SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOINFORMATICS

UNIT – 1- SBIA1303 – Proteomics and Interactomics

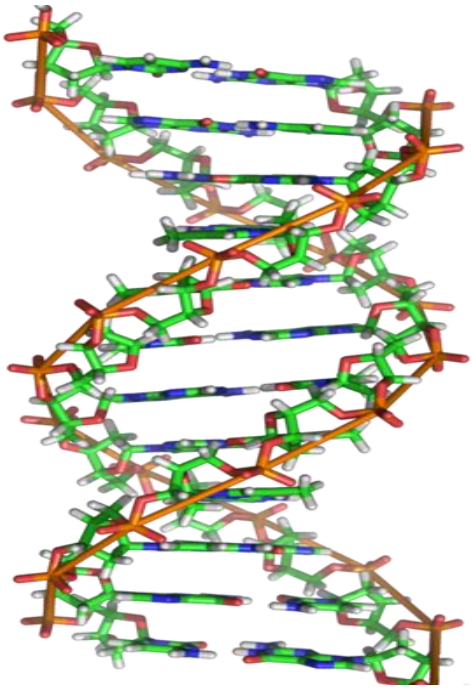
SBIA1303 -PROTEOMICS AND INTERACTOMICS

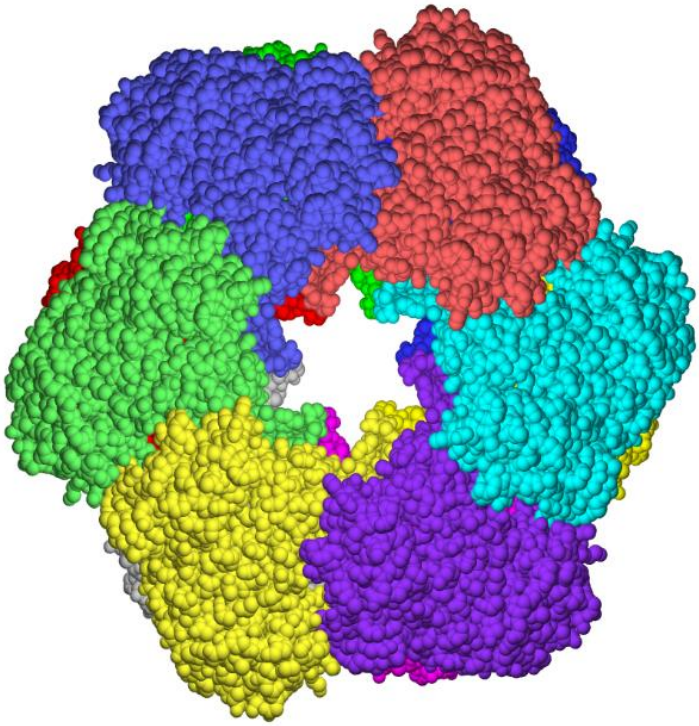
- *Unit I Introduction*
- *Protein structure organization. Introduction to proteomics. Branches of proteomics. Applications and Scope of Proteomics.*
- Structural Biology
- The study of structure-function relationships in macromolecules
- The importance of macromolecules

Macromolecules:

- Large (10^3 – 10^6 atoms)
- Polymeric – made of smaller building blocks that self-assemble into long chains with complex 3D structure
- Have key roles in living organisms

- Proteins – numerous roles?
- Nucleic acids (DNA, RNA) - encoding and expression of genetic information?





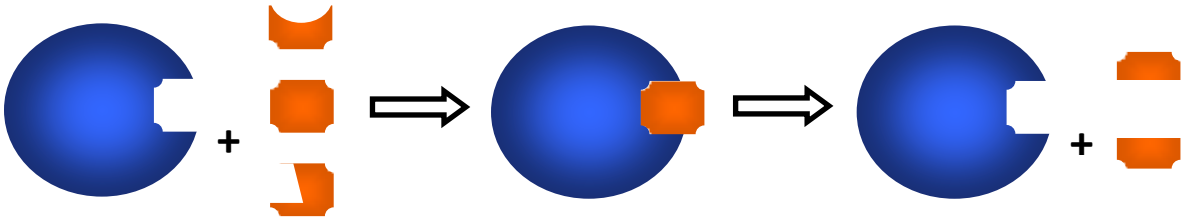
(Glutamine synthase)

Why Are Proteins Interesting?

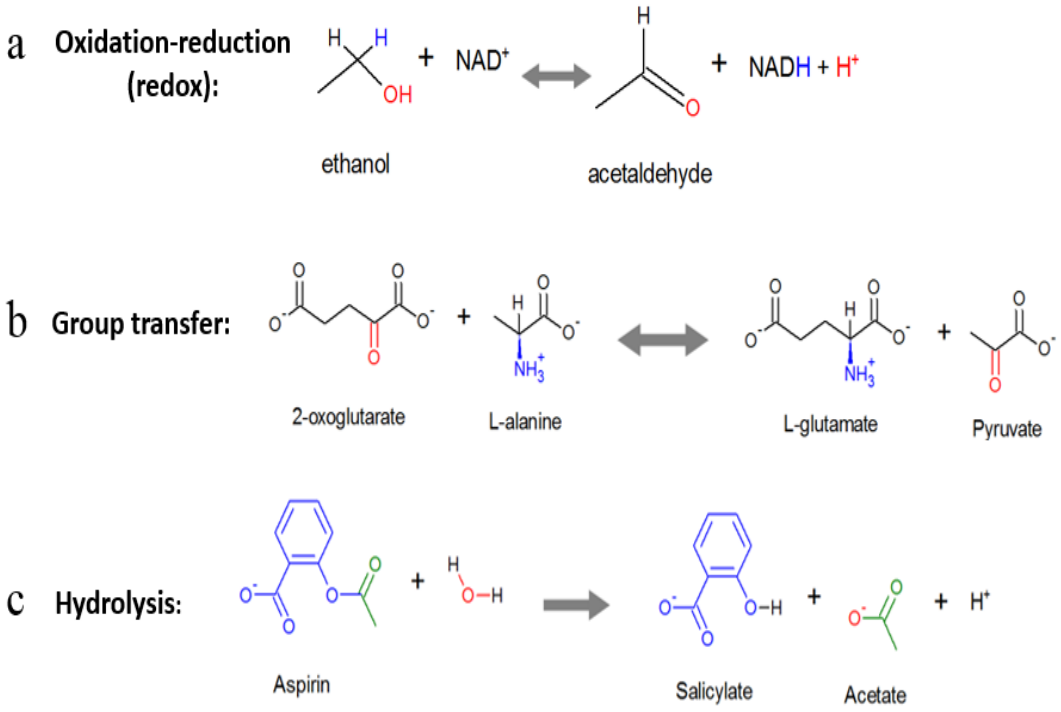
1. The most abundant macromolecules:
    - ~50% of the cell's dry mass
    - No. of proteins ≈ 4 times the number of their coding genes
    - Up to 15,000 different proteins in one cell
2. Functionally diverse:
    - Catalysis of metabolic processes
    - Energy transfer
    - Gene expression
    - Transport of solutes across membranes
    - Cellular communication
    - Molecular recognition
    - Defense
    - Forming intracellular & extracellular structures
3. Protein Roles: Enzymatic Catalysis
    - Life processes are based on thousands of chemical reactions (metabolic pathways, transport, mechanics, etc.)
    - Cellular needs: 10^{-5}-10^{2} seconds for each reaction
    - Chemical reactions: seconds to 10^{9} years (room temp)

Enzymes:

- Efficient (10^{11}-10^{16} acceleration, up to 10^{21})
- Specific (substrate, reaction)
- Can be controlled
- Can couple endergonic to exergonic reactions



Enzymes are reaction-specific



Examples of the six types of reactions catalyzed by enzymes, according to the EC method.

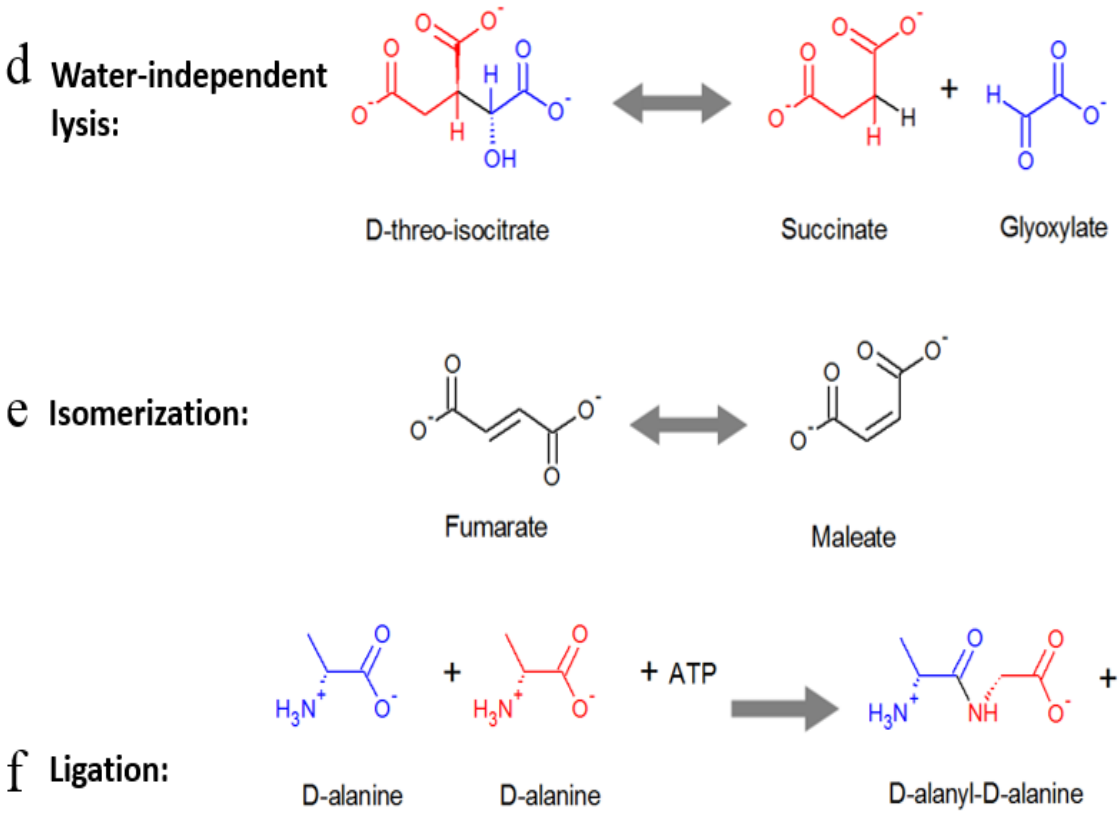
(The individual reactions are adapted from the MetaCyc database.)

For clarity, explicit hydrogens are not shown, except around centers in which the number of hydrogen atoms changes during the reaction.

(a) Oxidation-reduction (redox): oxidation of ethanol to acetaldehyde, catalyzed by *alcohol dehydrogenase*.The oxidation involves the transfer of a hydride species (blue) from ethanol's C_α to NAD^+ and the deprotonation of its hydroxyl group (red).

(b) Group transfer: amino transfer from alanine to α-ketoglutarate, catalyzed by *alanine aminotransferase*. The transfer of the amino group (blue) from the first co-substrate involves reduction of the corresponding carbon atom to a keto group (red), and vice versa in the other co-substrate. However, since these events involve internal electron transfer between amino and keto groups, the reaction is not considered to be redox (see more in subsection 9.1.5.2.1).

(c) Hydrolysis: breakdown of aspirin to salicylate (blue) and acetate (green) by using water (red), as catalyzed by *aspirin hydrolase* [Θ3].



(d) Water-independent cleavage of covalent bonds: breakdown of isocitrate to glyoxylate (blue) and succinate (red), catalyzed by *isocitrate lyase.*

(e) Isomerization: interconversion between fumarate (*trans* bond) and maleate (*cis* bond), catalyzed by maleate *cis-trans isomerase*.

(f) Ligation: the attachment of two D-alanine molecules (blue and red), catalyzed by *D-alanine-D-alanine ligase*

Protein Roles: Energy Transfer

Proteins extract energy from the environment

- *The action of virtually all proteins involves binding of a ligand (small molecule, another macromolecule).*
- Proteins' functionality results from their chemical and structural complexity
- Ligand binding occurs at specific proteins sites
- The binding involves non-covalent interactions of ligand with amino acids in the binding site

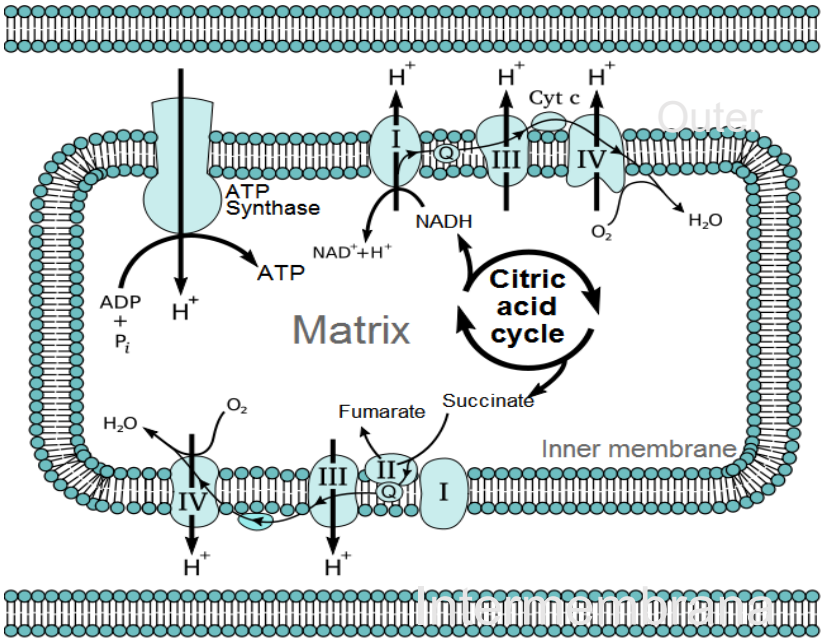
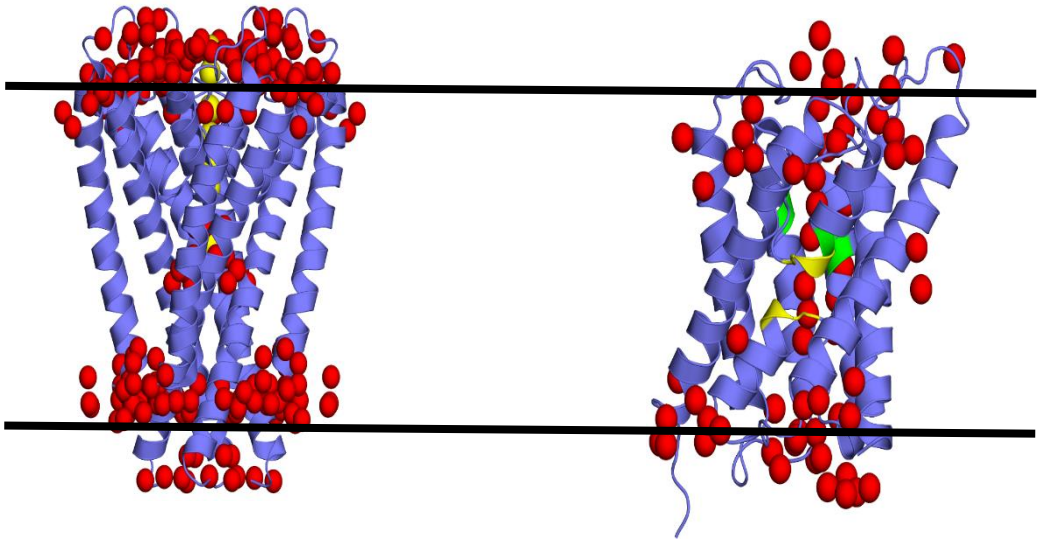


Figure :Cellular respiration within mitochondria.

The schematic illustration shows the main players of this multi-component system.

