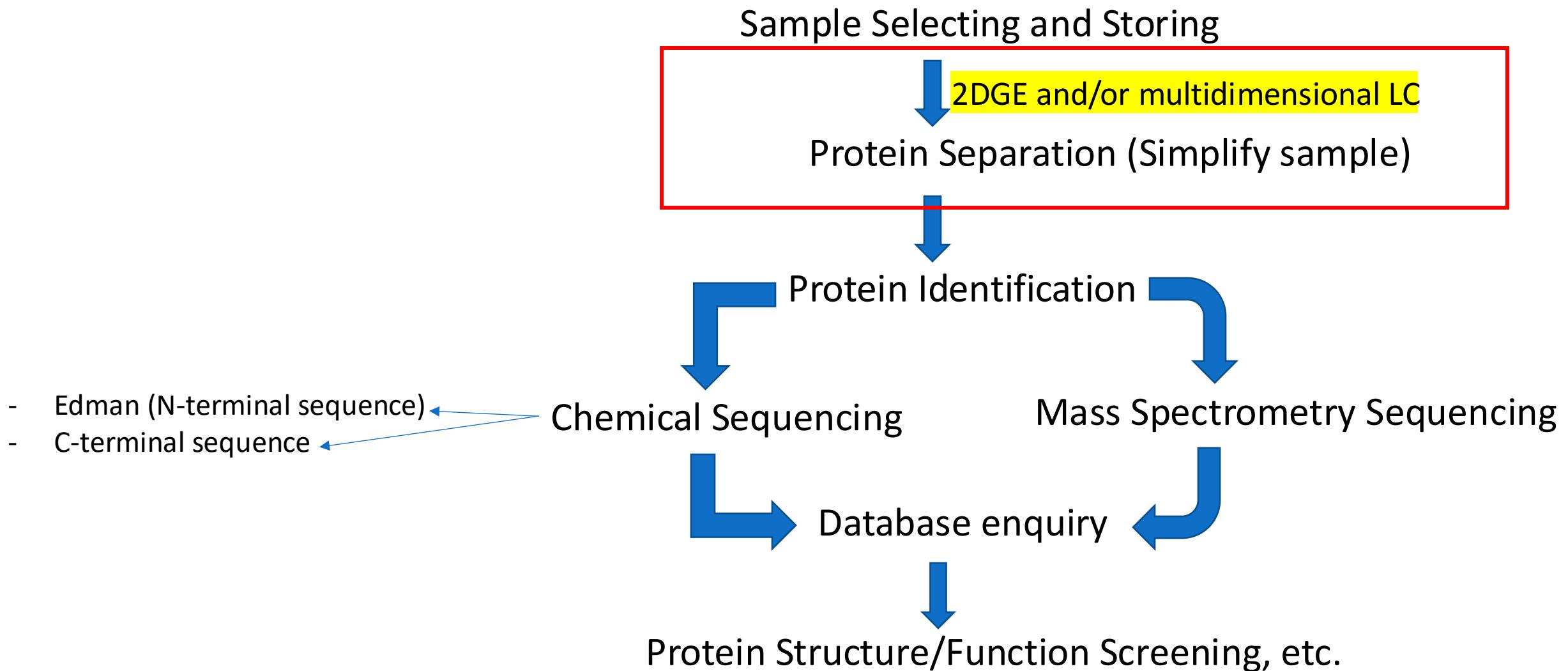
<https://www.researchgate.net/profile/Juan-Pelta/publication/342275076/figure/fig1/AS:904991556530176@1592778209035/Overview-of-proteome-complexity-Numerous-factors-contribute-to-the-generation-of-complex.ppm>

<https://www.researchgate.net/profile/Floris-Van-Den-Brink/publication/313535513/figure/fig4/AS:460194112118787@1486730234132/The-human-proteome-contains-many-more-species-compared-to-the-human-genome-making-it.png>

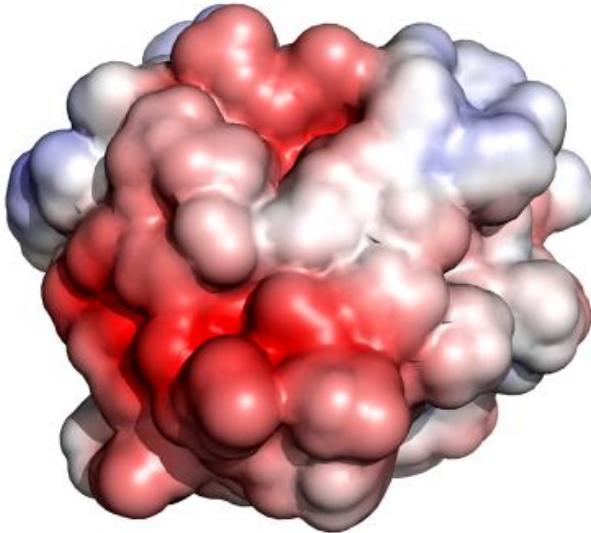
Proteomics Types and Application



General Workflow in Proteomics Analysis

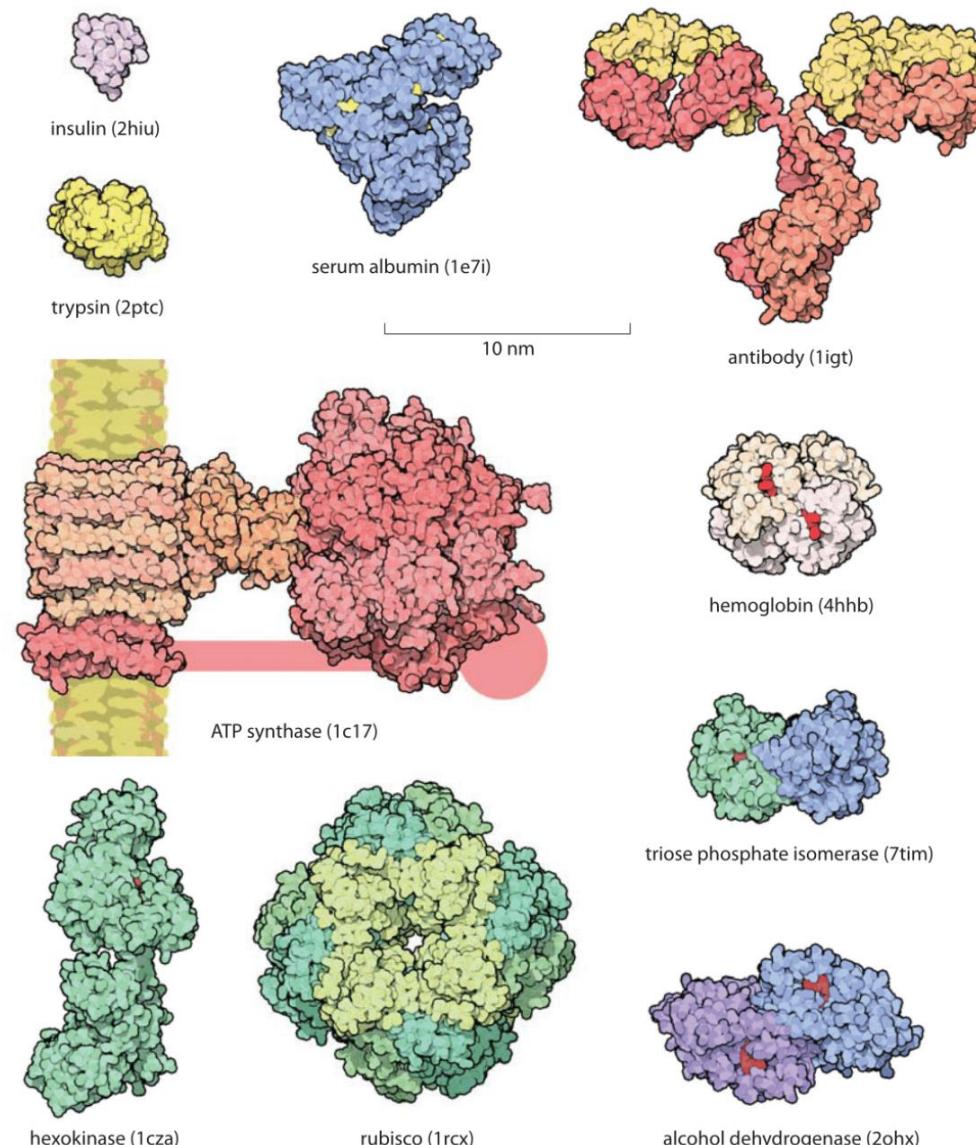


Proteins are Amphoteric Macromolecules with Different Sizes



- pH < pI, positively charged
- pH > pI, negatively charged

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



Methods for Protein Separation

Different sizes

- Ultracentrifugation
- Dialysis
- Size exclusion chromatography
- PAGE (SDS-PAGE or native PAGE)