- The citric acid (Krebs) cycle completes the aerobic oxidation of foodstuff inside the mitochodrial matrix.
- The high-energy electrons released by this process are first stored as NADH & $FADH_2$, then passed on to the first component of the electron transport chain, embedded within the inner mitochondrial membrane.
- The electrons pass through several cytochromes and finally passed to molecular oxygen, which is reduced to water.
- The electron chain contains 3 protein complexes (I, III, IV), which use the energy released upon electron transport to pump protons from the matrix into the intramembrane space.
- The energy stored in the gradient is released when the latter dissipates via ATP synthase, a protein complex that couples proton movement to ATP synthesis from ADP and inorganic phosphate.
- Protein Roles: Import and export of solutes
- Some membrane-bound proteins transfer ions and molecules across the cell's membrane
- These include channels and transporters (carriers)

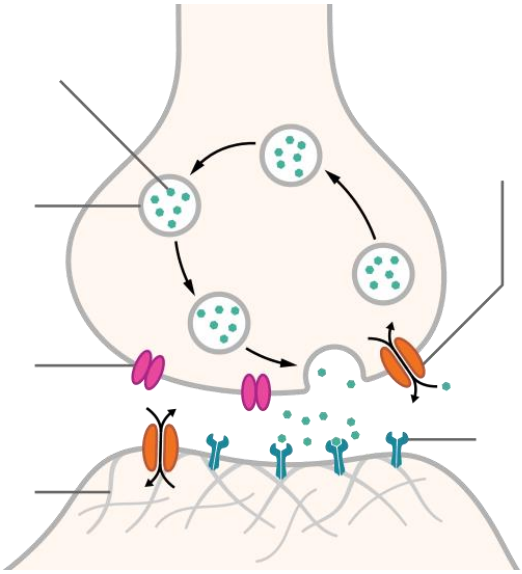


. KcsA (K^+) Aquaporin

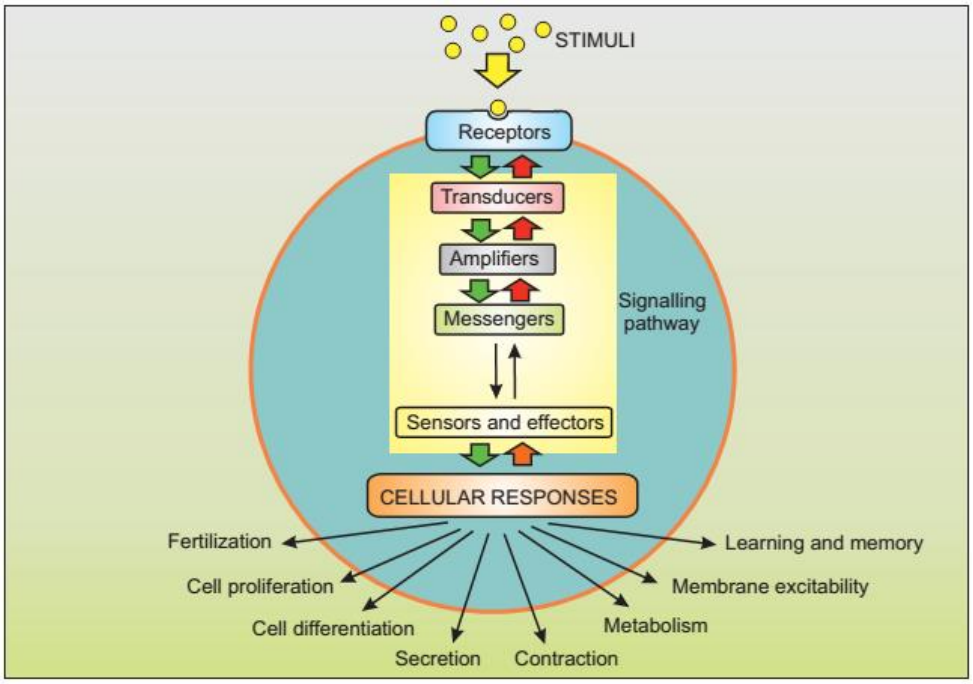
Protein Roles: Cellular Communication

Many cells communicate with one another via chemical messengers

Examples: neurotransmitters (nervous system), hormones (endocrine system) – between distant cells



The messengers bind to membrane receptors, which are proteins, and their message is relayed into the cell



The transduction of the external signal into the cell requires multiple elements, many of which are proteins

Protein Roles: Cellular Communication

Example: GPCR activation

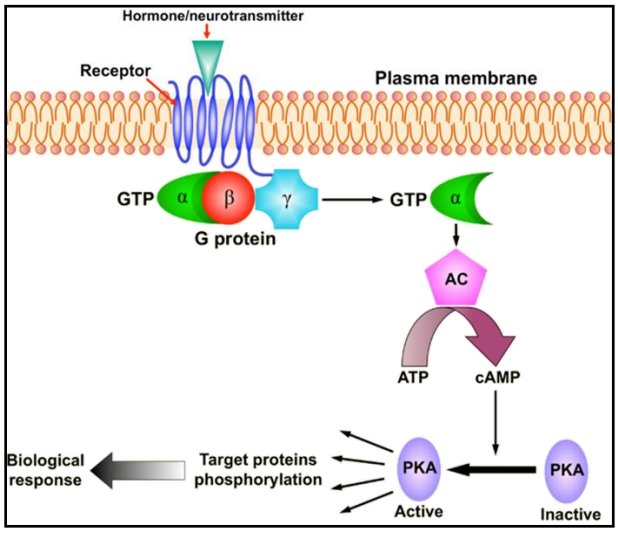


Figure . The main players in a typical signal transduction cascade.

The cAMP-PKA cascade.

- Binding of an external chemical messenger (hormone, neurotransmitter, etc.) to a membrane-bound protein receptor induces the activation of an enzyme called a G-protein, which acts as a transducer. This activation makes one of the protein's subunits detach from the other, bind to the enzyme adenylyl cyclase (AC), and activate it.
- Activated AC catalyzes the conversion of ATP into cyclic AMP, which acts as an intracellular messenger.
- It binds to and activates the enzyme amplifier PKA, which in turn phosphorylates a large set of cytoplasmic proteins.
- The phosphorylated proteins may activate other cellular components, or perform a certain function (that is, they may act as sensors and/or effectors).
- In any case, this signal transduction eventually leads to changes in the cell's behavior, i.e., to a biological response.

Figure . The main players in a typical signal transduction cascade.

The cAMP-PKA cascade.

- Binding of an external chemical messenger (hormone, neurotransmitter, etc.) to a membrane-bound protein receptor induces the activation of an enzyme called a G-protein, which acts as a transducer. This activation makes one of the protein's subunits detach from the other, bind to the enzyme adenylyl cyclase (AC), and activate it.
- Activated AC catalyzes the conversion of ATP into cyclic AMP, which acts as an intracellular messenger.
- It binds to and activates the enzyme amplifier PKA, which in turn phosphorylates a large set of cytoplasmic proteins.
- The phosphorylated proteins may activate other cellular components, or perform a certain function (that is, they may act as sensors and/or effectors).
- In any case, this signal transduction eventually leads to changes in the cell's behavior, i.e., to a biological response.
- Protein Roles: Cellular Communication
- Surface receptors also mediate direct cell-cell communication
- The receptor are chemically and physically adapted to recognize their target protein

- Protein Roles: Cellular Communication
- Protein-protein recognition is also part of cell's defense mechanisms

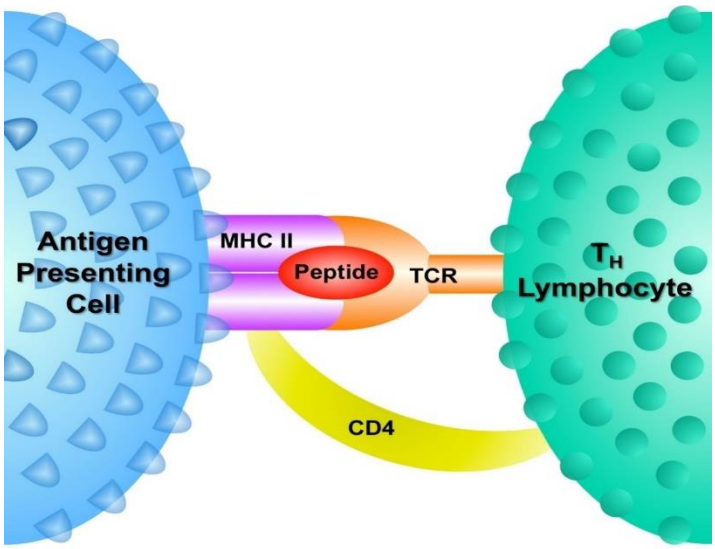


Figure.Proteins involved in the interactions between an antigen presenting cell and a T_H lymphocyte.

- Antigen-presenting cells (APC), such as macrophages, can internalize bacteria, cleave their proteins to peptides, and present these peptides onto their surface while bound to some of their own proteins, called MHC-II (major histocompatibility complex type II).
- T_H lymphocytes, a type of immune cells, can recognize the APC-foreign antigen complex using their own cell-surface proteins. Specifically, the T-cell receptor (TCR) of the lymphocyte specifically recognizes the foreign antigen presented onto the APC, whereas another protein, called CD4 (complementarity determinant number 4), recognizes the MHC-II protein.
- There are other recognition proteins that are not presented here. The recognition described above allows the T_H lymphocyte to alert the rest of the immune system about the presence of a pathogen inside the body, which leads to the amounting of a full-scale immune response against it.

Figure.Proteins involved in the interactions between an antigen presenting cell and a T_H lymphocyte.

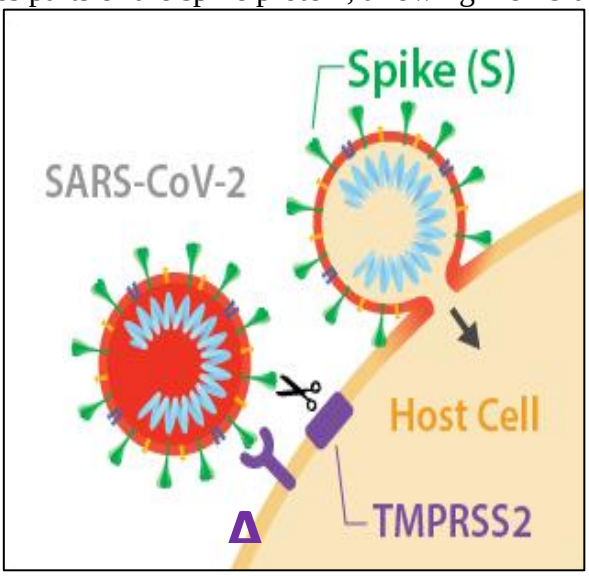
- Antigen-presenting cells (APC), such as macrophages, can internalize bacteria, cleave their proteins to peptides, and present these peptides onto their surface while bound to some of their own proteins, called MHC-II (major histocompatibility complex type II).

- T_H lymphocytes, a type of immune cells, can recognize the APC-foreign antigen complex using their own cell-surface proteins. Specifically, the T-cell receptor (TCR) of the lymphocyte specifically recognizes the foreign antigen presented onto the APC, whereas another protein, called CD4 (complementarity determinant number 4), recognizes the MHC-II protein.
- There are other recognition proteins that are not presented here. The recognition described above allows the T_H lymphocyte to alert the rest of the immune system about the presence of a pathogen inside the body, which leads to the amounting of a full-scale immune response against it.
- Viral proteins co-evolved with cellular surface proteins, facilitating cell entry and infection

<u>Example – SARS-CoV2</u>:

The spike protein binds to two surface receptors:

1. ACE2 – allows attachment to the cell
2. TMPRSS2 – cleaves parts of the spike protein, allowing membrane fusion



Protein Roles: Building Cells and Tissues

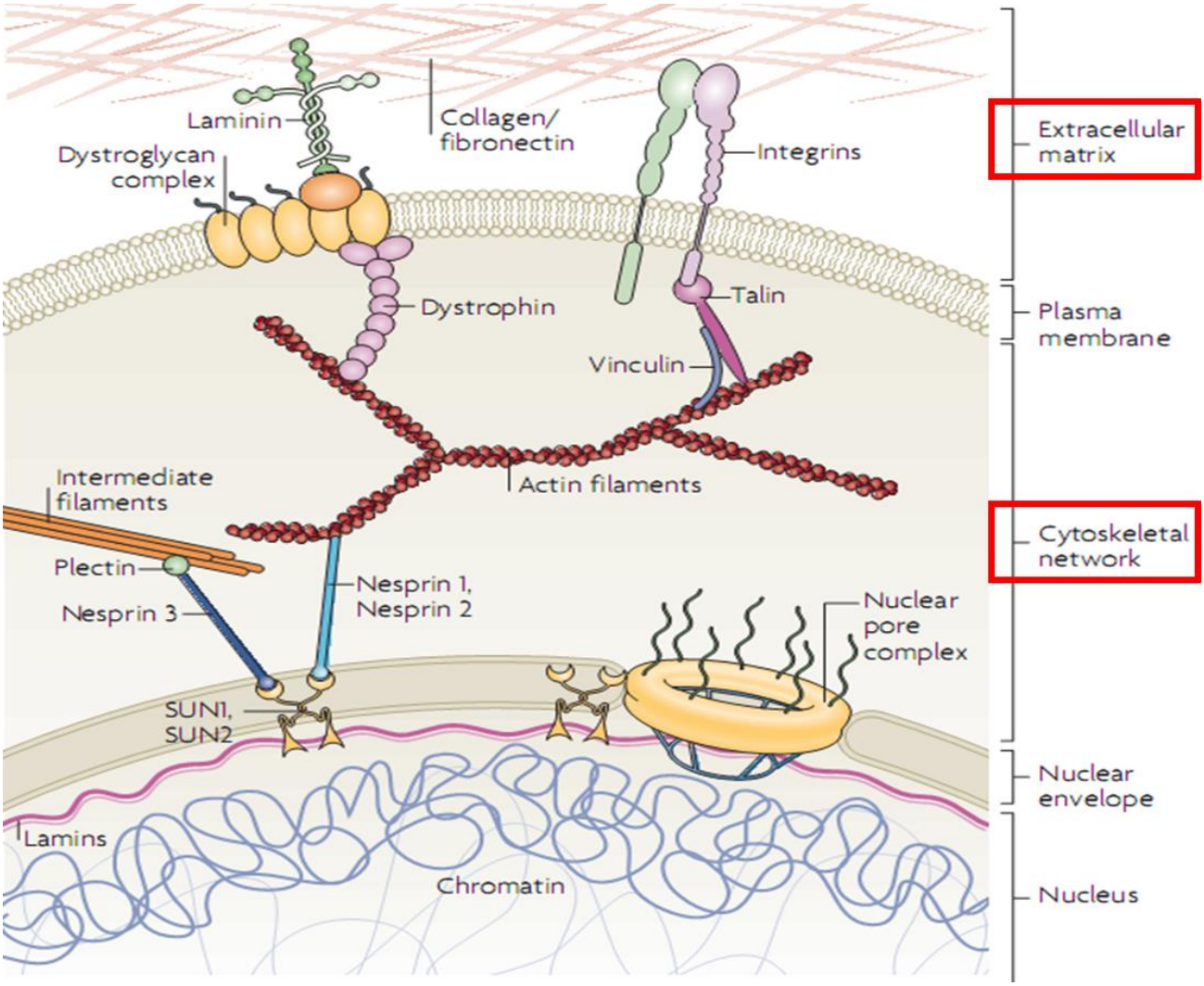


Figure Connectivity of the extracellular matrix and intracellular elements.

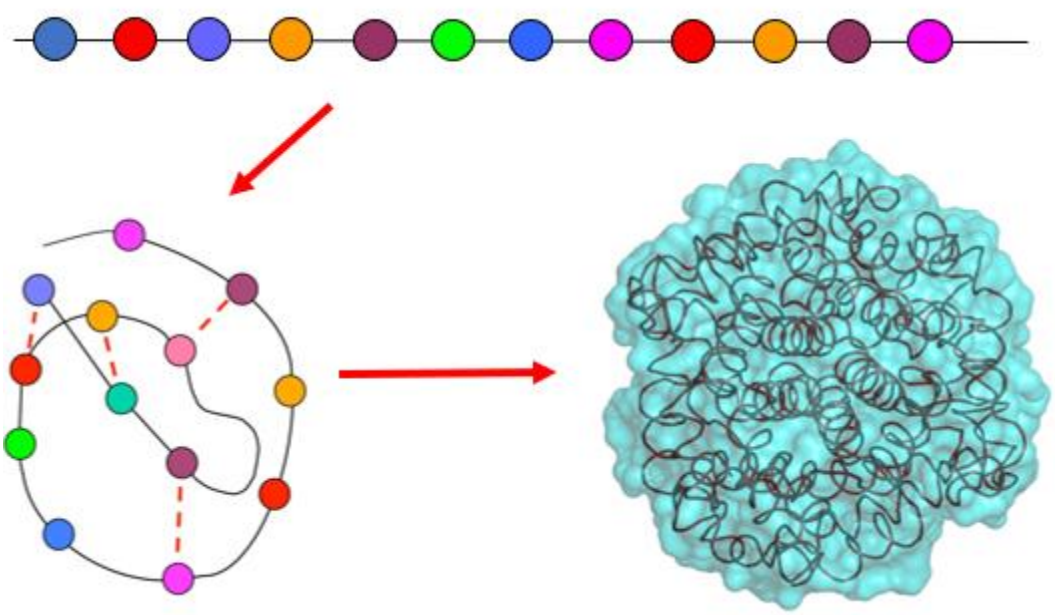
The figure shows the continuous protein network starting from collagen and fibronectin fibers outside the cell, to membrane-crossing elements such as integrins, to intracellular elements such as actin- and intermediate-filaments of the cytoskeleton, and finally to elements inside the cell nucleus, such as lamins.

Proteins' functionality results from their chemical and structural complexity

Each protein is built from a chain of amino acids that folds into a 3D structure due to attraction and repulsion forces

- How can proteins, which are made from the same material, have so many different functions?

- The answer has to do with the structural and chemical diversity of proteins.
- Proteins are made of chains of amino acids. There are 20 types of amino acids, and each protein chain is a combination of these types.
- The chemical properties of the amino acids makes some attract to each other whereas others are repelled from each other.
- This drives the chain to fold into a 3D structure that maximizes attraction and minimizes repulsion.
- Since each protein has a different sequence of amino acids, each folds into a different structure.
- What does that have to do with functional diversity?

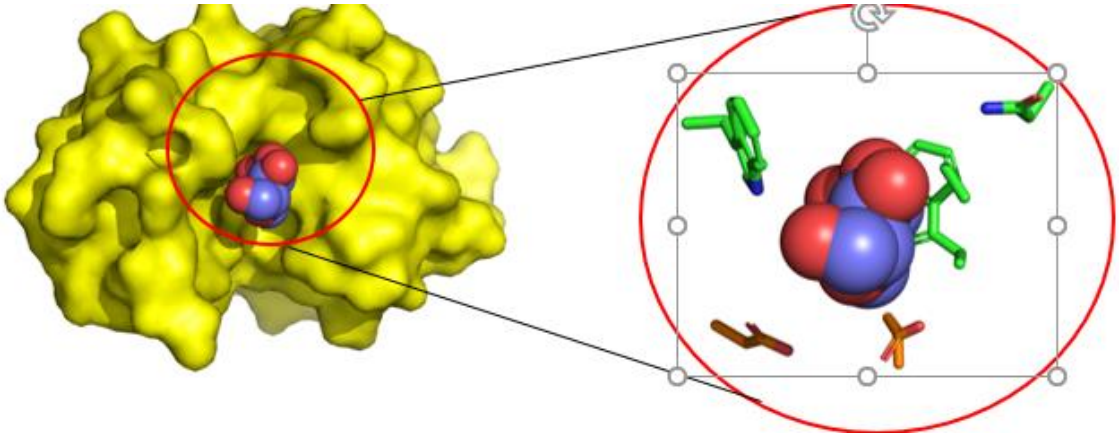


Proteins' functionality results from their chemical and structural complexity

The action of all proteins involves ligand binding:

- Enzyme-substrate
- Receptor-hormone/neurotransmitter
- G-protein-GTP
- Histone-DNA
- Hemoglobin-O_2

- Channel/transporter-ion/solute
- Cytoskeleton proteins
- Extracellular proteins
- *The action of virtually all proteins involves binding of a ligand (small molecule, another macromolecule).*
-



- Fig .Structure-function relationship in the binding site of an enzyme (lysozyme). Left: The overall structure of an enzyme (PDB entry 9lyz). The contours of the enzyme are shown as a yellow surface. The substrate (blue spheres) is shown within the binding site (red circle). Right: A magnification of the binding site, showing the substrate and some of the enzyme's amino acids that participate in the binding (green sticks) and catalysis (orange sticks) of the substrate.

Proteins' functionality results from their chemical and structural complexity

Each protein:

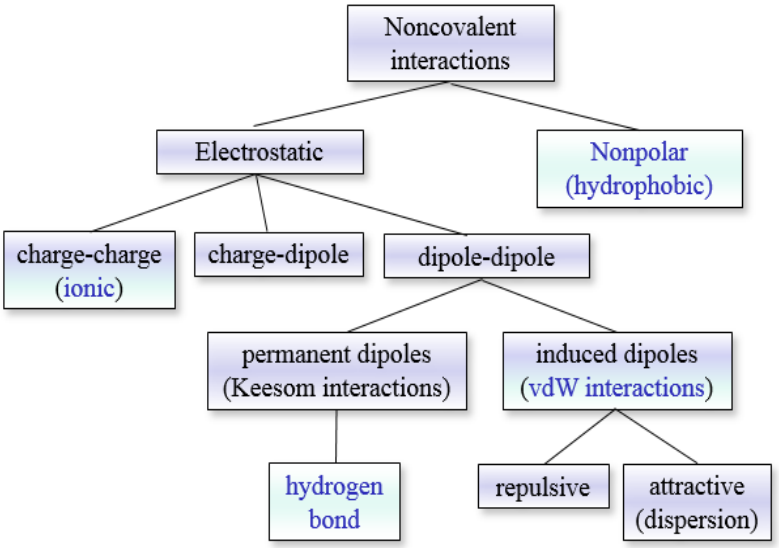
Different amino acid sequence → different 3D structure → different binding/catalytic sites

Non-covalent interactions

Non-covalent interactions:

- Drive the folding of the protein chain into a specific 3D structure
- Drive the binding of ligands (substrate, hormone, messenger, allosteric modulator) to a binding/catalytic site in the proteins

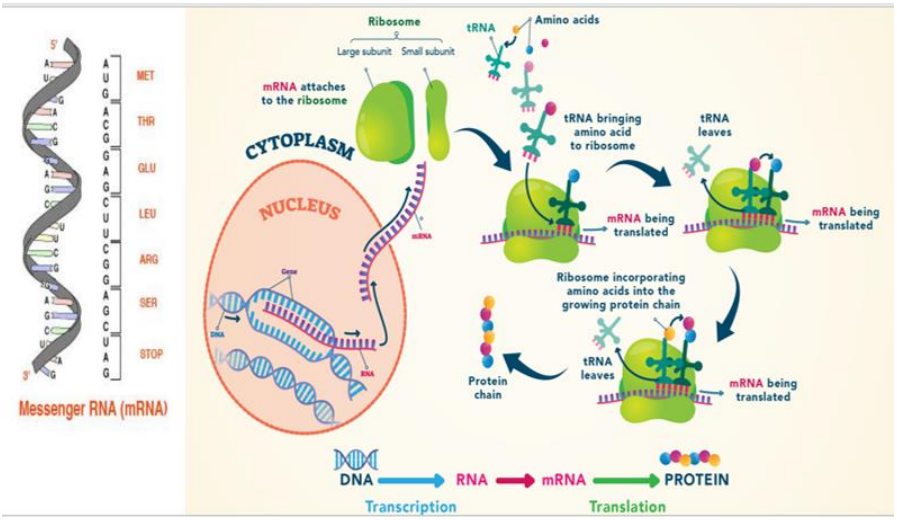
- Much weaker than covalent bonds
- Make biomolecules flexible, and allow them to interact reversibly with each other
- Non-covalent interactions

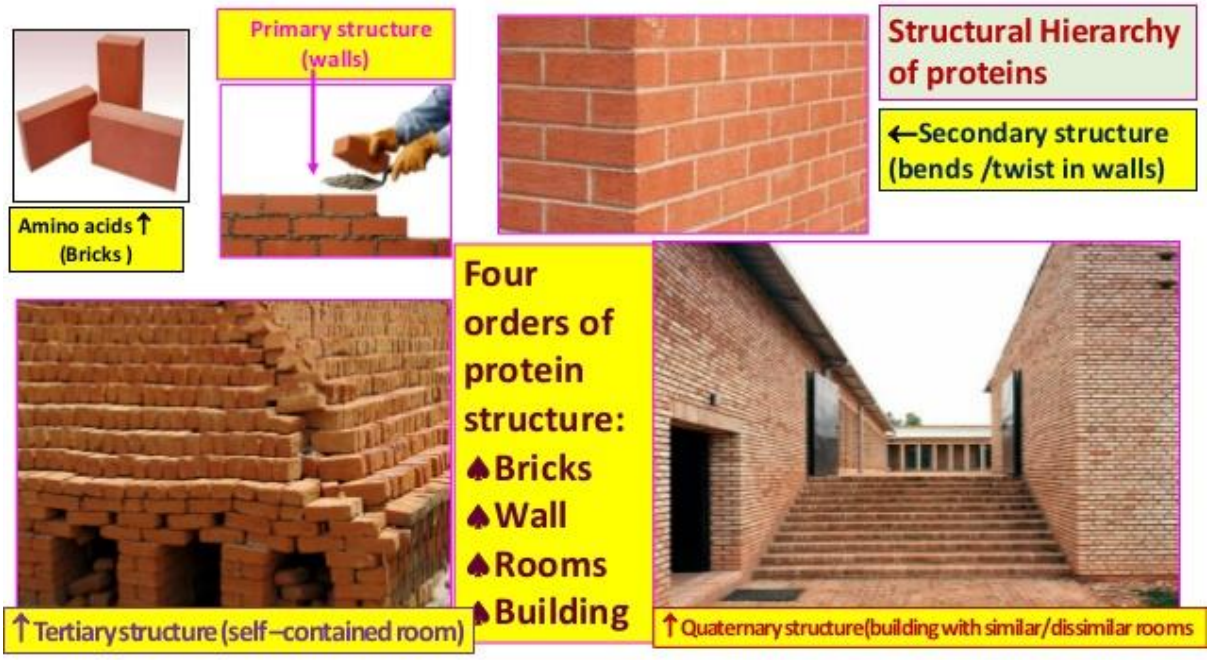


- Figure. Main non-covalent interactions found in proteins
- The relationship between the interaction types
- The dominant interactions in proteins are colored in blue. vdW: van der Waals.

- Proteins
- Proteins are the workhorses of the cell.
- Virtually everything that goes on inside of cells happens as a result of the actions of proteins.
- Among other things, protein enzymes catalyze the vast majority of cellular reactions, mediate signaling, give structure both to cells and to multicellular organisms, and exert control over the expression of genes.
- Proteins are made by linking together amino acids, with each protein having a characteristic and unique amino acid sequence.
- To get a sense for the diversity of proteins that can be made using 20 different amino acids, consider that the number of different combinations possible with 20 amino acids is 20n, where n=the number of amino acids in the chain. It becomes apparent that even a dipeptide made of just two amino acids joined together gives us 202 = 400 different combinations.

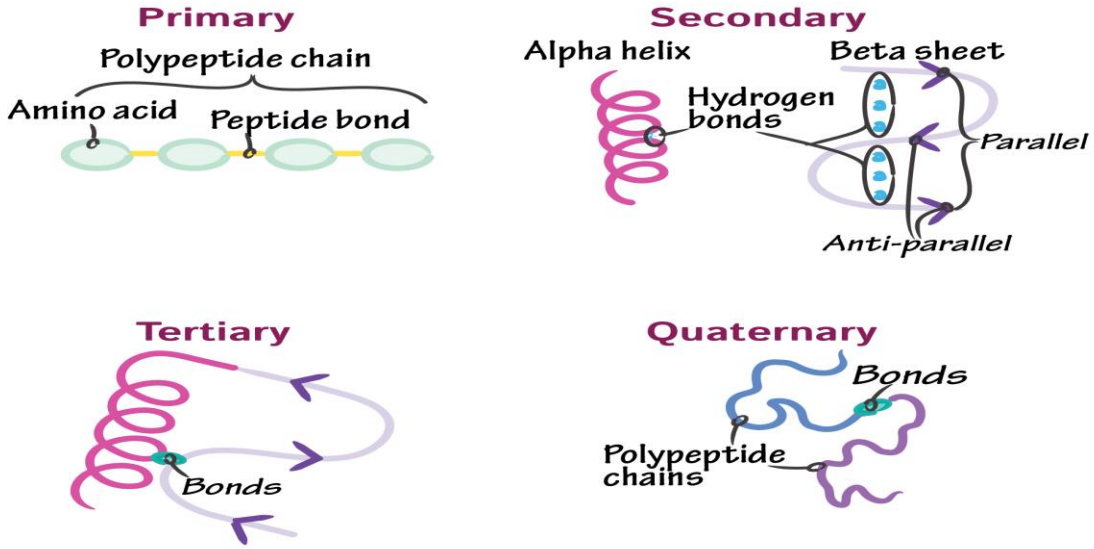
- If we do the calculation for a short peptide of 10 amino acids, we arrive at an enormous 10,240,000,000,000 combinations.
- Most proteins are much larger than this, making the possible number of proteins with unique amino acid sequences unimaginably huge.

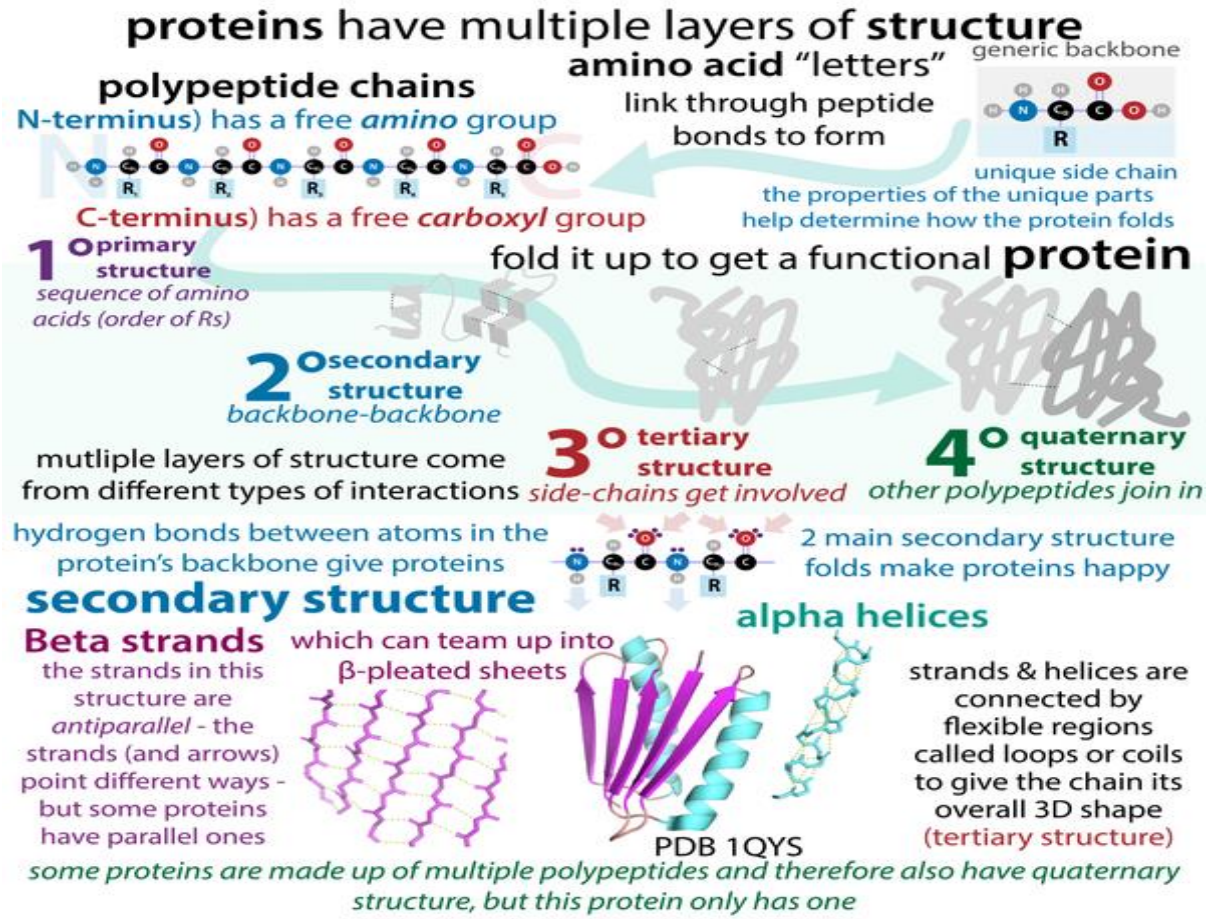




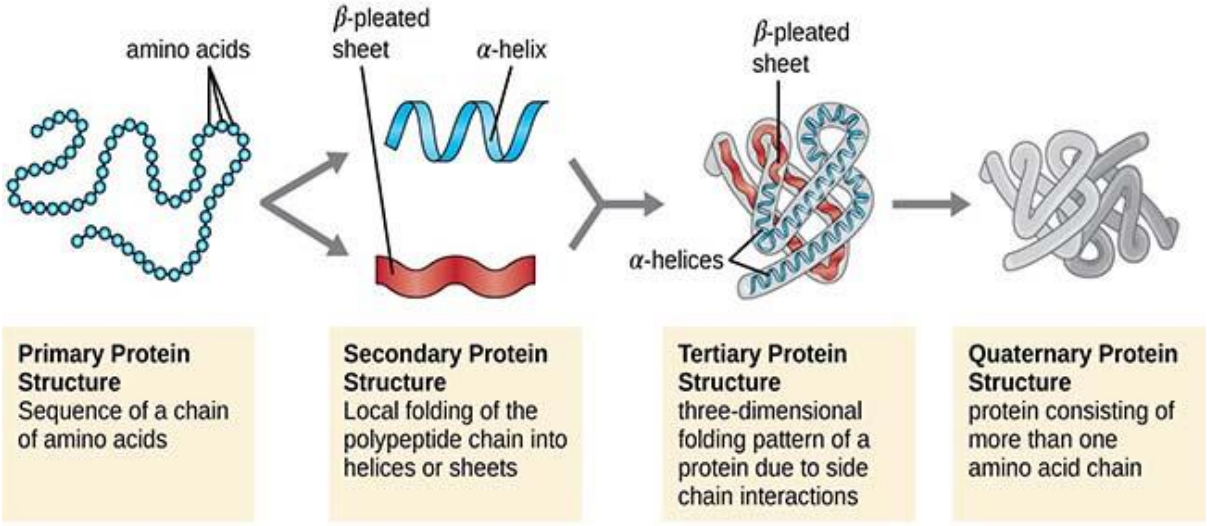
Structure of proteins: primary, secondary, tertiary and quaternary, biological significance

CLASSES OF PROTEIN STRUCTURE

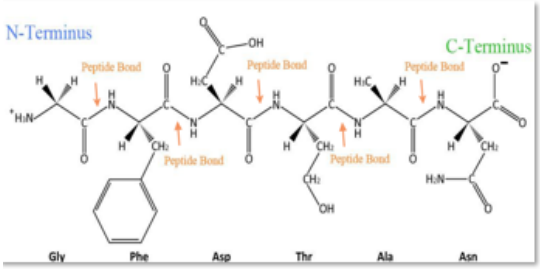


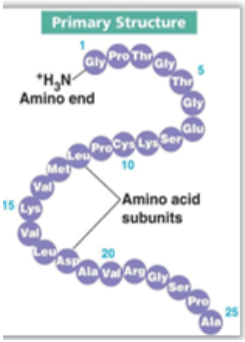


Structure of Proteins



Primary structure of proteins



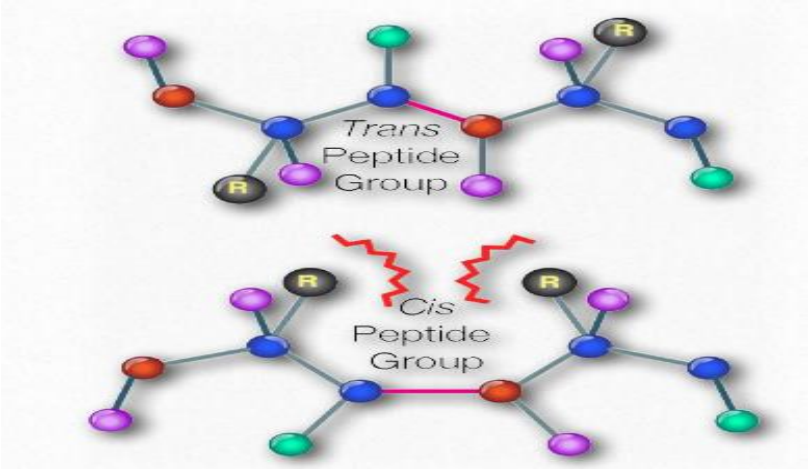


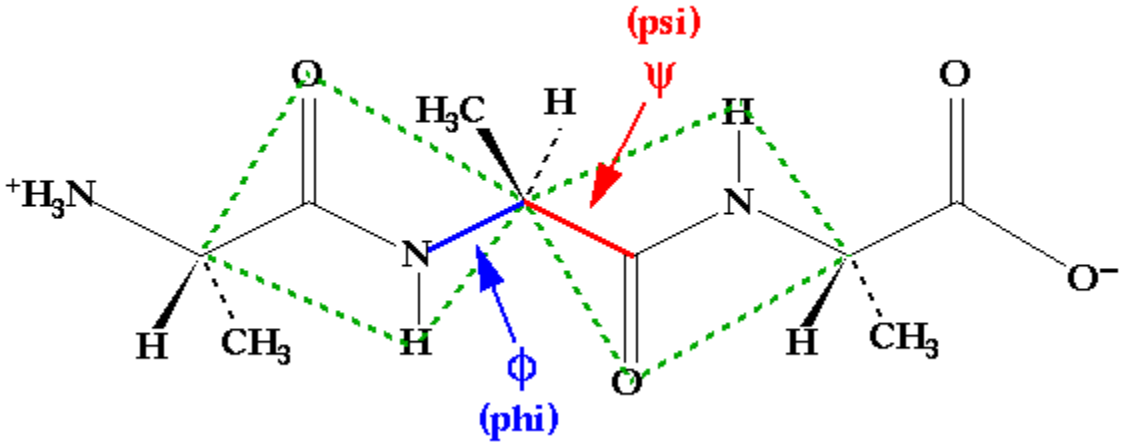
The "R" groups come from the 20 amino acids which occur in proteins. The peptide chain is known as the **backbone**, and the "R" groups are known as **side chains**.