Different solubility

Salting out
e.g. Ammonium sulfate

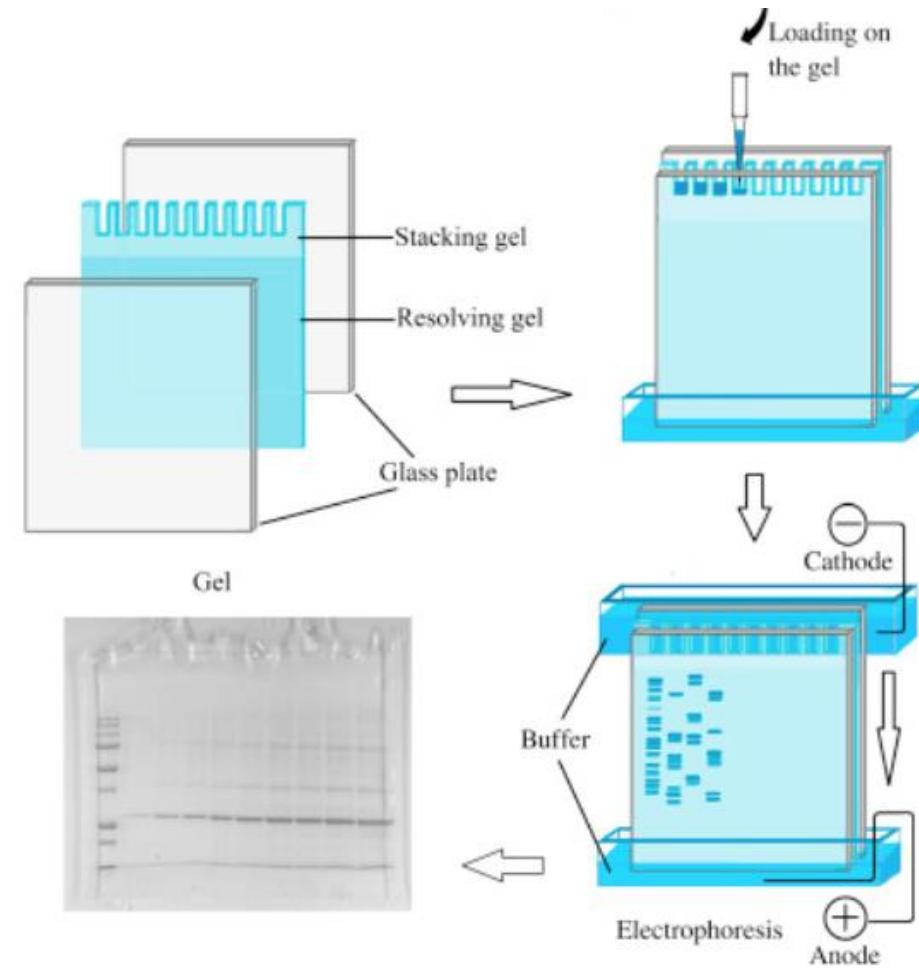
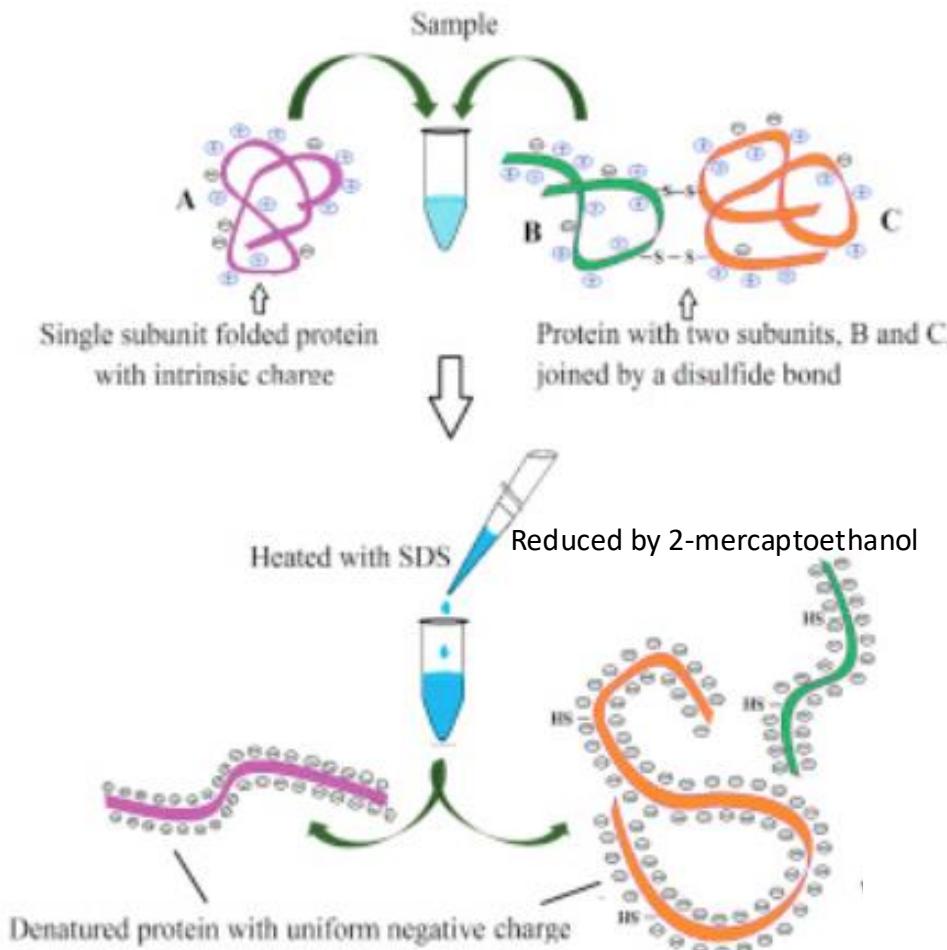
Different charges