- The primary structure of proteins refers to the specific sequence of amino acids that make up that protein.
- Every protein contains its own unique sequence of amino acids that determines the three-dimensional structure of that protein.
- The linear polymer of amino acids, which are held together by peptide bonds, has polarity. This is because one end of the polypeptide chain contains a full positive charge while the other end contains a full negative charge.
- By convention, the beginning of the polypeptide chain is always at the positively-charged alpha amino group while then end is at the negatively-charged alpha carboxyl group.
- Each amino acid in the polypeptide chain has the ability to donate a hydrogen atom to form a hydrogen bond via the N-H group and accept a hydrogen atom to form a hydrogen bond via the C=O group. This will play an important role in determining the secondary structure of proteins.

- The peptide bond holding each adjacent pair of amino acids is resonance stabilized, which means that it has a double bond character. Therefore the peptide bond is planar and does not rotate in space.

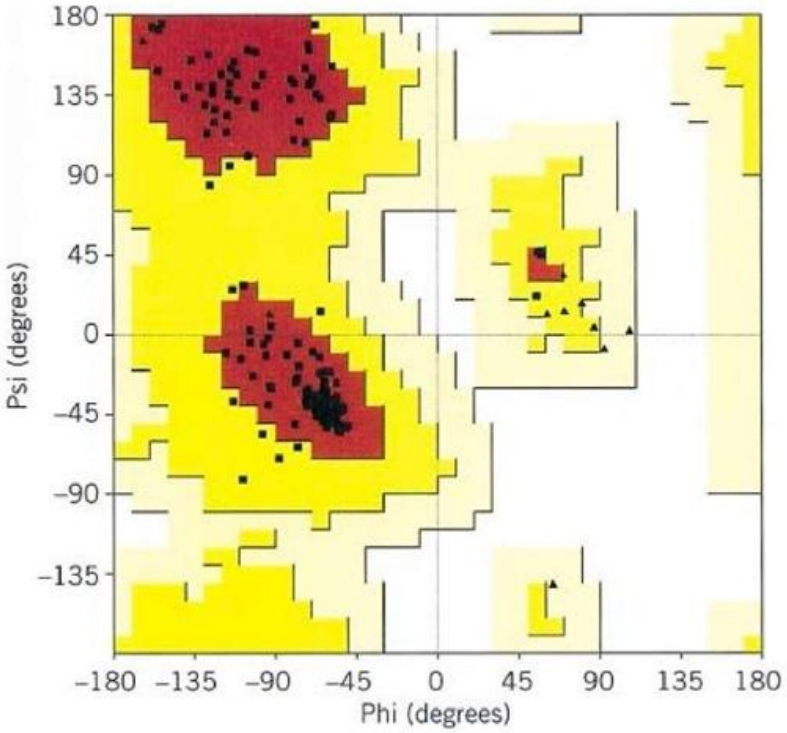
- The cis configuration of the peptide bond is typically more stable than the trans peptide because of steric hinderance.

- 



Each amino acid contains two bonds that can readily rotate - this includes the phi angle and the psi angle.

- The phi angle is the angle between the alpha carbon atom and the nitrogen while the psi angle is the angle between the alpha carbon and the carbon of the carbonyl group.
- These angles, known as the torsion angles, are responsible for rotating the entire linear polymer and ultimately transforming the linear polymer into a three-dimensional molecule.
- Two torsion angles in the polypeptide chain, also called Ramachandran angles (after the Indian physicist who worked on modeling the interactions in polypeptide chains, Ramachandran, GN,

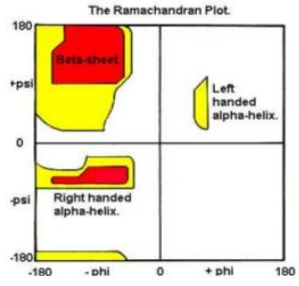


- The Ramachandran plot provides a convenient way to view the distribution of torsion angles in a protein structure.
- It also provides an overview of excluded regions that show which rotations of the polypeptide are not allowed due to steric hindrance (collisions between atoms).
- The Ramachandran plot of a particular protein may also serve as an important indicator of the quality of its three-dimensional structures

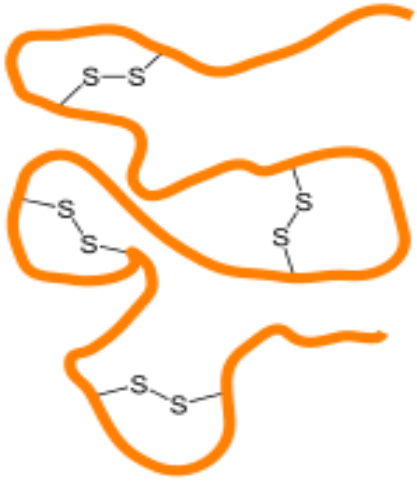
The Ramachandran Plot

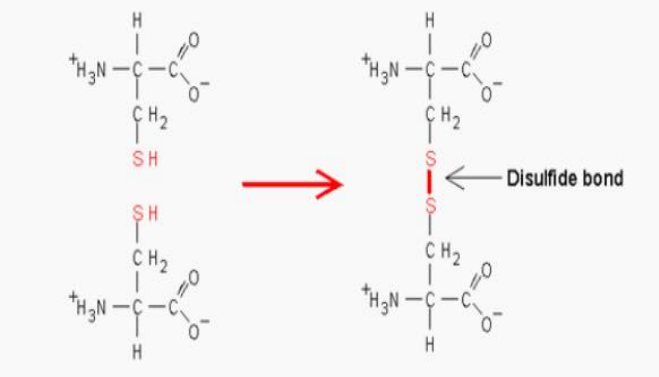
The red regions correspond to conformations where there are no steric clashes, ie these are the allowed regions namely the alpha-helical and beta-sheet conformations

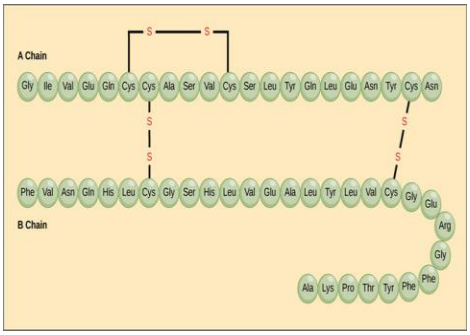
- **The yellow areas show the allowed regions if slightly shorter van der Waals radi are used in the calculation, ie the atoms are allowed to come a little closer together**
- **This brings out an additional region which corresponds to the left-handed alpha-helix**



- If two cysteine side chains end up next to each other because of folding in the peptide chain, they can react to form a sulphur bridge. This is another covalent link and so some people count it as a part of the primary structure of the protein.
- Because of the way sulphur bridges affect the way the protein folds, other people count this as a part of the tertiary structure

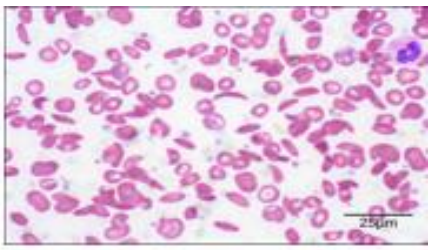




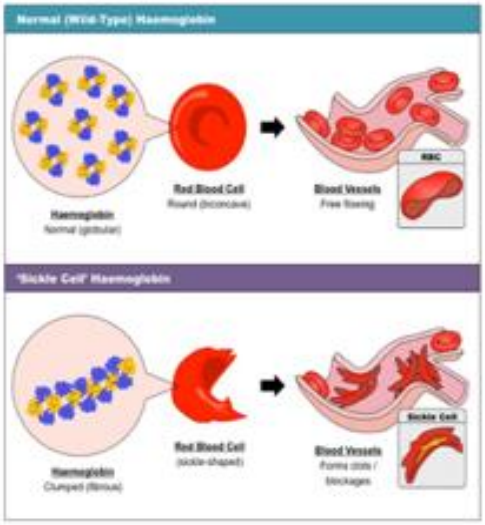


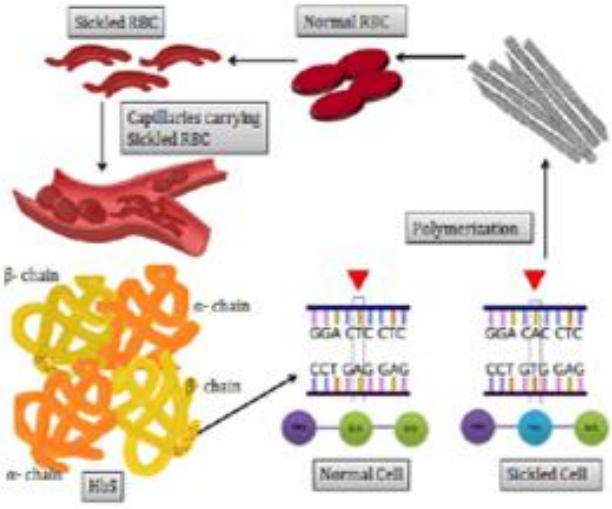
The oxygen-transport protein hemoglobin consists of four polypeptide chains, two identical α chains and two identical β chains. In sickle cell anemia, a single amino substitution in the hemoglobin β chain causes a change the structure of the entire protein. When the amino acid glutamic acid is replaced by valine in the β chain, the polypeptide folds into a slightly-different shape that creates a dysfunctional hemoglobin protein. So, just one amino acid substitution can cause dramatic changes. These dysfunctional hemoglobin proteins, under low-oxygen conditions, start associating with one another, forming long fibers made from millions of aggregated hemoglobins that distort the red blood cells into crescent or “sickle” shapes, which clog arteries. People affected by the disease often experience breathlessness, dizziness, headaches, and abdominal pain.

The A chain of insulin is 21 amino acids long and the B chain is 30 amino acids long, and each sequence is unique to the insulin protein.



Sickle cells are crescent shaped, while normal cells are disc-shaped.





Characteristics of Primary structure of proteins:1

1. A peptide contains two or more amino acid residues joined together by a peptide bonds. Individual amino acids can be considered as bricks.
2. **Formation of peptide bond** = a covalent bond is formed by amide linkage between the α-amino group of one amino acid and α- carboxyl group of another amino acid by removal of a water molecule.
3. **Characteristics of a Peptide bond:**

a. Rigid, covalent, stable, strong and can be hydrolyzed by the action of proteolytic enzymes, acids and alkalis

b. Planer and with partial double bond (no freedom of rotation) in character

c. –C=O, NH– Exist in trans-configuration. Both groups are polar and involved in hydrogen bonds.

d. Impart stability to the primary structure of proteins(disulphide bonds are also responsible for the stability)

e. The side chain are free to rotate on either side of the peptide bond.

f. Distance between amino acids is 1.32 A° which is midway between that of single bond (1.49 A°) and double bond (1.27A°)

g. **Ramachandran angles**: are angle of rotation → determine the spatial orientation of peptide chain

Characteristics of Primary structure of proteins:2

5. In polypeptide chain , at one end there will be one free alpha amino group → N-terminal end and protein biosynthesis starts from amino terminal end (the first amino acid).
6. The other end of polypeptide chain , **is carboxy** terminal end (the last amino acid) where there is free alpha **carboxy** group.
7. All other alpha-amino acids and alpha-**carboxy** groups are involved in peptide bond formation.
8. **Writing of peptide structures** –by convention ,the amino acid sequence is written from left to right with the free amino end (N- terminal acid/residue→ number 1 by tradition) on left and ending with the free carboxyl end (C –terminal amino acid / residue).
9. **Shorthand to read peptides : three** letter or **one** letter abbreviation/ short hand form of amino acids in protein to be read from N-terminal residue on left of peptide e.g. (Gly-Ala–Val)→Glycyl –Alanyl –Valine → G A V

Or NH_2-Gly-Ala-Val-COOH

10. Naming peptides : for naming peptides the amino acid Suffixes **ine** (Glycine → to glycyl) , **an** (Tryptophan), **ate** (Glutamate→ Glutamyl) by **yl** with the exception of C-terminal amino acid.

Use of symbols in representing a peptide

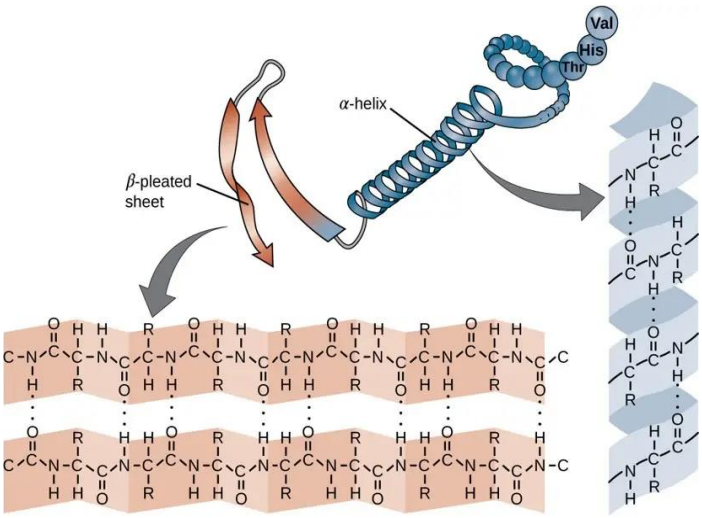
- A tripeptide → 3 amino acids and 2 peptide bonds is shown.

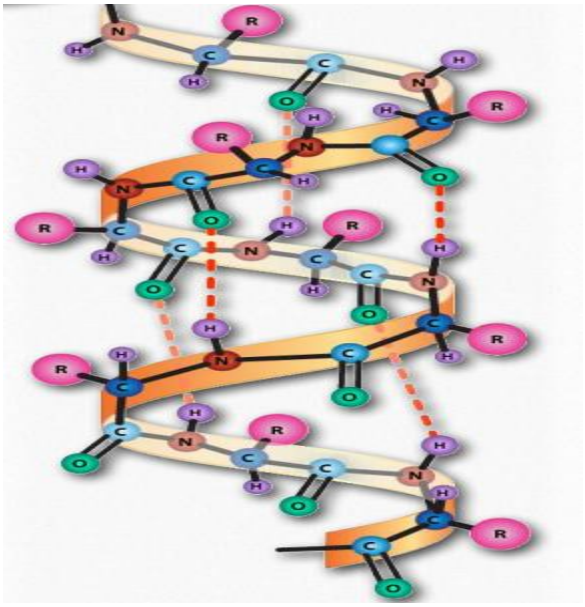
H_3N^+ Glutamate – Cysteine –Glycine $-COO^-$	← Amino acids in a peptide
E – C – G	←one letter abbreviation
Glu – Cys – Gly	←three letter abbreviation
Glutamyl – cysteinyl –Glycine	← Peptide name

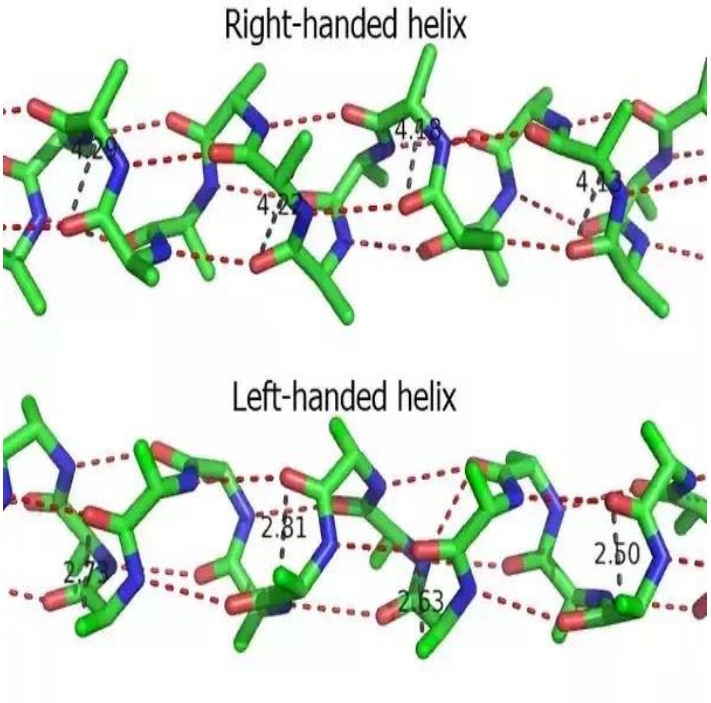
- Free amino end (N- terminal acid/residue) →$-N^+H_3$ is on the left. Free carboxyl end (C –terminal amino acid / residue)→ $-COO^-$ is on right .
- The amino acid sequence is written and read from left to right. This is the chemical shorthand to write proteins.