- Ion exchange chromatography
- Electrophoresis

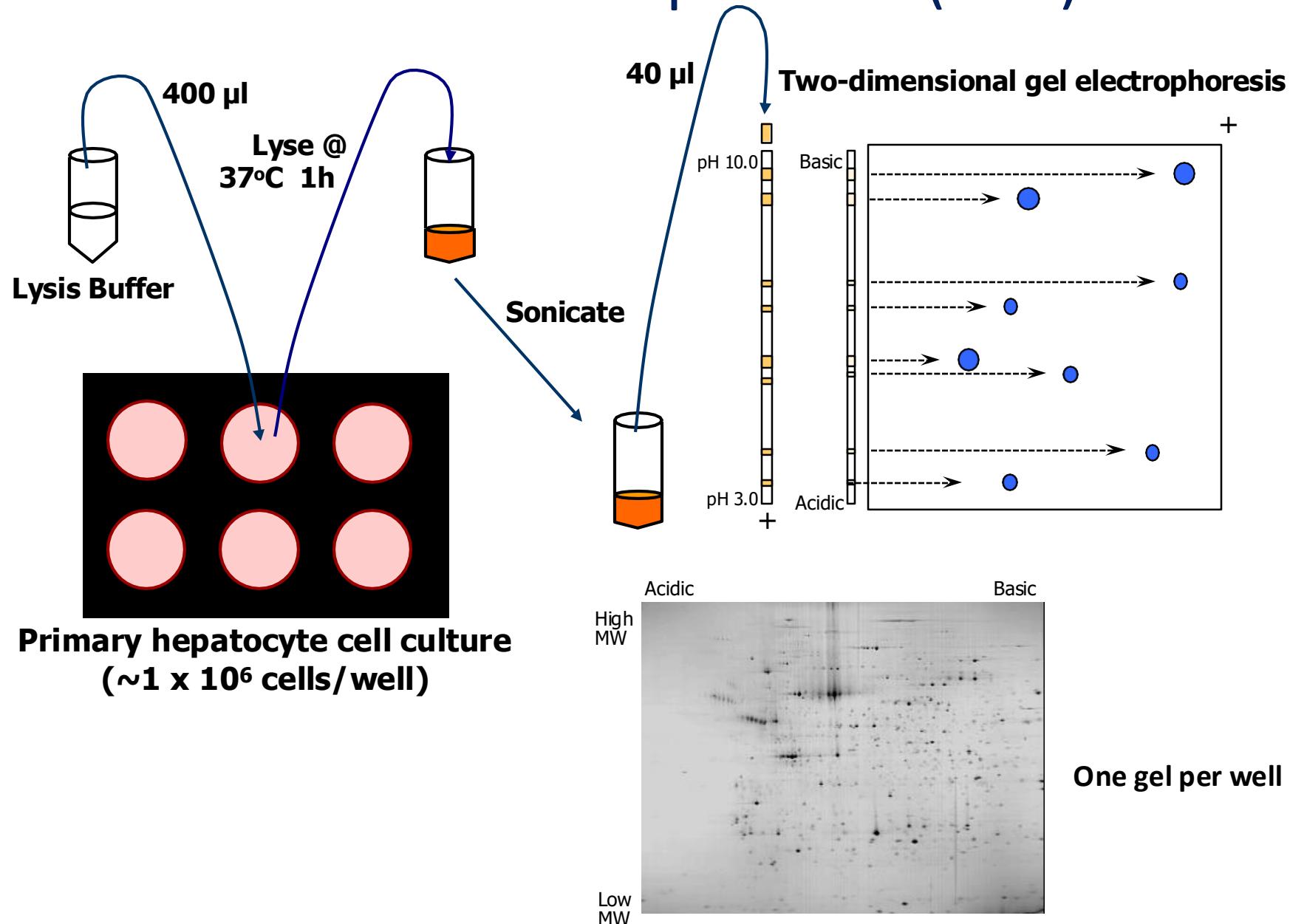
Different Ligand Binding

Affinity chromatography

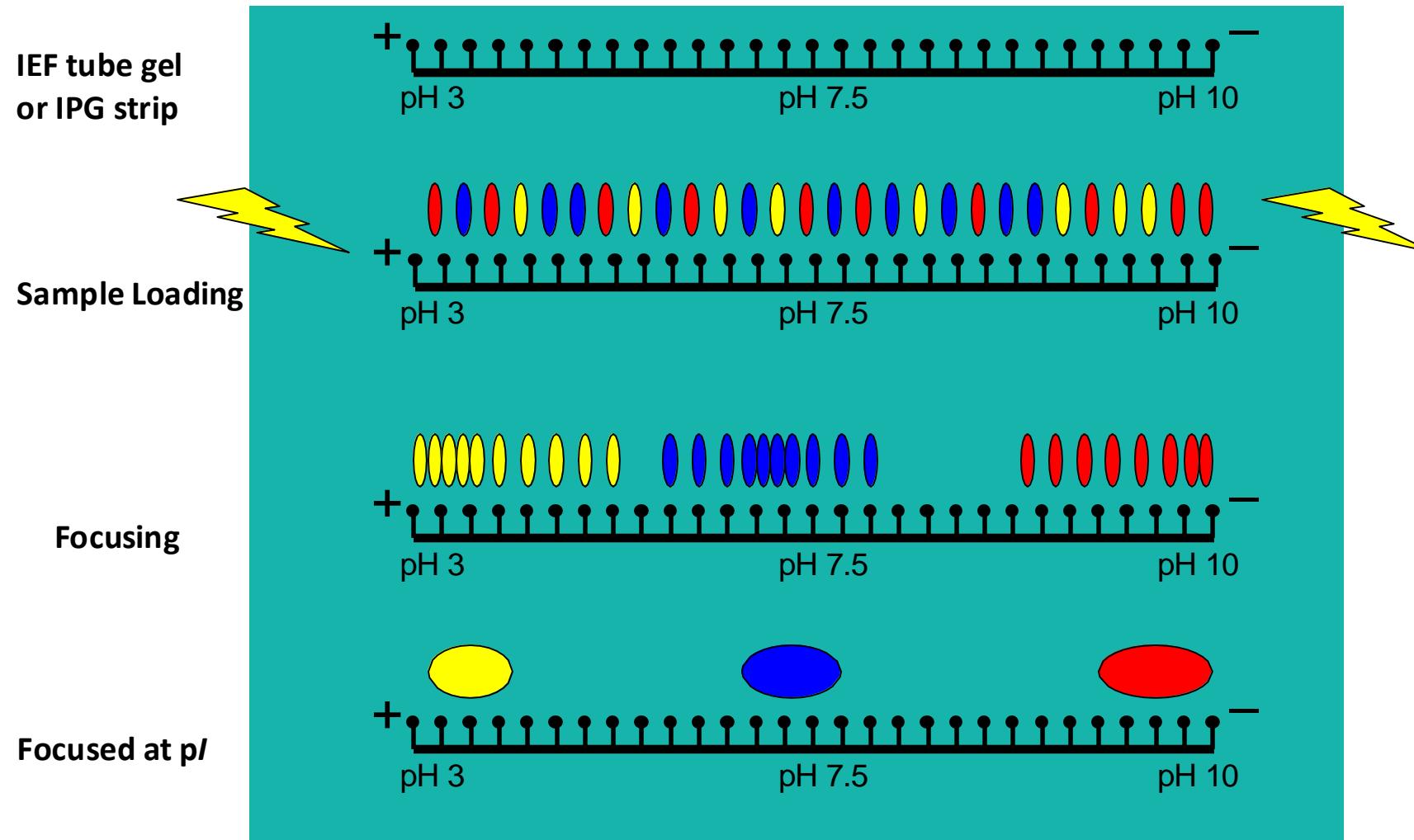
SDS-PAGE (sodium dodecyl sulphate–polyacrylamide gel electrophoresis)