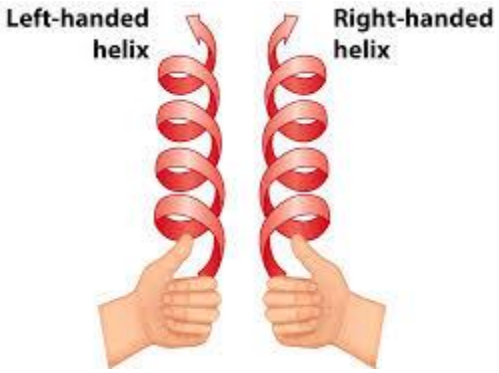
Secondary Structural Features in Protein Structure

- As protein synthesis progresses, interactions between amino acids close to each other begin to occur, giving rise to local patterns called secondary structure.
- These secondary structures include the well known α- helix and β-strands.
- Both were predicted by Linus Pauling, Robert Corey, and Herman Branson in 1951.
- Each structure has unique features.

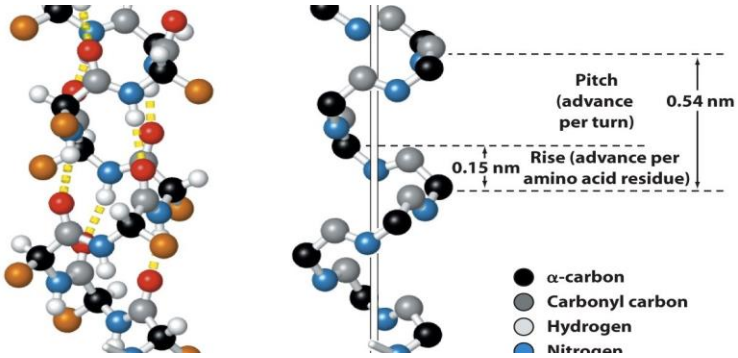

Val
His
Thr
α-helix
β-pleated
sheet

- 

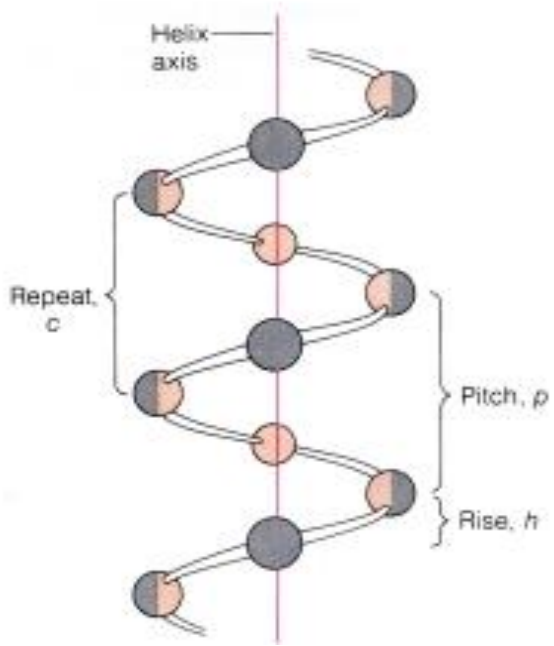
- 



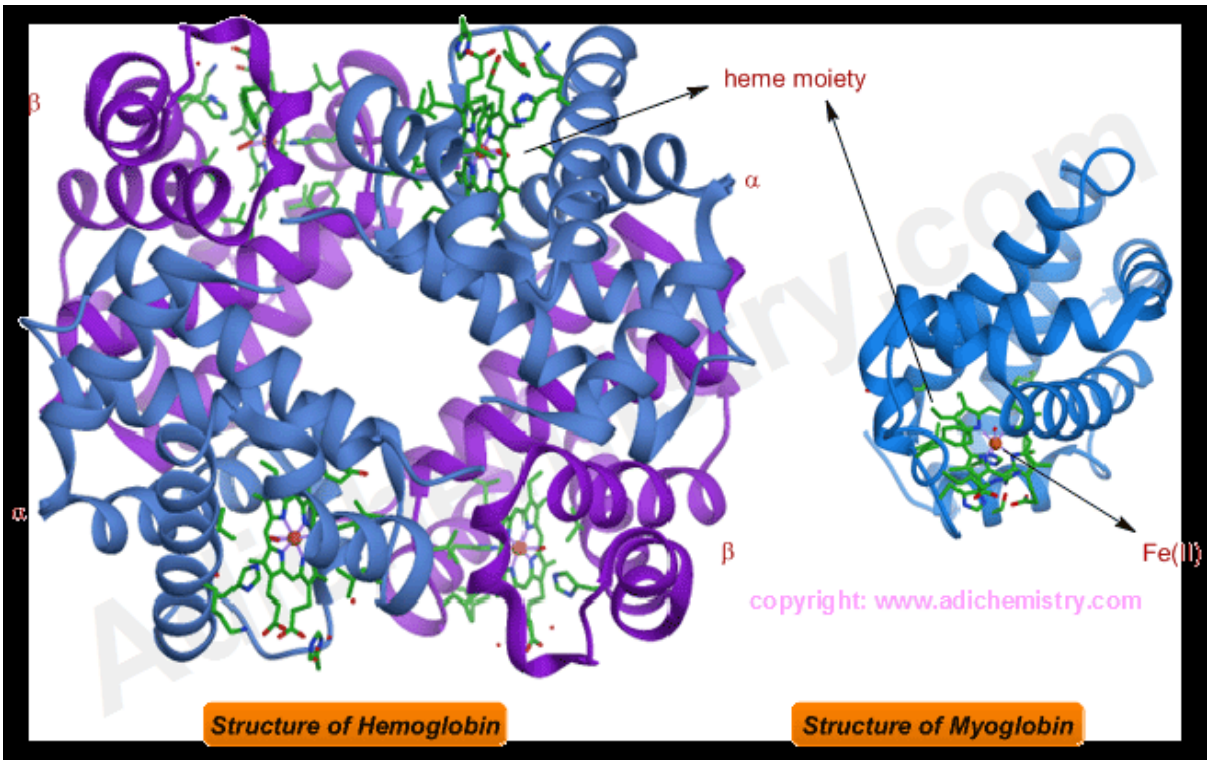
- The α-helix has a coiled structure, with 3.6 amino acids per turn of the helix (5 helical turns = 18 amino acids).
- Helices are predominantly right handed - only in rare cases, such as in sequences with many glycines can left handed α- helices form.
- In the α-helix, hydrogen bonds form between C=O groups and N-H groups in the polypeptide backbone that are four amino acids distant.
- These hydrogen bonds are the primary forces stabilizing the α-helix.



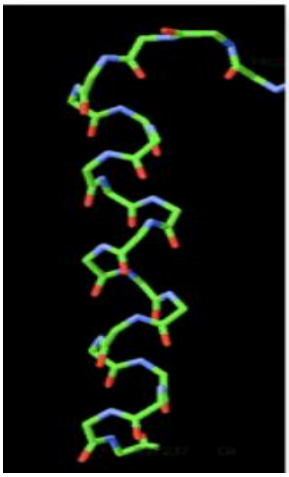
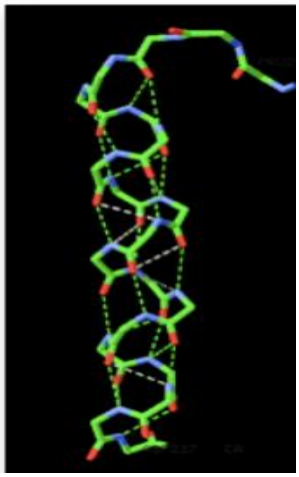
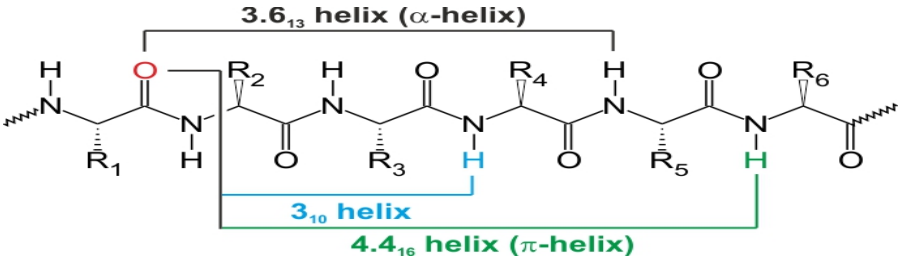
- We use the terms rise, repeat, and pitch to describe the parameters of any helix.
- The repeat is the number of residues in a helix before it begins to repeat itself. For an α-helix, the repeat is 3.6 amino acids per turn of the helix.
- The rise is the distance the helix elevates with addition of each residue. For an α-helix, this is 0.15 nm per amino acid. The pitch is the distance between complete turns of the helix. For an α-helix, this is 0.54 nm.

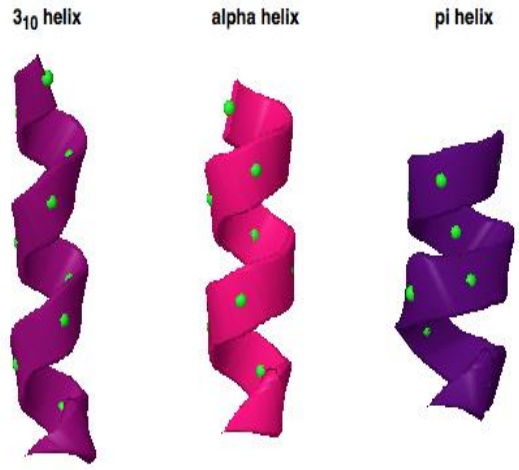


- The most common type of secondary structure in proteins is the α-helix.
- Linus Pauling was the first to predict the existence of α-helices.
- The prediction was confirmed when the first three-dimensional structure of a protein, myoglobin (by Max Perutz and John Kendrew) was determined by X-ray crystallography.

- 

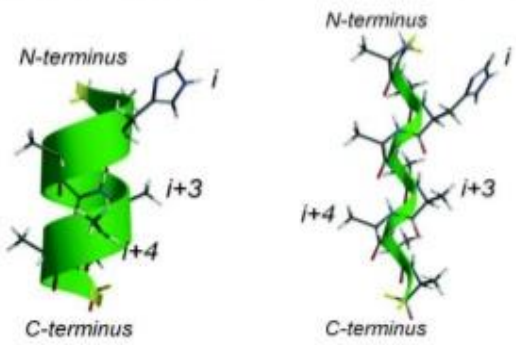
  The stability of an α-helix is enhanced by the presence of the amino acid aspartate-justify
- The repeating structural pattern in helices is a result of repeating φ values and ψ values, which is reflected in the clustering of the torsion angles within the helical region of the Ramachandran plot.
- When looking at the helix in the figure below, notice how the carbonyl (C=O) oxygen atoms (shown in red) point in one direction, towards the amide NH groups 4 residues away (i, i+4).
- Together these groups form a hydrogen bond, one of the main forces in the stabilization of secondary structure in proteins. The hydrogen bonds are shown on the right figure as dashed lines.

-  
- The α-helix is not the only helical structure in proteins.
- 3_10 helix, which is stabilized by hydrogen bonds of the type (i, i+3)
- π-helix, which is stabilized by hydrogen bonds of the type (i, i+5).
- The 3_10 helix has a smaller radius, compared to the α-helix, while the π-helix has a larger radius.
- The first detailed analysis of the occurrence of the π-helix in proteins, which was based on the analysis of entries in the Protein Data Bank (PDB), was published by Fodje & Al-Karadaghi, 2002.
- 

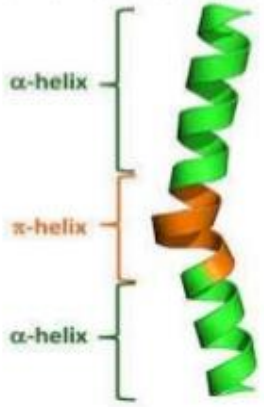


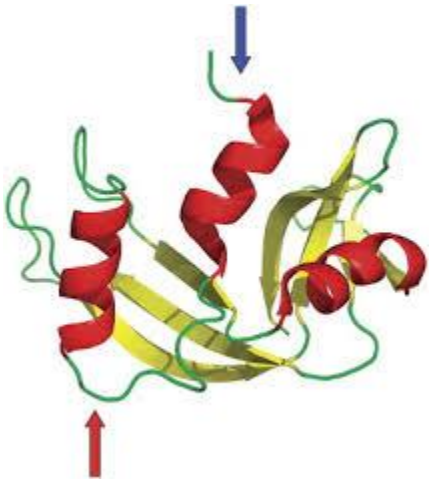
Type	Angle			Aminoacid residues involved in the ring	Hydrogen bonds
	ϕ	ψ	ω		
α	-57	-47	180	3.6	i+4
3_{10}	-60	-30	180	3.0	i+3
π	-57	-70	180	4.4	i+5

- The residues of the component amino acids stick out at right angle from the helix, thus minimizing the steric interactions and stabilizing the structure.
- Amino acids that prefer to adopt α-helical conformations in proteins include methionine, alanine, leucine, glutamate and lysine, as their small side chains minimize the steric hindrance in helix formation.
- Other less common helices found in proteins are 3(10)-helix which is more stretched than ideal α – helix and π – helix which is more compact and extremely rare.



***α*-helix** **3_{10}-helix**



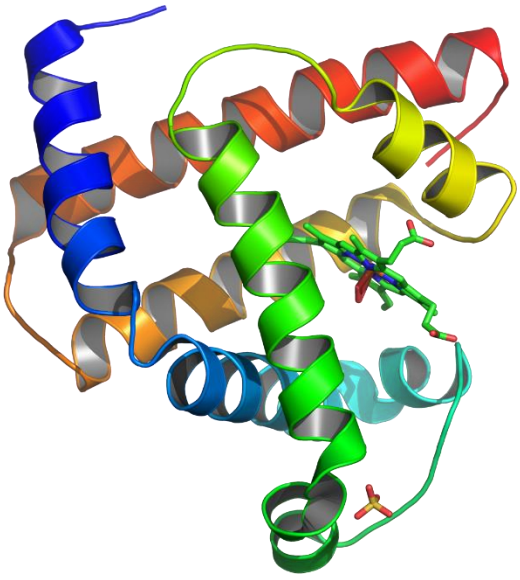


3_{10} helix

- A 3_{10} helix is a type of secondary structure
- found (often) in proteins and polypeptides
- Top view of the same helix shown to the right
- Three carbonyl groups are pointing upwards towards the viewer
  - spaced roughly 120° apart on the circle
  - corresponding to 3.0 amino-acid residues per turn of the helix

voltage sensor segment of voltage-gated potassium channels

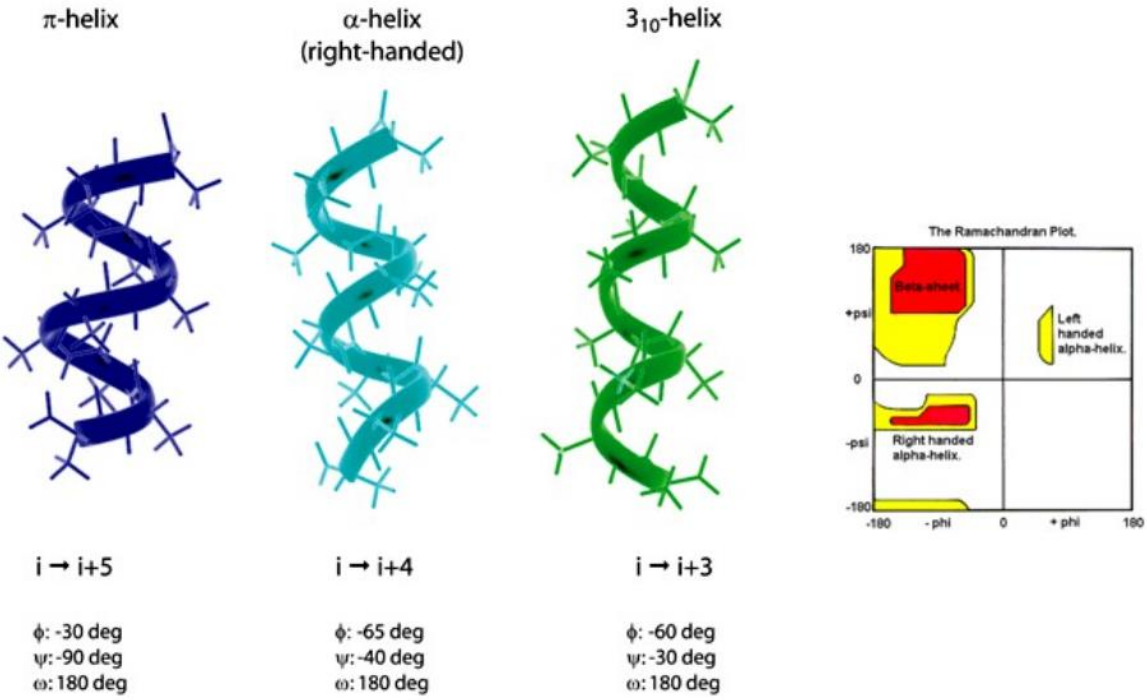
- Soluble methane monooxygenase is the current record holder for the greatest number of π-helices in a single enzyme with 13 (PDB code 1MTY)



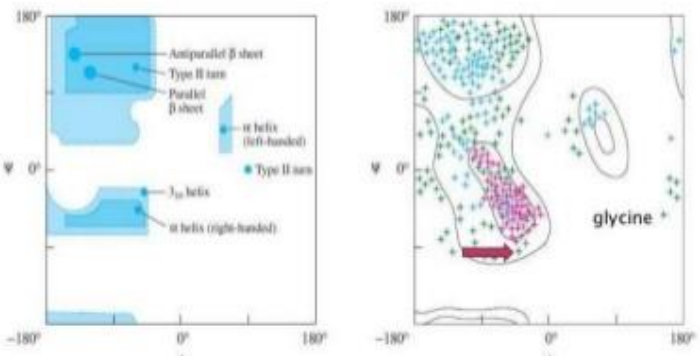
- A pi helix (or π-helix) is a type of secondary structure found in proteins.

- Discovered by crystallographer Barbara Low in 1952 and once thought to be rare, short π-helices are found in 15% of known protein structures and are believed to be an evolutionary adaptation derived by the insertion of a single amino acid into an α-helix.
- Because such insertions are highly destabilizing, the formation of π-helices would tend to be selected against unless it provided some functional advantage to the protein.
- π-helices therefore are typically found near functional sites of proteins.

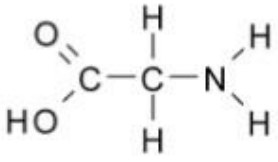
Types of helix	No. of residue per turn	Pitch	Radius of helix	Rise per residue
α helix	3.6	5.4Å	2.3Å	1.5Å
3_{10} helix	3	5.8Å	1.9Å	2.0Å
π helix	4.4	5.2Å	4.8Å	1.1Å



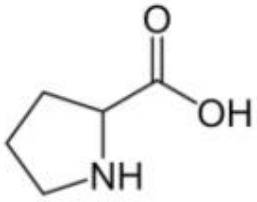
- Glycine and Proline act as α helix breakers:
  - ❑Glycine, with no side chain, has an unconstrained rotation around both Φ and Ψ and thus may have angles in the "Disallowed" regions, which hinder in the formation of hydrogen bonds in the α – helix, thus acting as a helix breaker.
  - ❑Proline side chain is jammed into the space that should be occupied by the backbone of the α – helix, a methylene group is in the space that would normally be occupied by a hydrogen bonding amide proton, thus disrupting the H – bond network and sterics of the helix, leading to helix break.



Glycine Ramachandran plot



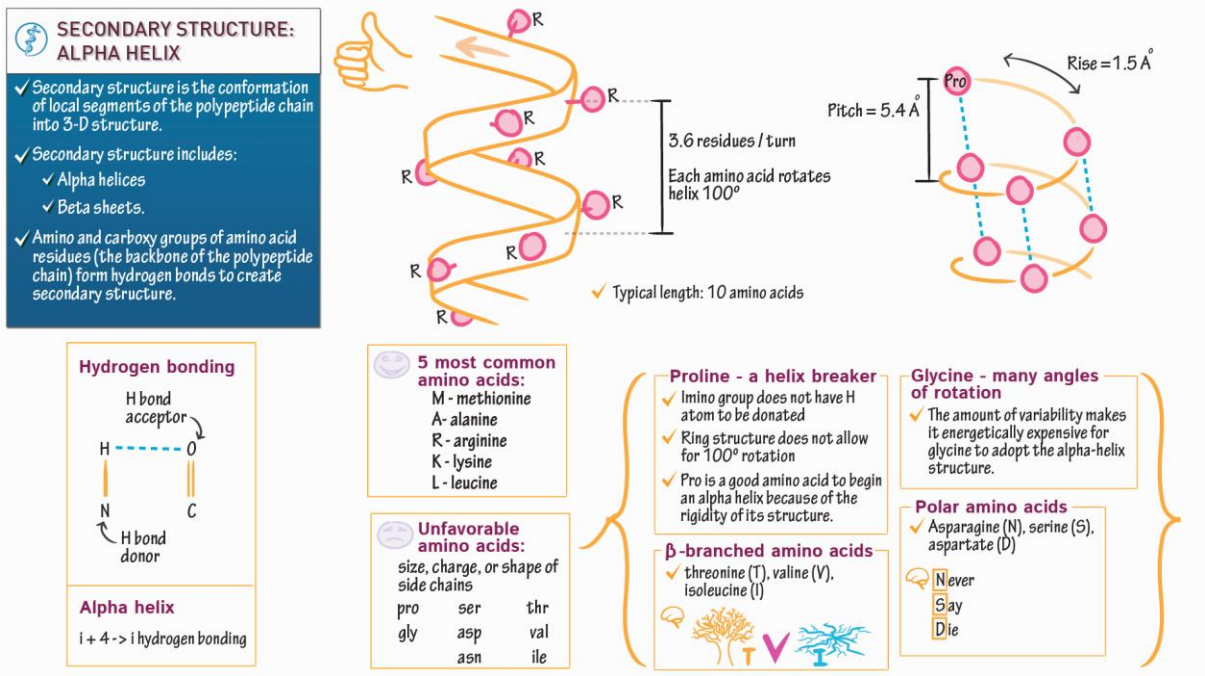
Glycine



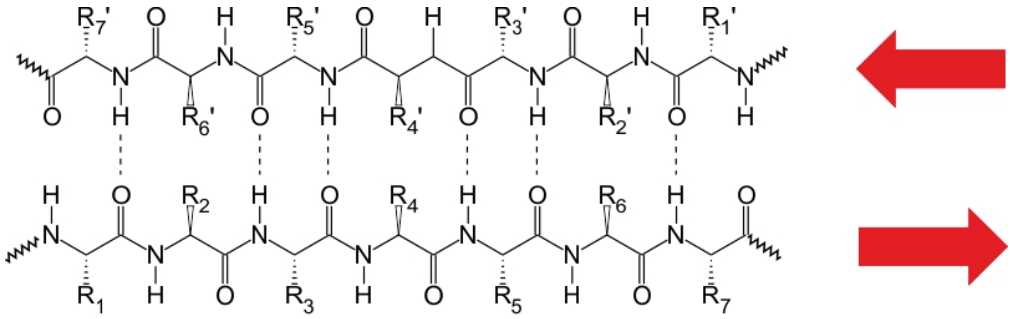
Proline

Amino Acid		Helix Behavior*	
A	Ala	H	(I)
C	Cys	Variable	
D	Asp	Variable	
E	Glu	H	
F	Phe	H	
G	Gly	I	(B)
H	His	H	(I)
I	Ile	H	(C)
K	Lys	Variable	
L	Leu	H	
M	Met	H	
N	Asn	C	(I)
P	Pro	B	
Q	Gln	H	(I)
R	Arg	H	(I)
S	Ser	C	(B)
T	Thr	Variable	
V	Val	Variable	
W	Trp	H	(C)
Y	Tyr	H	(C)

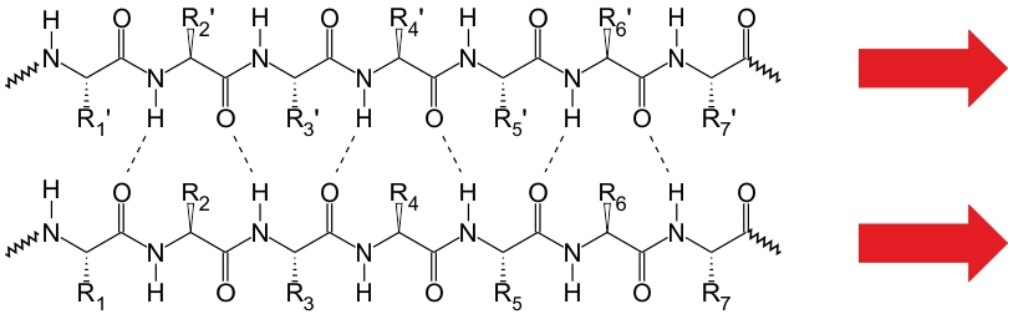
I: indifferent
C: random coil
B: helix breaker
H: helix former



β pleated sheet



antiparallel β-sheet



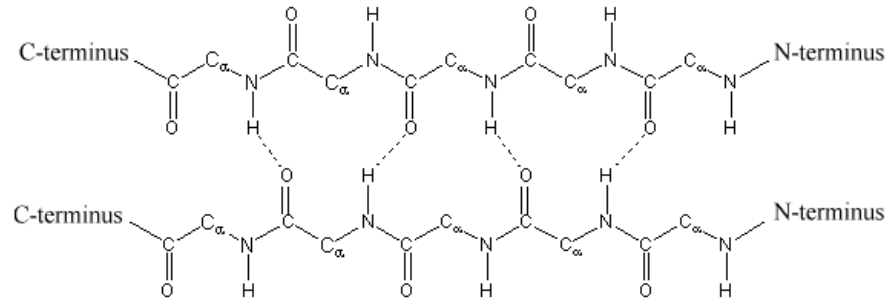
parallel β-sheet

-
-

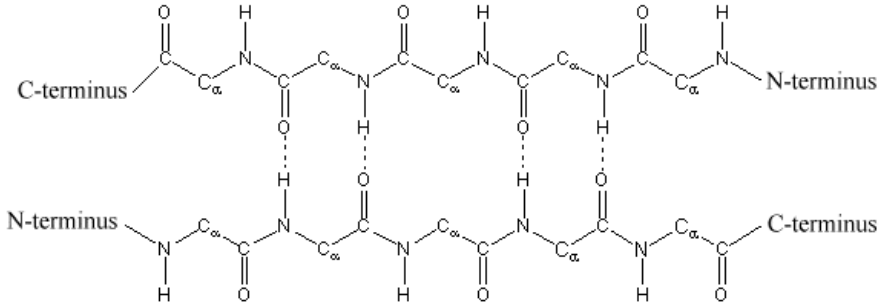
- In a β pleated sheet, two or more segments of a polypeptide chain line up next to each other, forming a sheet-like structure held together by hydrogen bonds.
- The hydrogen bonds form between carbonyl and amino groups of backbone, while the R groups extend above and below the plane of the sheet.
- The strands of a β pleated sheet may be parallel, pointing in the same direction (meaning that their N- and C-termini match up), or antiparallel, pointing in opposite directions (meaning that the N-terminus of one strand is positioned next to the C-terminus of the other).

- Parallel and anti-parallel arrangement is the direct consequence of the directionality of the polypeptide chain.
- In anti-parallel arrangement, the C-terminus end of one segment is on the same side as the N-terminus end of the other segment.
- In parallel arrangement, the C-terminus end and the N-terminus end are on the same sides for both segments.
- The "pleat" occurs because of the alternating planes of the peptide bonds between amino acids; the aligned amino and carbonyl group of each opposite segment alternate their orientation from facing towards each other to facing opposite directions.
- The parallel arrangement is less stable because the geometry of the individual amino acid molecules forces the hydrogen bonds to occur at an angle, making them longer and thus weaker.
- Contrarily, in the anti-parallel arrangement the hydrogen bonds are aligned directly opposite each other, making for stronger and more stable bonds.

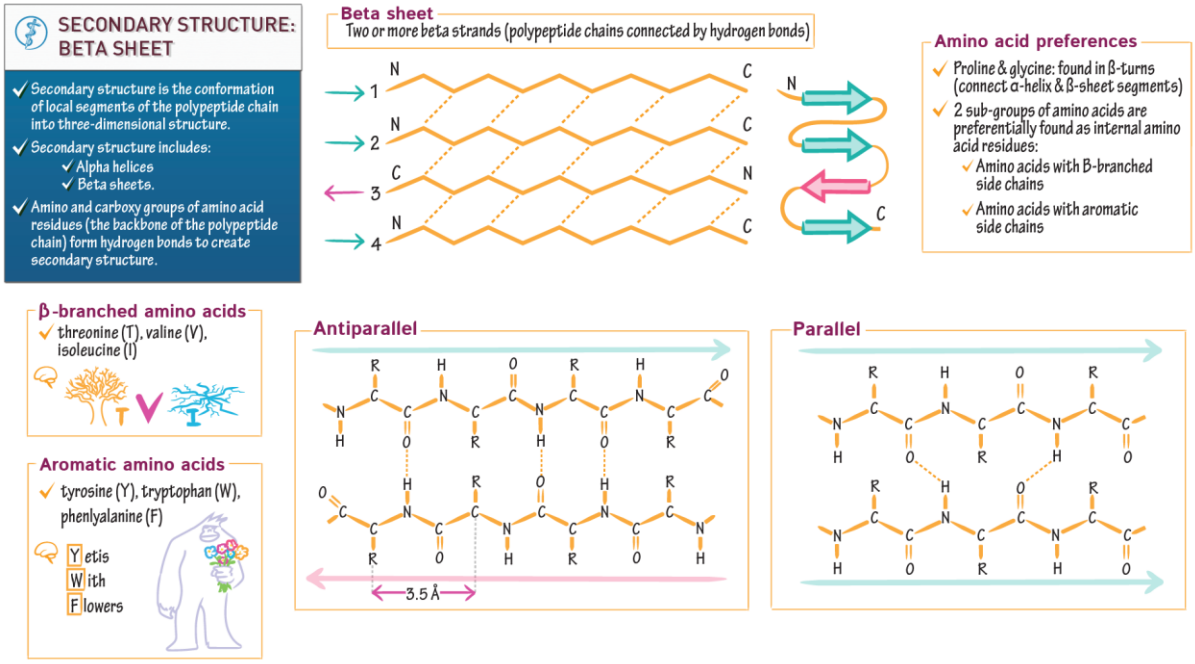
Parallel β Sheet



Antiparallel β Sheet

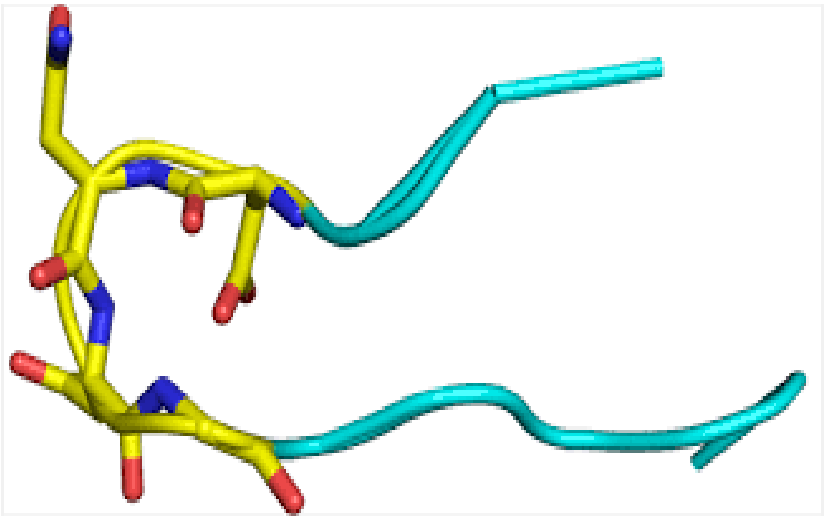
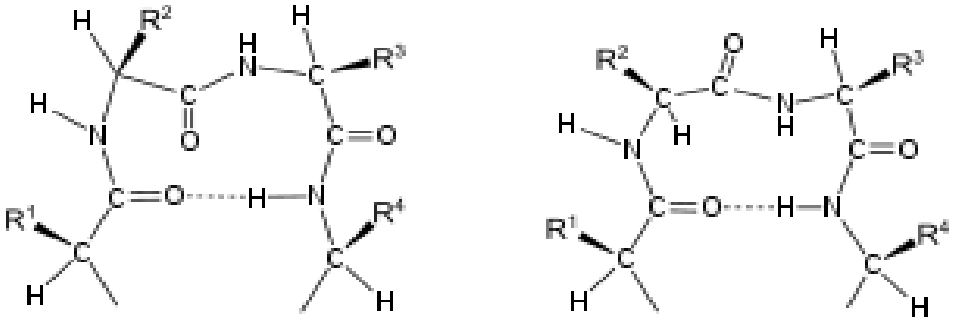


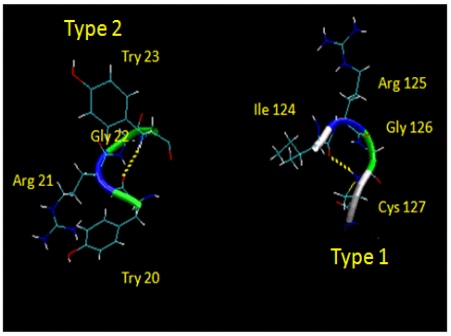
- Commonly, an anti-parallel beta-pleated sheet forms when a polypeptide chain sharply reverses direction.
- This can occur in the presence of two consecutive proline residues, which create an angled kink in the polypeptide chain and bend it back upon itself.
- This is not necessary for distant segments of a polypeptide chain to form beta-pleated sheets, but for proximal segments it is a definite requirement.



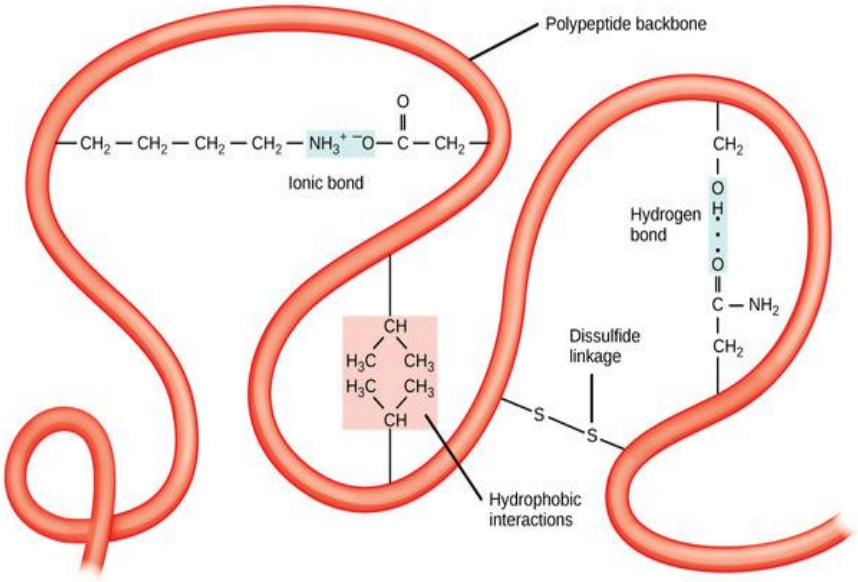
Turns

- Turns are the third of the three "classical" secondary structures with approximately one-third of all residues in globular proteins are contained in turns that serve to reverse the direction of the polypeptide chain.
- Turns are located primarily on the protein surface and accordingly contain polar and charged residues.
- Antibody recognition, phosphorylation, glycosylation, hydroxylation, and intron/exon splicing are found frequently at or adjacent to turns.
- Most proteins have compact, globular shapes, requiring reversals in the direction of their polypeptide chains. Many of these reversals are accomplished by a common structural element called the reverse turn (also known as the β turn or hairpin bend)

- 
- Turns were first identified by Venkatachalam (1968) who found three types each containing a hydrogen bond between the carbonyl oxygen of residue i and the amide nitrogen of i+3.
- These three types of turns are designated I, II, and III.
- Type III is simply a single turn of 3.10 helix
- Type I turns occur most frequently (2-3 times more frequently than type II).
- The mirror-image types I' and II' are rare but type I' appears to be preferred in beta-hairpins
- 



- Tertiary structure of protein
- The tertiary structure of a polypeptide chain is its overall three-dimensional shape, once all the secondary structure elements have folded together among each other.
- Interactions between polar, nonpolar, acidic, and basic R group within the polypeptide chain create the complex three-dimensional tertiary structure of a protein.
- When protein folding takes place in the aqueous environment of the body, the hydrophobic R groups of nonpolar amino acids mostly lie in the interior of the protein, while the hydrophilic R groups lie mostly on the outside.
- Cysteine side chains form disulfide linkages in the presence of oxygen, the only covalent bond forming during protein folding.
- All of these interactions, weak and strong, determine the final three-dimensional shape of the protein.
- When a protein loses its three-dimensional shape, it will no longer be functional.