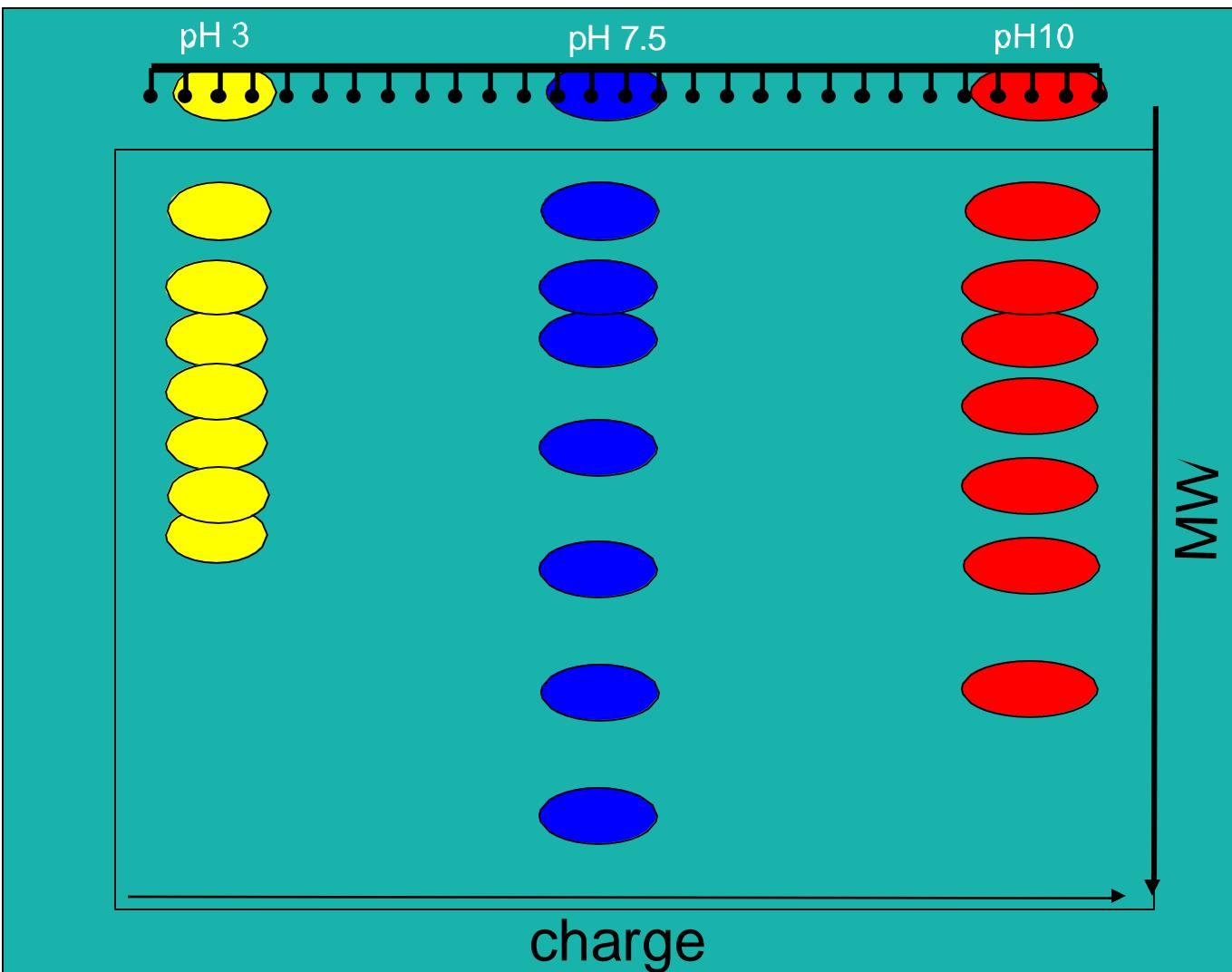
Two-dimensional Gel Electrophoresis (2DE)



First Dimension: Isoelectric Focusing



Second Dimension: SDS-PAGE



+

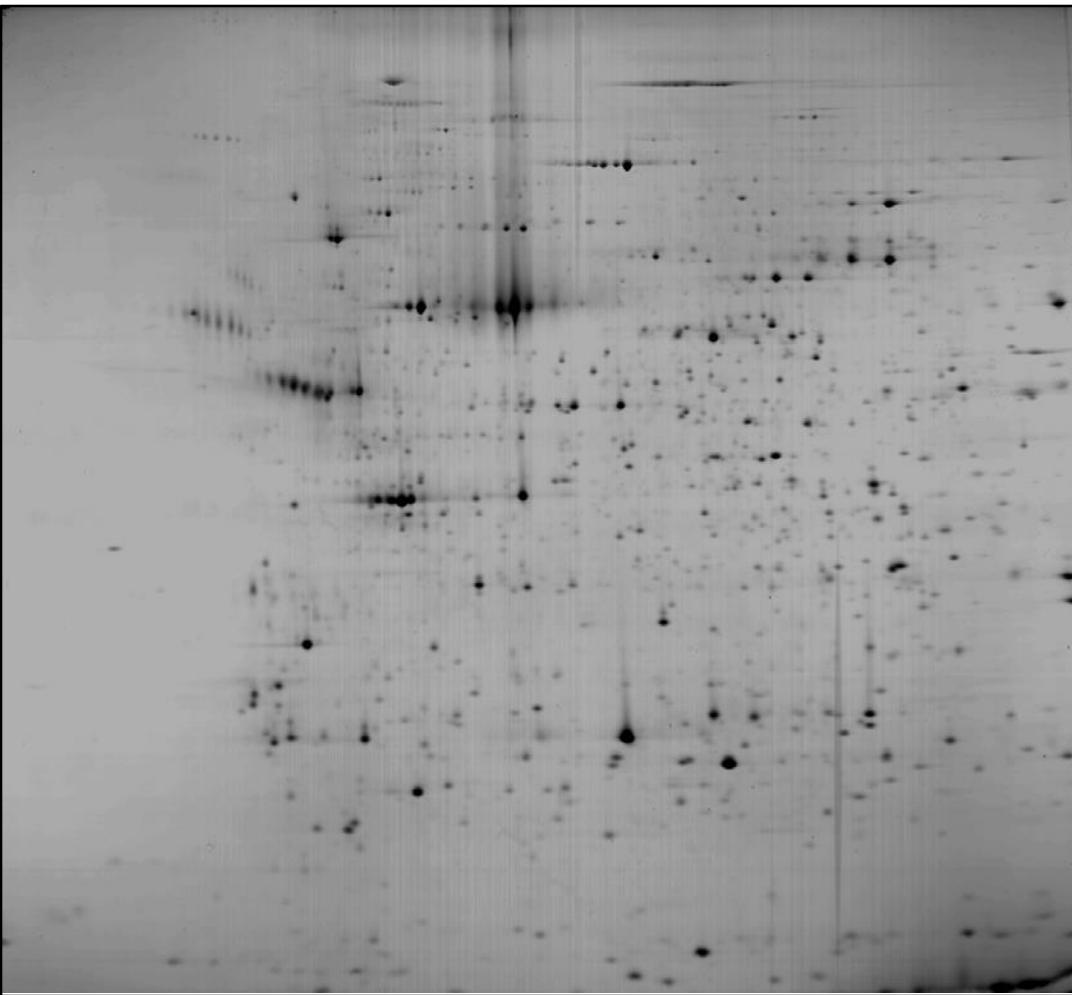
pH 3.0

-

pH 10.0

Isoelectric (pI) Focusing (pH gradient)

SDS-PAGE (gradient)