Forces that stabilize tertiary structure

Disulfide bonds:

covalent bond between 2 SH groups of 2 cysteine residues forming an S~S bond of cystine residue.

Hydrophobic interaction:

non covalent bonds between amino acids with non-polar side chains that are located in the interior of polytpeptide chain away from water.

Hydrogen bonds:

non covalent bond between a hydrogen atom attached to nitrogen or oxygen and another oxygen or nitrogen atom.

Ionic interaction:

non covalent bonds between negatively charged groups in acidic amino acids (as carboxilic group in the side chain of aspartate or glutamate) and positively charged groups in basic amino acids (as amino group in the side chain of lysine)

***Van der Waal's* forces:**

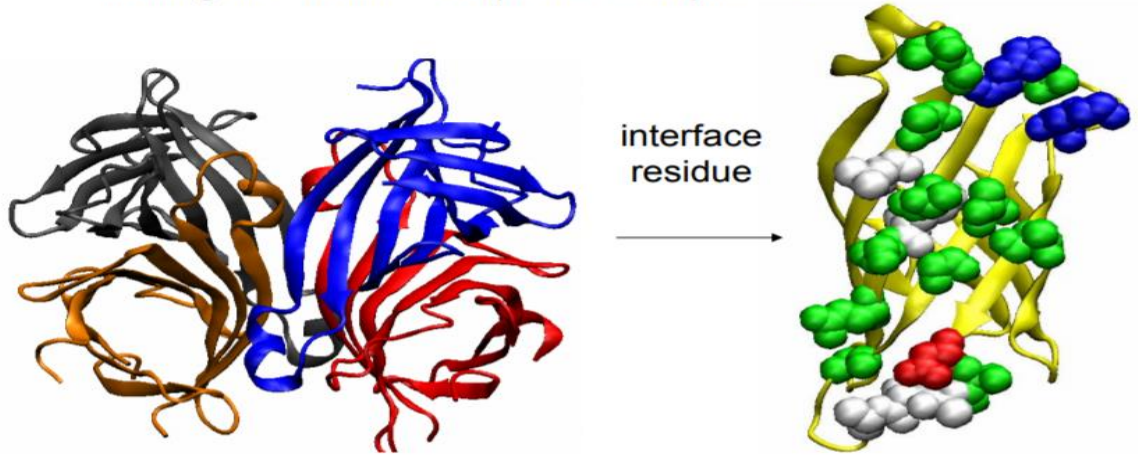
non covalent bonds occurring when two adjacent atoms come into closer distance.

Quaternary structure

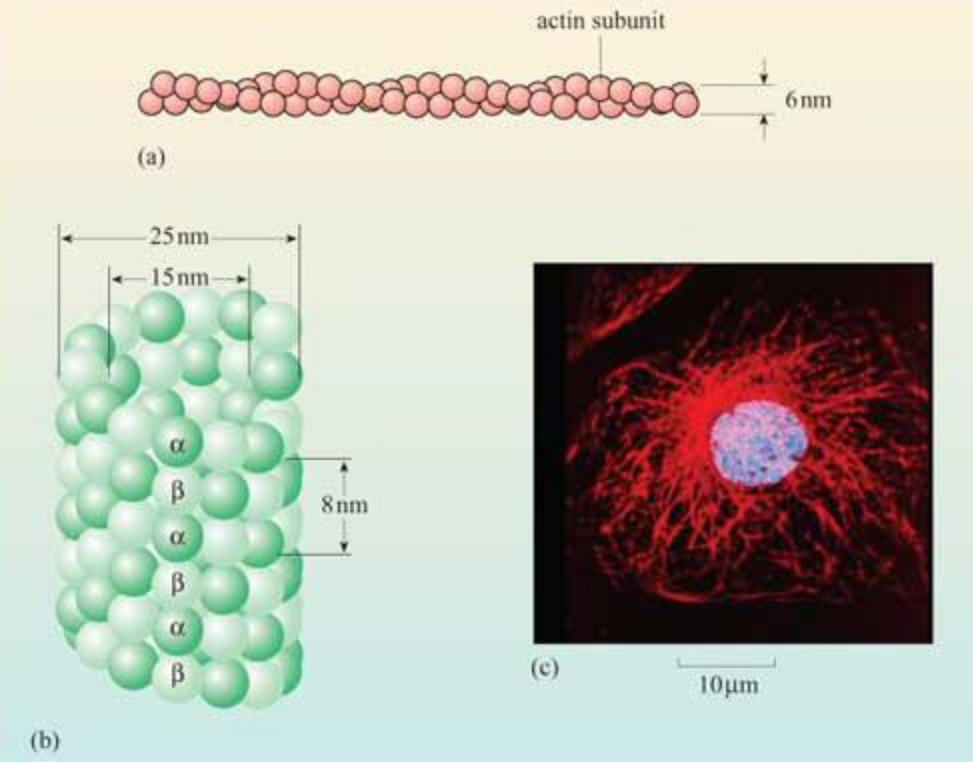
- This level of protein structure applies only to those proteins that consist of more than one polypeptide chain, termed subunits.
- In such proteins, sometimes referred to as multi subunit proteins, the same kinds of non-covalent interaction that stabilise the folded polypeptides also specify the assembly of complexes of subunits.
- Quaternary structure refers to the way in which the subunits of such proteins are assembled in the finished protein.
- Multi subunit proteins can have a number of identical (homomeric) or non-identical (heteromeric) subunits.
- The simplest multi subunit proteins are homodimers – two identical polypeptide chains that are independently folded but held together by non-covalent interactions.
- An example of a homodimeric protein is the Cro repressor protein from bacteriophage lambda , which turns off expression of specific genes in its bacterial host.
- Haemoglobin, the red blood cell protein responsible for carrying molecular oxygen, contains two each of two different subunits, termed α and β globin.

Formation of quaternary structure

- The subunits are held together by both **hydrophobic interactions** and **ionic interactions** between polar/charged amino side chains
  - a quaternary structure may fall apart in high salt environment
- A monomer typically buries 600 – 5000 Å of surface
- Understanding protein-protein interaction is key to understanding and controlling the formation of protein complexes



- Some proteins can assemble to form long filaments. Two such proteins are actin and tubulin
- . These proteins exist in a soluble globular form that can assemble into long helical filaments called microfilaments (actin) and microtubules (tubulin).
- Both these proteins are important components of the cytoskeleton, and the filaments that they form can extend from one end of a cell to another.
- Dynamic assembly and disassembly of microfilaments and microtubules is integral to the responsive nature of the cytoskeleton during many cell processes such as cell division, intracellular transport, and cell movement and adhesion.



Based on the chemical nature, structure, shape and solubility, proteins are classified as:

1. Simple proteins: They are composed of only amino acid residue. On hydrolysis these proteins yield only constituent amino acids. It is further divided into:
    1. Fibrous protein: Keratin, Elastin, Collagen
    2. Globular protein: Albumin, Globulin, Glutelin, Histones
    3. Conjugated proteins: They are combined with non-protein moiety.

Eg. Nucleoprotein, Phosphoprotein, Lipoprotein, Metalloprotein etc.

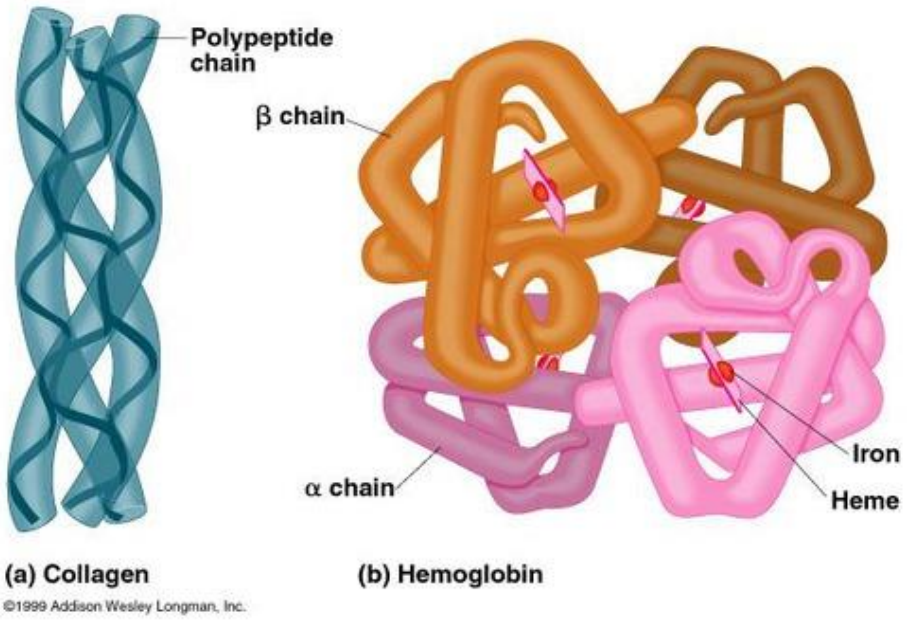
3. Derived proteins: They are derivatives or degraded products of simple and conjugated proteins. They may be :

    1. Primary derived protein: Proteans, Metaproteins, Coagulated proteins
    2. Secondary derived proteins: Proteosesn or albunoses, peptones, peptides.

Fibrous or Fibrillar Proteins

- These proteins are long ribbons or fibres in shape.
- These are mainly of animal origin and are insoluble in all common solvents such as water, dilute acids, alkalis and salts and in organic solvents.
- Most fibrous proteins serve in a structural or protective role.

- The are found exclusively in animals where they serve roles in flesh, connective tissues and hardened external structures, such as hair.
- They also contain the three common fibrous protein structures α -helices (keratins), β-strands/sheets (fibroin & elastin) and triple helices (collagen).



The fibrous proteins have some commonality of amino acid sequence. Each possesses an abundance of repeating sequences of amino acids with small, non-reactive side groups. Many contain short repeats of sequences, often with glycine.

Conn and Stumpf (1976) have classified globular proteins as follows:

Globular Proteins

- Cytochrome C
- Blood proteins
  - Serum albumin
  - Glycoproteins
  - Antibodies (= Immunoglobulins)
  - Hemoglobin
  - Hormones
- Enzymes
- Nutrient proteins

FIBROUS PROTEINS	GLOBULAR PROTEINS
Do not have a tertiary structure.	Have tertiary structure. Quaternary may or may not be present
Long fibres or sheets in shape.	Spherical in shape
Insoluble in water	Dissolve in water to form colloidal solution
The length of polypeptide chain may vary in two samples of the same fibrous protein	The length of polypeptide chain is always identical in two samples of the same globular protein.
Perform structural functions	Perform metabolic functions.
eg. Keratins, collagen, elastin and fibroin.	Egg albumin, serum globulin etc.

What is proteomics?

- Proteomics is the *large-scale study of proteomes.*
- A proteome *is a set of proteins produced in an organism, system, or biological context*.
- We may refer to, for instance, the proteome of a species (for example, *Homo sapiens*) or an organ (for example, the liver).
- The *proteome is not constant*; it differs from cell to cell and changes over time.
- To some degree, *the proteome reflects the underlying transcriptome.*
- However, protein activity (often assessed by the reaction rate of the processes in which the protein is involved) is also modulated by many factors in addition to the expression level of the relevant gene.

- The cell responds to internal and external changes by regulating the activity and level of its proteins; therefore changes in the proteome (a collection of all the proteins coded in our genes) provide a snapshot of the cell in action.
- Proteomics enables the understanding of the structure, function, and interactions of the entire protein content in a specific organism.

History of proteomics

- The term "protein" was initially introduced in 1938 by the Swedish chemist Jöns Jakob Berzelius, an accomplished experimenter in the field of electrochemistry. He wanted to describe a particular class of macromolecules that are plentiful in living organisms and made up of linear chains of amino acids.
- The first protein studies that can be called proteomics began in 1975 with the introduction of the two-dimensional gel and mapping of the proteins from the bacterium *Escherichia coli*, guinea pig, and mouse. Albeit many proteins could be separated and visualized, they could not be identified.
- The terms "proteome" and "proteomics" were coined in the early 1990s by Marc Wilkins, a student at Australia's Macquarie University, in order to mirror the terms "genomics" and "genome", which represent the entire collection of genes in an organism.
- Since the first use of the term "proteome", its meaning and scope have narrowed.
- Post-translational modifications, alternative splice products, and proteins intractable to classic separation techniques have presented a challenge towards the realization of the conventional definition of the word.
- Today, many different areas of study are explored by proteomics.
- Amongst them are protein-protein interaction studies, protein function, protein modifications, and protein localization studies.
- The fundamental goal of proteomics is not only to pinpoint all the proteins in a cell but also to generate a complete three-dimensional map of the cell indicating their exact location.
- In many ways, proteomics runs parallel to genomics.

- The starting point for genomics is a gene in order to make inferences about its products (*i.e.* proteins), whereas proteomics begins with the functionally modified protein and works back to the gene responsible for its production.
- Unlike the genome (the complete set of genes within each organism), the composition of the proteome is in a constant state of flux over time and throughout the organism.
- Therefore, when scientists refer to the proteome, they are also sometimes referring to the proteome at a given point in time (such as the embryo versus the mature organism), or to the proteome of a particular cell type or tissue within the organism.
- What is proteomics?
-

  Proteomics is the study of the proteome—investigating how different proteins interact with each other and the roles they play within the organism.
- Although protein expression can be inferred by studying the expression of mRNA, which is the middle man between genes and proteins, mRNA expression levels do not always correlate well with protein expression levels.
- Furthermore, the study of mRNA does not consider protein posttranslational modifications, cleavage, complex formation and localization, or the many variant mRNA transcripts that can be produced; all of which are key to protein function.

Proteomics is used to investigate:

1. When and where proteins are expressed
2. Rates of protein production, degradation, and steady-state abundance
3. How proteins are modified (for example, post-translational modifications (ptms) such as phosphorylation)
4. The movement of proteins between subcellular compartments
5. The involvement of proteins in metabolic pathways
6. How proteins interact with one another
7. What are the key questions that proteomics can answer?

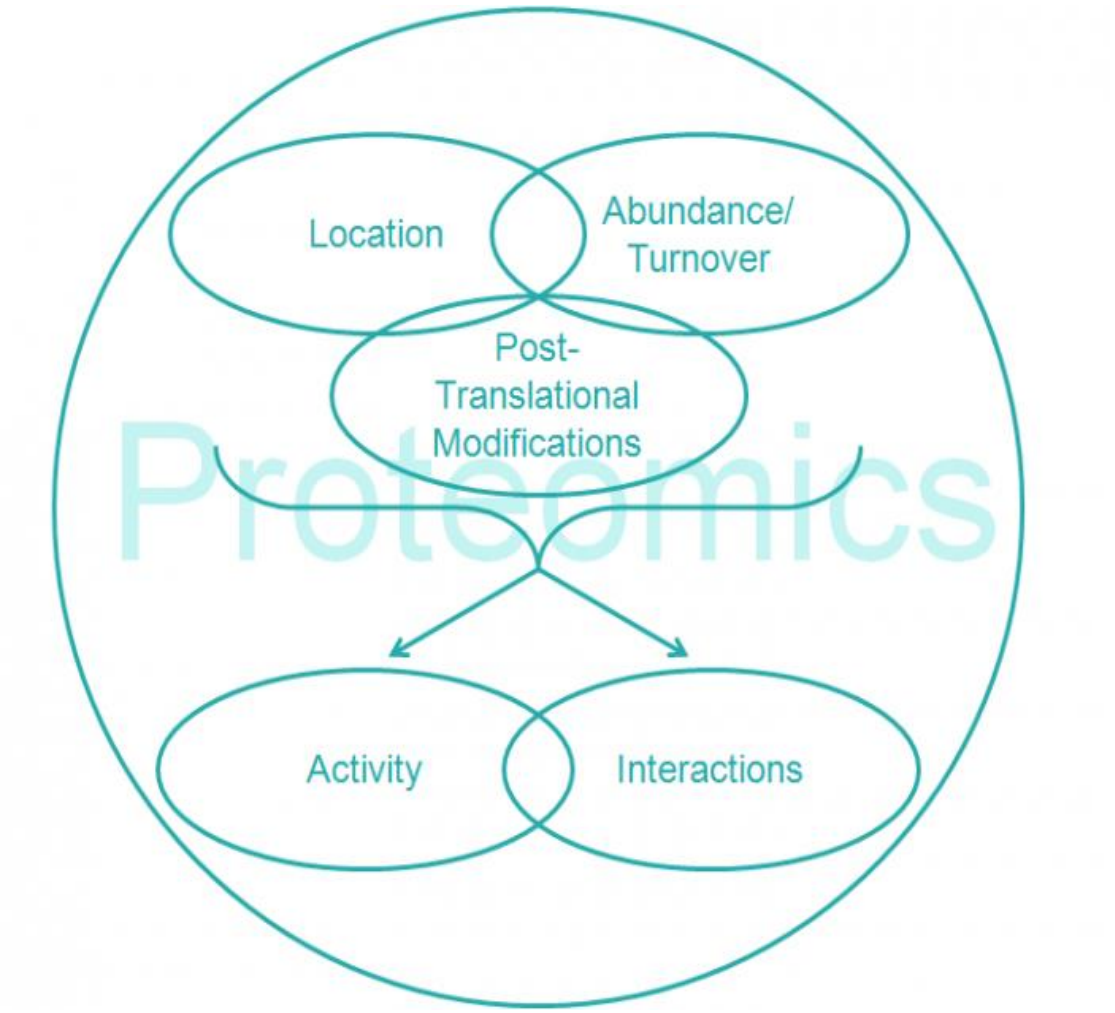
   Broadly speaking, proteomic research provides a global view of the processes underlying healthy and diseased cellular processes at the protein level.[3,4]

8. To do this, each proteomic study typically focuses on one or more of the following aspects of a target organism's proteome at a time to slowly build on existing knowledge:

Protein identification	Which proteins are normally expressed in a particular cell type, tissue or organism as a whole, or which proteins are differentially expressed?
Protein quantification	Measures total ("steady-state") protein abundance, as well as investigating the rate of protein turnover (i.e., how quickly proteins cycle between being produced and undergoing degradation).
Protein localization	Where a protein is expressed and/or accumulates is just as crucial to protein function as the timing of expression, as cellular localization controls which molecular interaction partners and targets are available.
Post-translational modifications	Post-translational modifications can affect protein activation, localization, stability, interactions and signal transduction among other protein characteristics, thereby adding a significant layer of biological complexity.
Functional proteomics	This area of proteomics is focused on identifying the biological functions of specific individual proteins, classes of proteins (e.g., kinases) or whole protein interaction networks.
Structural proteomics	Structural studies yield important insights into protein function, the "druggability" of protein targets for drug discovery, and drug design.
Protein-protein interactions	Investigates how proteins interact with each other, which proteins interact, and when and where they interact.