4%

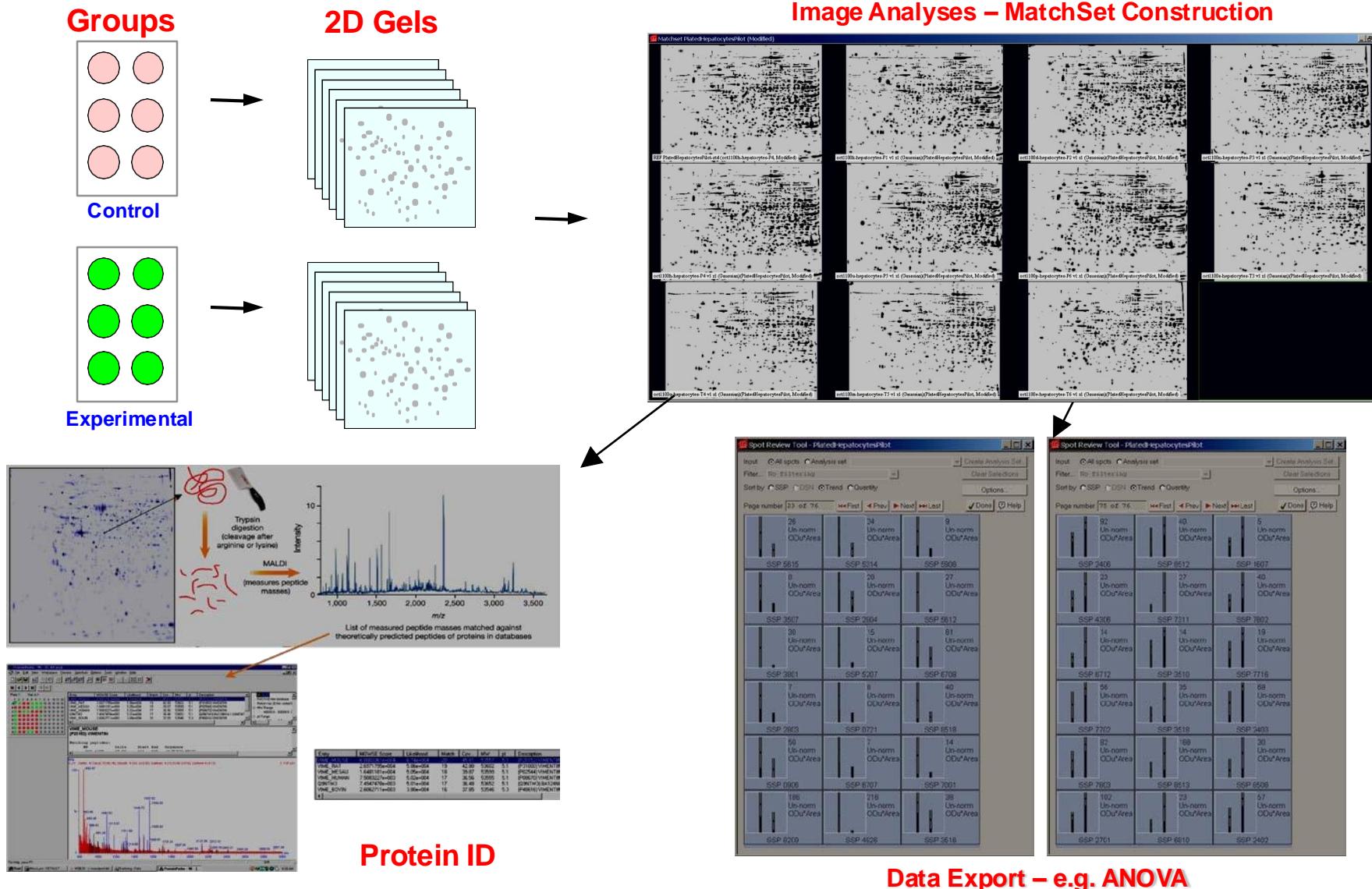
20%

+

Can resolve up to
2000+ proteins /gel

Amount of proteins
varies up to 10^9 -fold
in a cell lysate.

Two-dimensional Gel Electrophoresis (2DE)

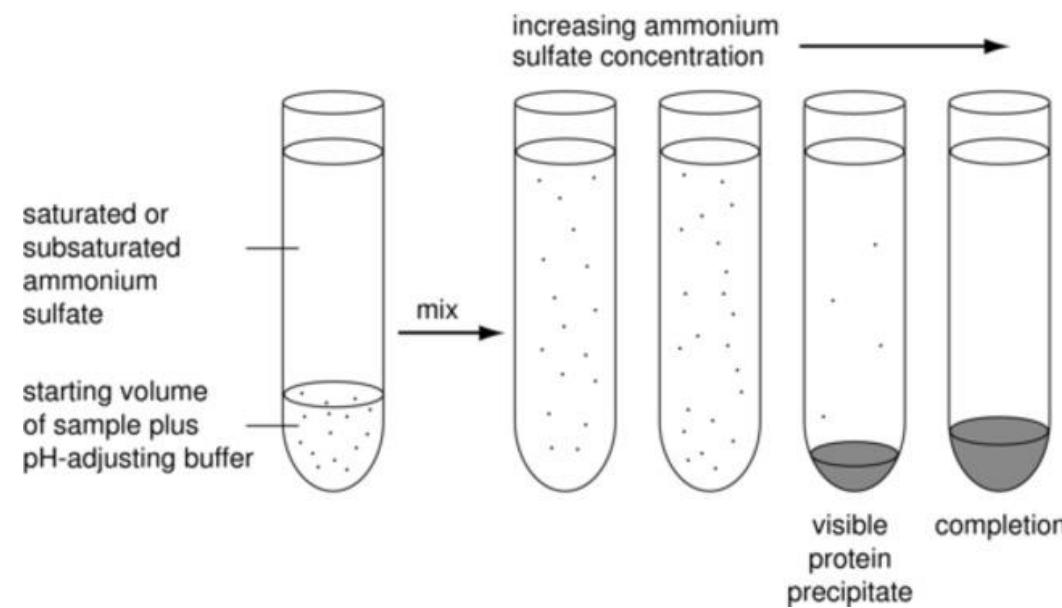


Application and Weakness of 2DE

- **Broad-based Screening of Protein Expression**
 - up- and down-regulation
 - post-translational modification
 - protein identification
- **WEAKNESSES**
 - Labor Intensive and time-consuming (~2 days, only one sample per gel)
 - Many large or hydrophobic proteins can not enter the 1st dimension gel or proteins with extreme acidity or basicity are not well represented.
 - Dynamic range of protein expression is problematic (e.g. low copy proteins)

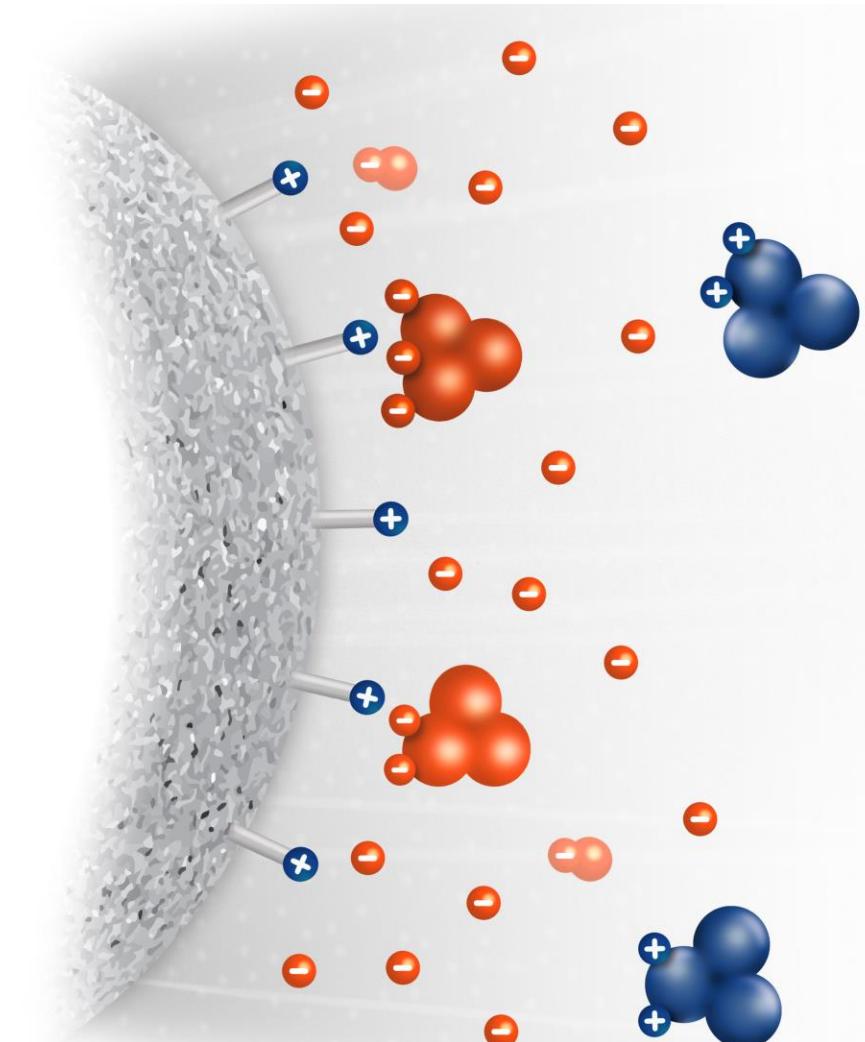
Salting out

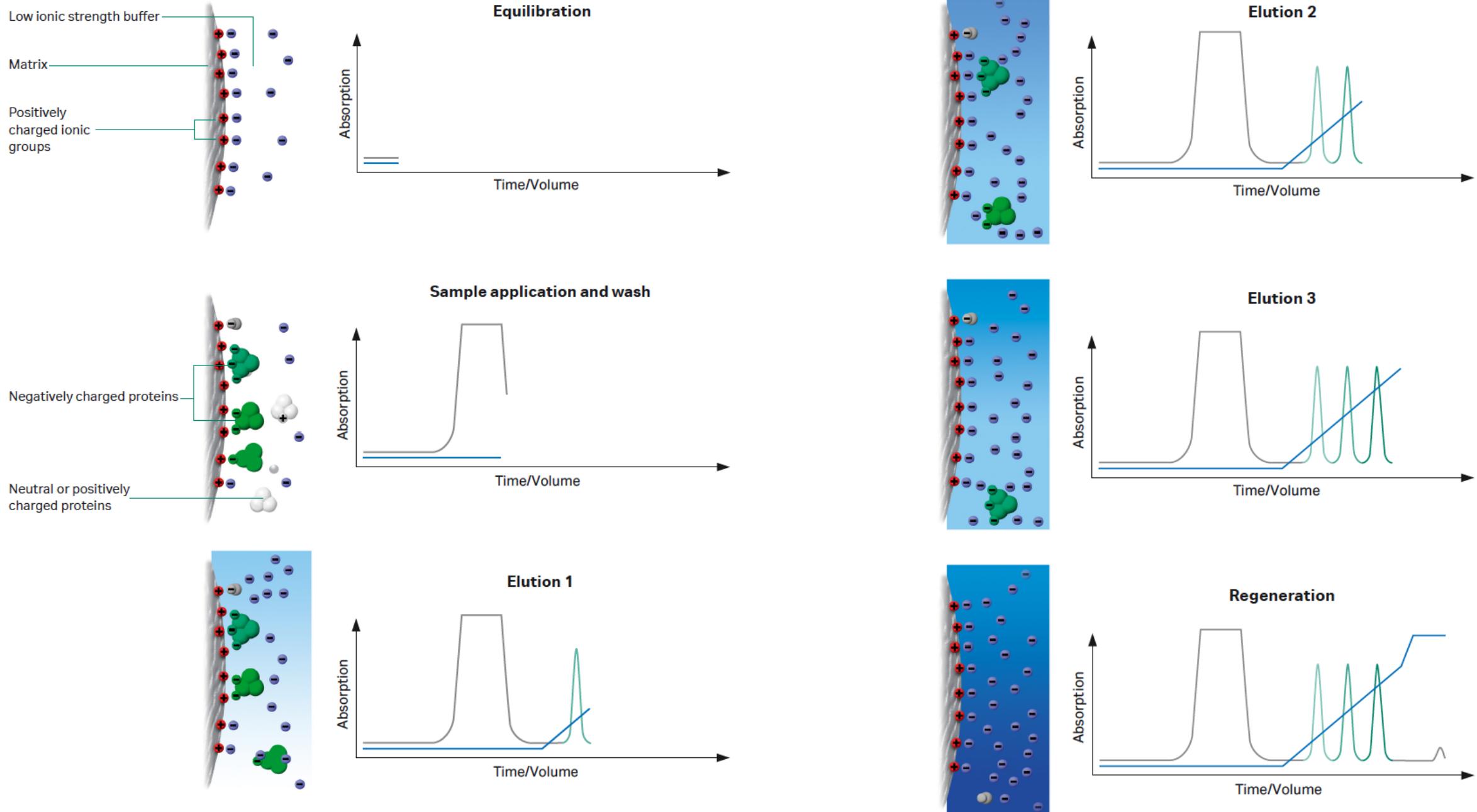
- A purification method that relies on the basis of protein solubility.
 - Most proteins are less soluble in solutions of high salt concentrations because the addition of salt ions shield proteins with multi-ion charges.
- Most common method is ammonium sulfate precipitation
 - Cheaper
 - Water soluble
 - No denaturation (change solubility)
 - But it requires prior knowledge of the protein's solubility.
- Useful in concentrating proteins



Ion Exchange Chromatography (IEC)

- Net surface charge is highly pH dependent
 - $\text{pH} < \text{pI}$, positively charged
 - $\text{pH} > \text{pI}$, negatively charged
- 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)





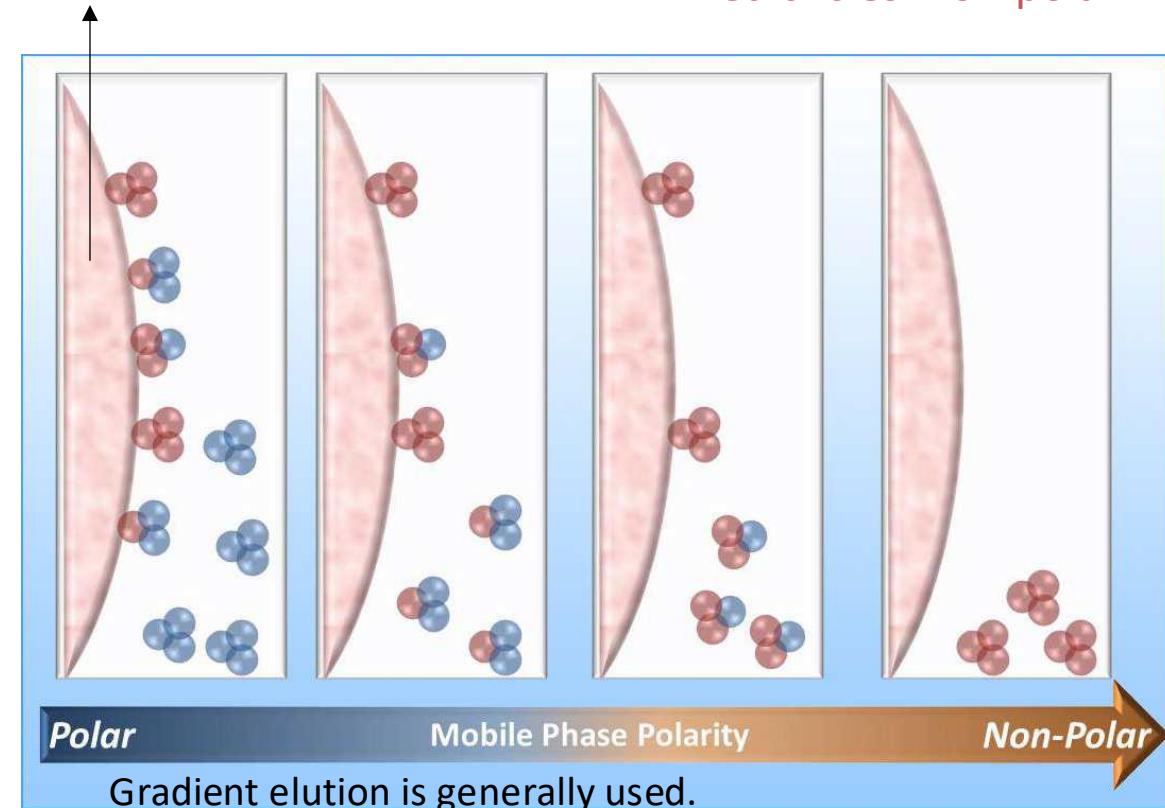
Reverse-Phase Chromatography

- Based on the surface hydrophobicity of molecules.
 - Shorter alkyl chains (C4 and C8) are typically preferred for intact protein separation because they are less retentive
 - RPLC is most commonly applied as the final dimension of separation. This is due to the solvent used in RPLC is compatible with MS.

Solid phase: nonpolar carbon chains (from C2-C18) with various modifications

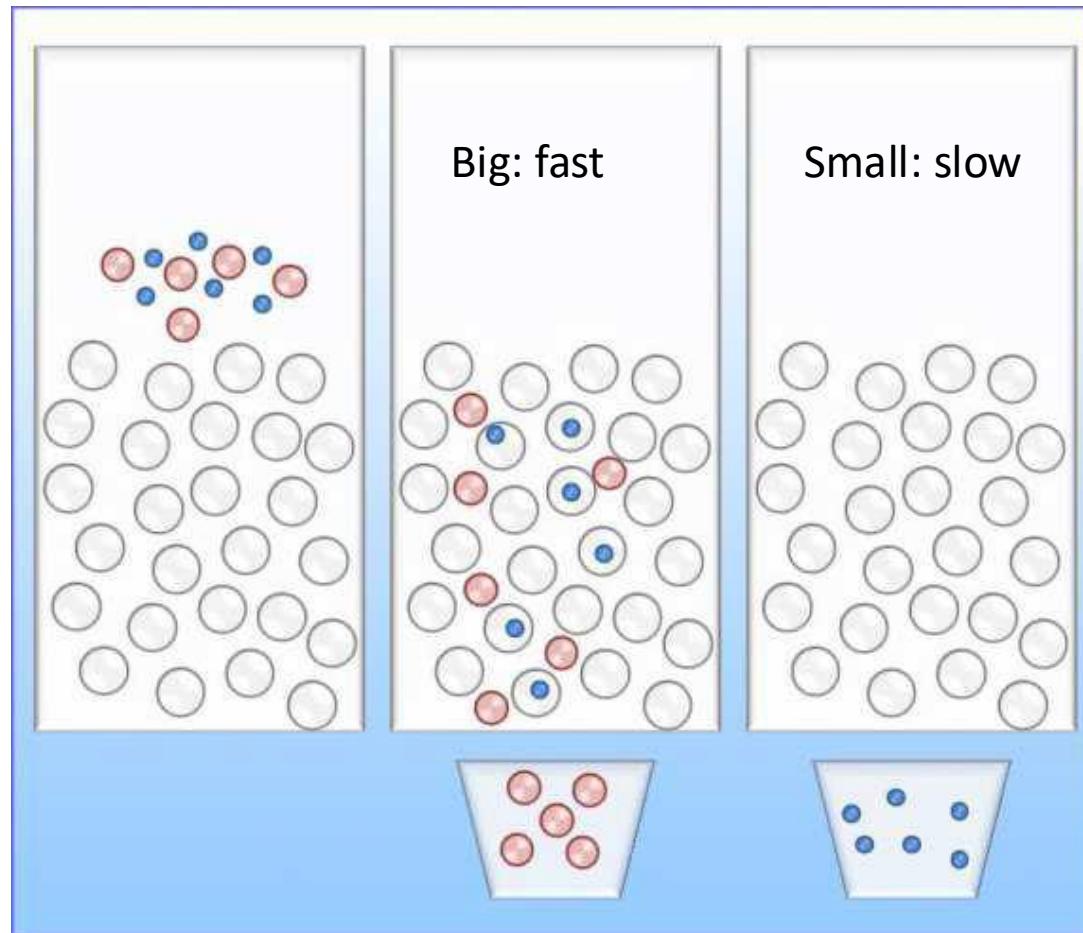
Blue circles: polar

Red circles: non-polar

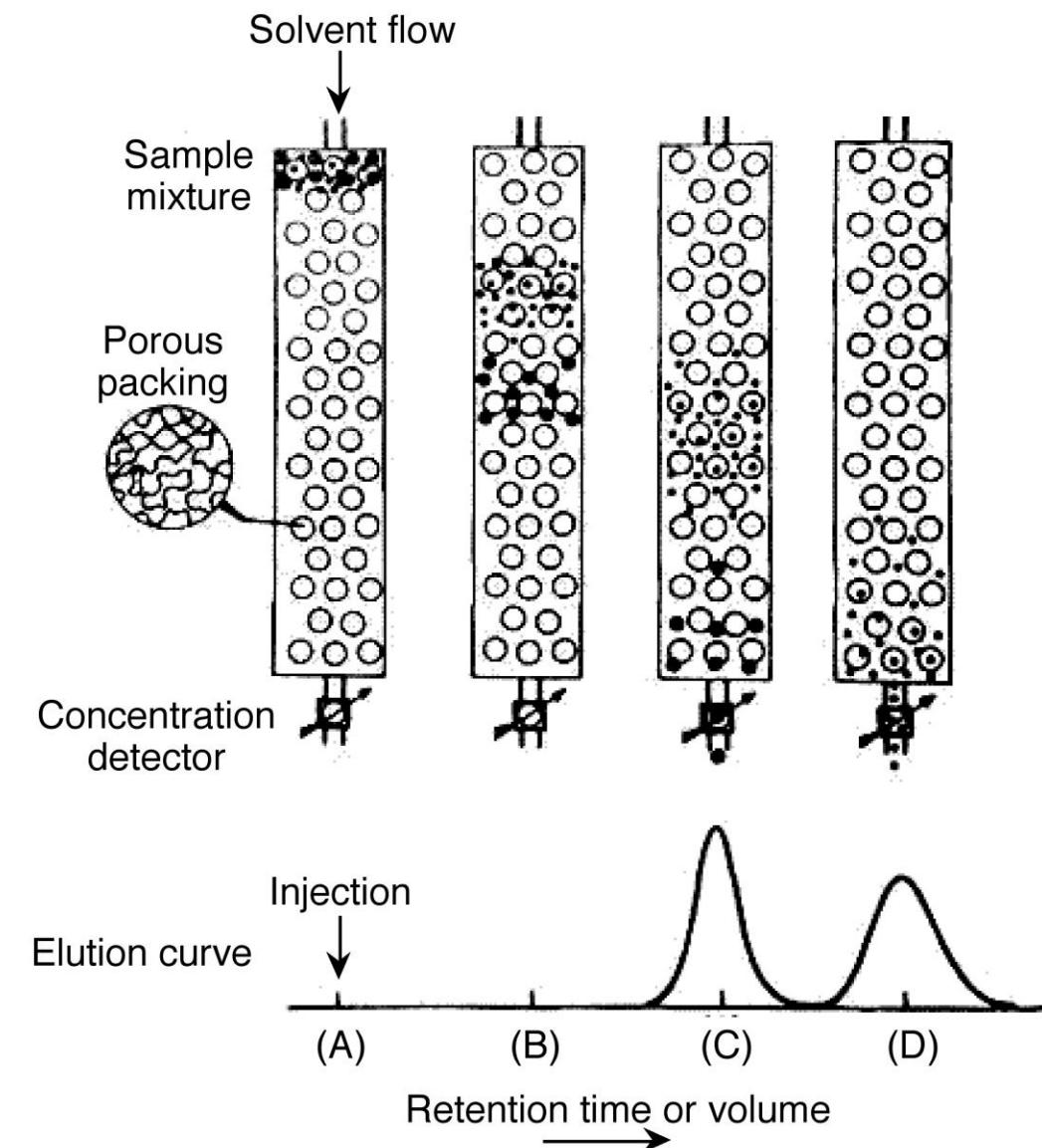


0.1% formic acid in water, acetonitrile with 0.1% formic acid

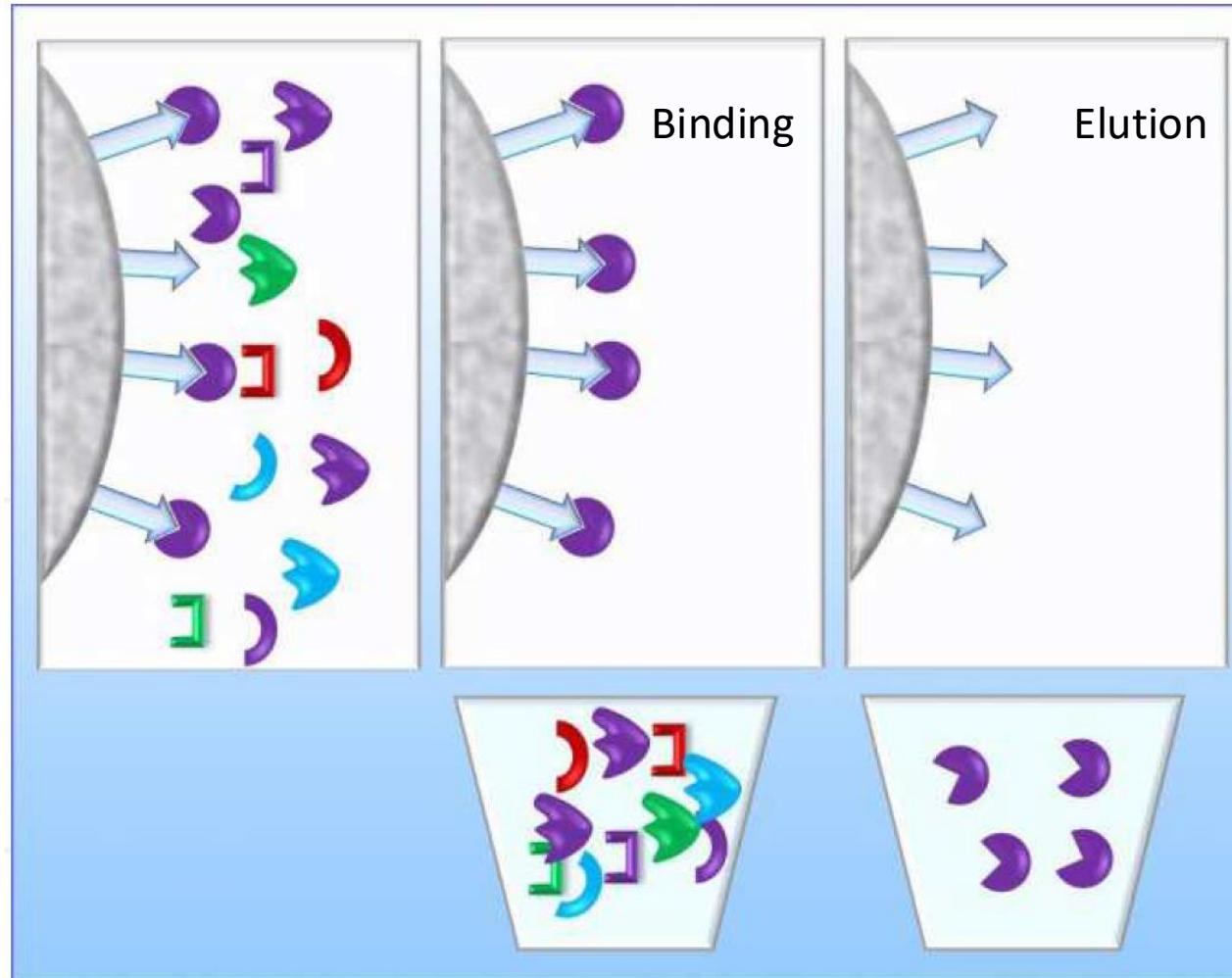
Gel Filtration/Size Exclusion Chromatography



Time sequence → (A) sample injected (B) size separation (C) large solutes eluted (D) small solutes eluted

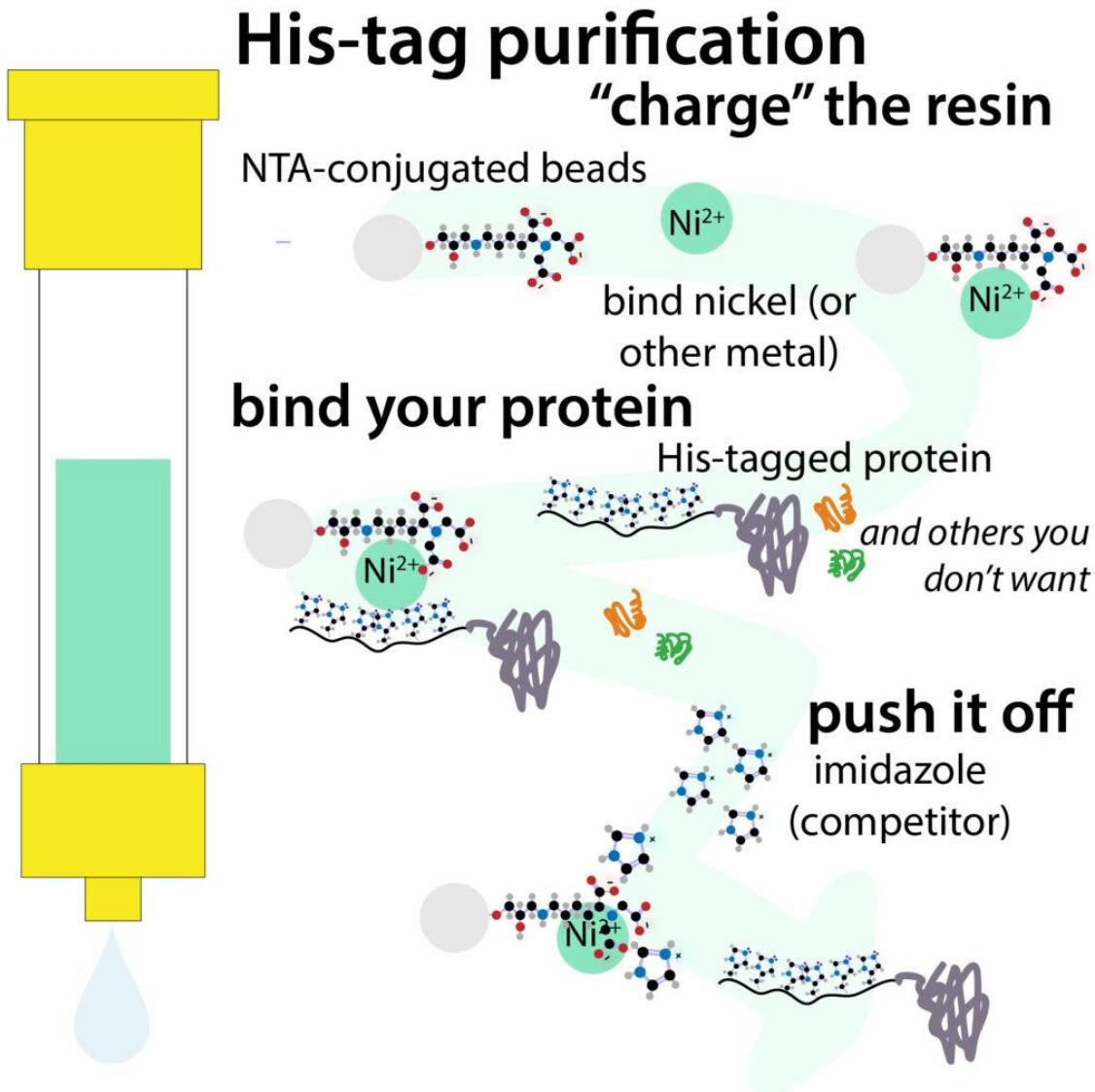


Affinity Chromatography



- Separate proteins or peptides based on their specific, ligand binding affinity.
- Metals (Ni^{2+}): 6x His tag
- Phosphoprotein/peptide
- Proteins binds to specific drug or substrate
- Isolate proteins that interact to form a complex

Ni-Affinity Chromatography



Different Types of Chromatography: Summary

1. **Gel filtration/size exclusion** - separates by size (molecular weight) of proteins
- first 2. **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
3. **Hydrophobic interaction** - separates by hydrophobicity of Proteins
4. **Affinity** - separates by some unique binding characteristic of protein of interest for affinity matrix in column
No! (do not know the protein)