Areas of proteomics:

- Proteomic experiments generally collect data on three properties of proteins in a sample:
- location, abundance/turnover and post-translational modifications.
- Depending on the experimental design, researchers may be directly interested in these data, or may use them to infer additional information.
- For example, it may be possible to infer a protein's interaction partners among others that are co-localised with it, or to assess whether a protein is active from its post-translational modifications.

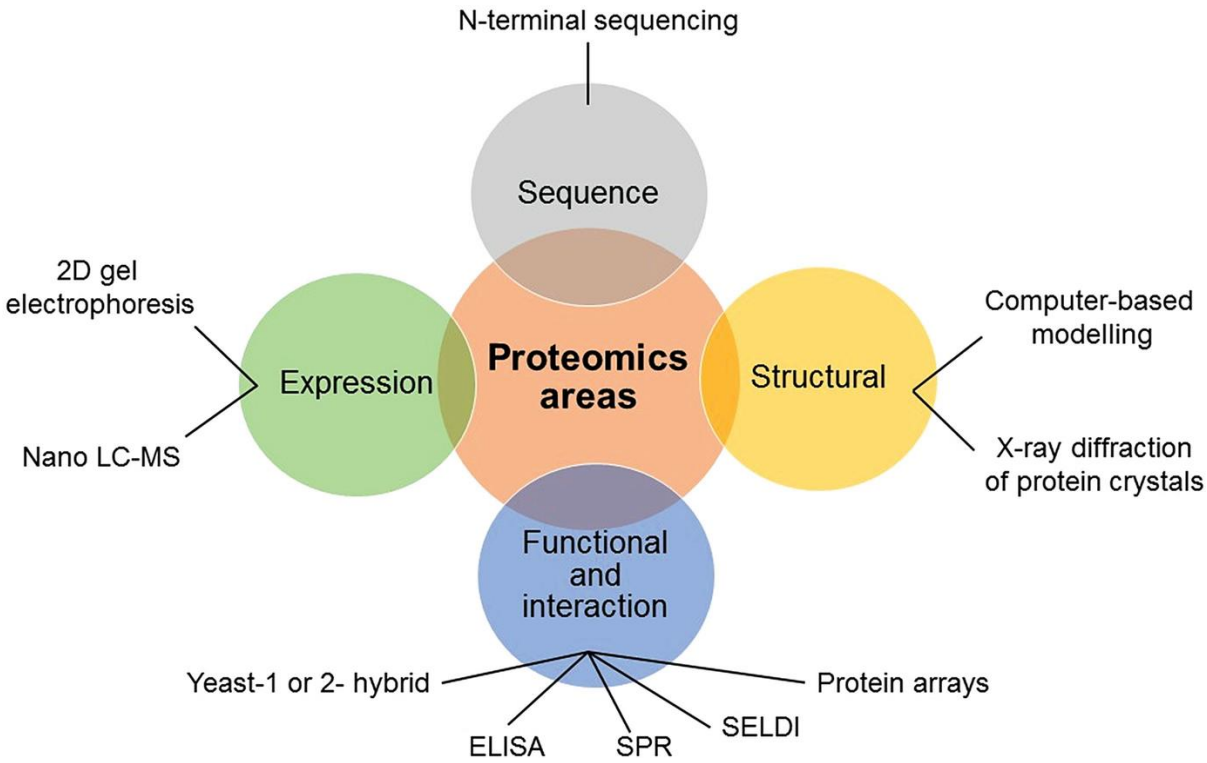


Several high-throughput technologies have been developed to investigate proteomes in depth.

The most commonly applied are mass spectrometry (MS)-based techniques such as Tandem-MS and gel-based techniques such as differential in-gel electrophoresis (DIGE).

These high-throughput technologies generate huge amounts of data.

Databases are critical for recording and carefully storing this data, allowing the researcher to make connections between their results and existing knowledge.



Data analysis in proteomics

- Proteomic studies, particularly those employing high-throughput technologies, can generate huge amounts of data.
- In addition to the sheer quantity of data produced, proteomic data analysis can also be relatively complex for certain techniques such as shotgun MS.
- Adding to this complexity is the range of bioinformatics tools available for proteomic analyses.

Approach	Description	Method
Top-down proteomics	Proteins in a sample of interest are first separated before being individually characterized.	Protein separation is performed based on mass and charge with e.g., 2DE, DIGE or MS. When using 2D electrophoresis techniques, the proteins are first resolved on the gel and then individually digested into peptides that are analyzed by a mass spectrometer. When using MS directly, the undigested sample containing the whole proteins is

		injected into the mass spectrometer, the proteins are separated, and individual proteins are then selected for digestion and a further round of MS for analysis of the peptides.
Bottom-up proteomics, or "shotgun proteomics"	All the proteins in the sample are first digested into a complex mixture of peptides, and these peptides are then analyzed to identify which proteins were present in the sample.	Proteins are first digested, and the digested peptide mixture is fractionated and subjected to MS, frequently in an LC-MS/MS configuration. The resulting peptide sequences are compared to existing databases using automated search algorithms. These search engines match the experimentally obtained peptide spectra to the predicted spectra of proteins produced by in silico digestion (this is called "peptide-spectrum matching"). There are several different bottom-up workflows possible, including data-dependent and data-independent methods, as well as hybrids of these.

High-throughput methods:

1. Analytical, functional and reverse-phase microarrays

Protein microarrays apply small amounts of sample to a "chip" for analysis (this is sometimes in the form of a glass slide with a chemically modified surface). Specific antibodies can be immobilized to the chip surface and used to capture target proteins in a complex sample. This is termed an analytical protein microarray, and these types of microarray are used to measure the expression levels and binding affinities of proteins in a sample. Functional protein microarrays are used to characterize protein functions such as protein–RNA interactions and enzyme-substrate turnover. In a reverse-phase protein microarray, proteins from e.g., healthy vs. diseased tissues or untreated vs. treated cells are bound to the chip, and the chip is then probed with antibodies against the target proteins.

The differences between forward phase and reverse phase protein microarrays.

2. Mass spectrometry-based proteomics

There are several "gel-free" methods for separating proteins, including isotope-coded affinity tag (ICAT), stable isotope labeling with amino acids in cell culture (SILAC) and isobaric tags for relative and absolute quantitation (iTRAQ). These approaches allow for both quantitation and comparative/differential proteomics. There are also other, less quantitative techniques such as multidimensional protein identification technology (MudPIT), which offer the advantages of being faster and simpler. Other gel-free, chromatographic techniques for protein separation include gas chromatography (GC) and liquid chromatography (LC).

Mass spectrometry workflow

Regardless of how the protein sample is separated, the downstream MS workflow comprises three main steps:

1. The proteins/peptides are ionized by the ion source of the mass spectrometer.
2. The resulting ions are separated according to their mass to charge ratio by the mass analyze.
3. The ions are detected.

When using gel-free techniques upstream of MS such as iTRAQ or SILAC, the samples are used directly for input into the mass spectrometer. When using g

el-based techniques, the protein spots are first cut out of the gel and digested before being either separated by LC or directly analyzed by MS.

There are two main ionization sources, namely:

- Matrix-assisted laser desorption/ionization (MALDI)
- Electrospray ionization (ESI)

2. Mass spectrometry-based proteomics

There are several "gel-free" methods for separating proteins, including isotope-coded affinity tag (ICAT), stable isotope labeling with amino acids in cell culture (SILAC) and isobaric tags for relative and absolute quantitation (iTRAQ). These approaches allow for both quantitation and comparative/differential proteomics. There are also other, less quantitative techniques such as multidimensional protein identification technology (MudPIT), which offer the advantages of being faster and simpler. Other gel-free, chromatographic techniques for protein separation include gas chromatography (GC) and liquid chromatography (LC).

Mass spectrometry workflow

Regardless of how the protein sample is separated, the downstream MS workflow comprises three main steps:

1. The proteins/peptides are ionized by the ion source of the mass spectrometer.
2. The resulting ions are separated according to their mass to charge ratio by the mass analyze.
3. The ions are detected.

When using gel-free techniques upstream of MS such as iTRAQ or SILAC, the samples are used directly for input into the mass spectrometer. When using gel-based techniques, the protein spots are first cut out of the gel and digested before being either separated by LC or directly analyzed by MS.

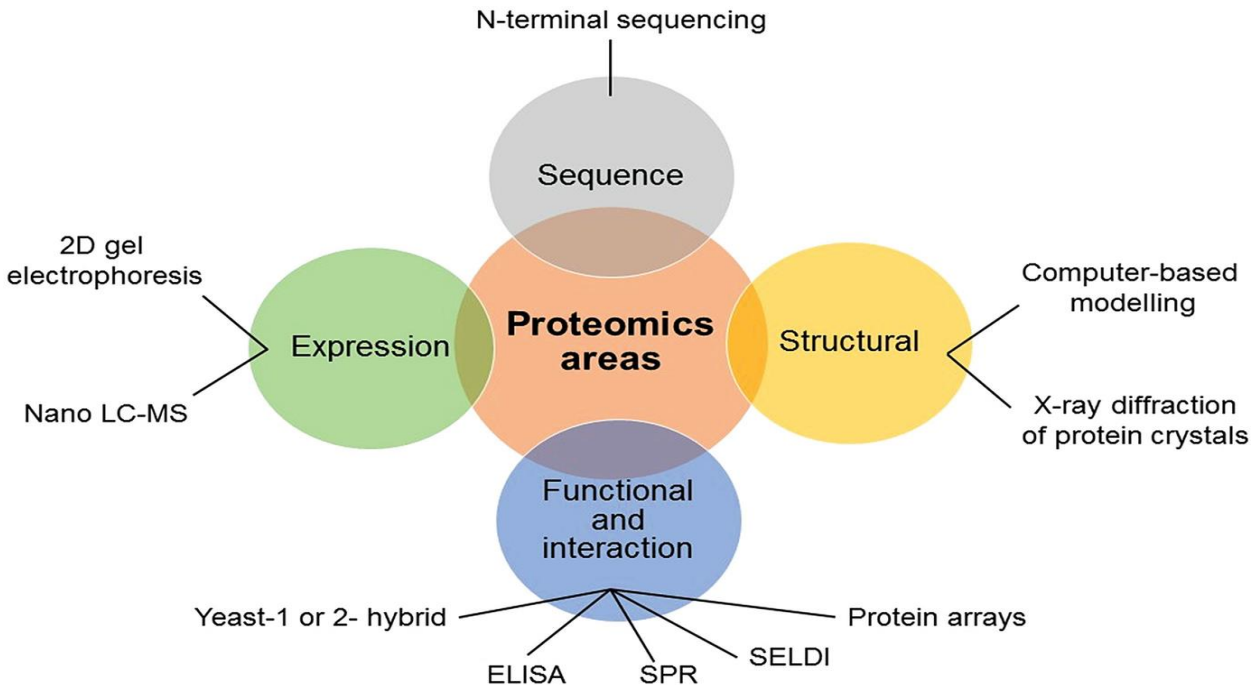
There are two main ionization sources, namely:

- Matrix-assisted laser desorption/ionization (MALDI)
- Electrospray ionization (ESI)

  The applications of proteomics are incredibly numerous and varied. The table below lists just some of these applications and provides links to examples of studies using these approaches.

Proteomics application	Description and examples
Personalized medicine	Tailoring disease treatment to each patient based on their genetic and epigenetic makeup, so as to improve efficacy and reduce adverse effects. While genomics and transcriptomics have been the main focus of such studies to date, proteomics data will likely add a further dimension for patient-specific management.
Biomarker discovery	Identification of protein markers for e.g., the diagnosis and prognosis of glioblastoma, and evaluating patients' response to therapeutic interventions such as stem cell transplantation.
Drug discovery and development	Identifying potential drug targets, examining the druggability of selected protein targets, and developing drugs aimed at candidate therapeutic protein targets (e.g., for hepatocellular carcinoma).

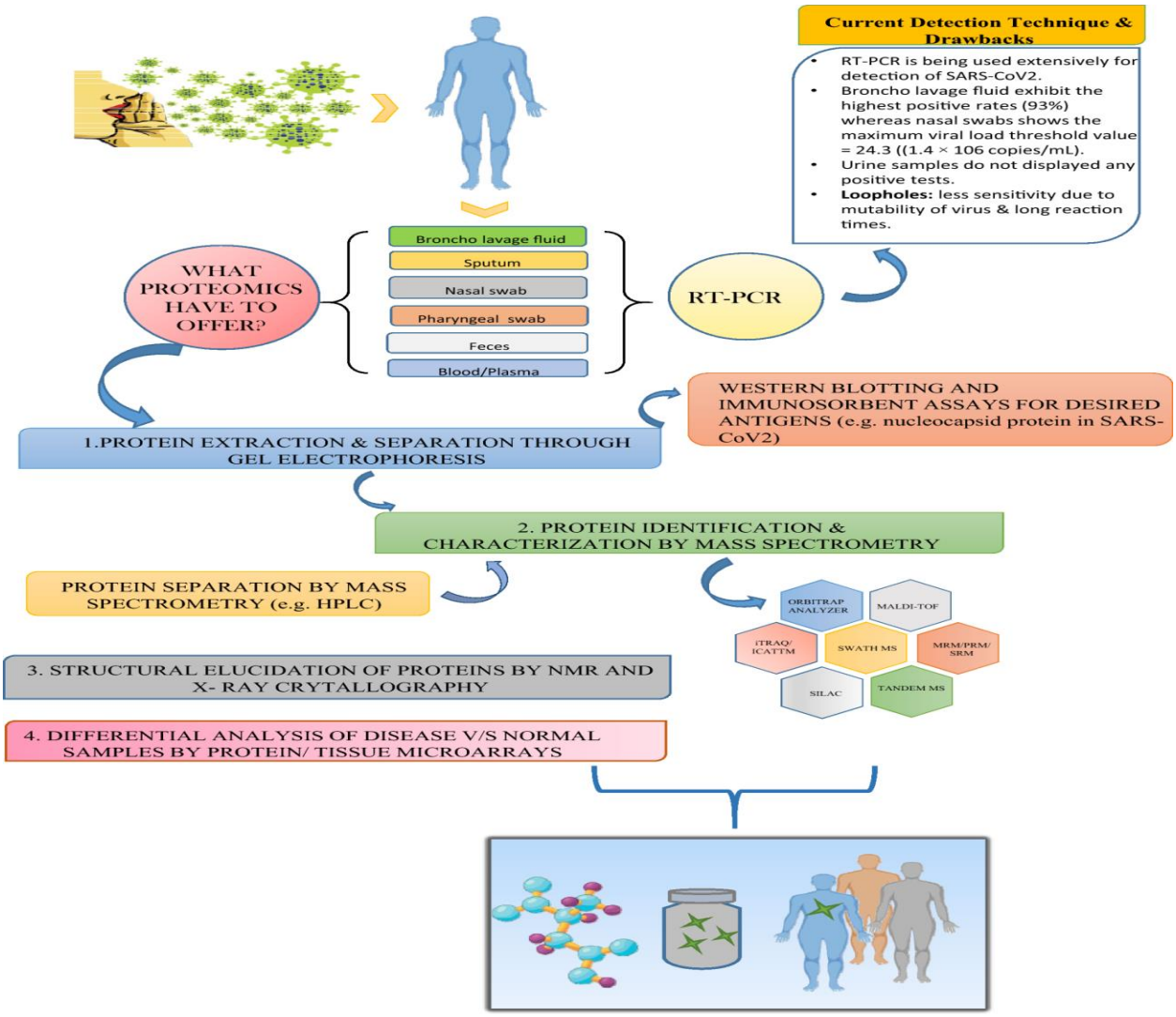
Systems biology	Investigations of plant–pathogen interactions, crop engineering for increased resilience to e.g., flooding, drought and other environmental stresses.
Agriculture	Food safety and quality control, allergen detection and improving the nutritional value of foods.
Food science	The study of ancient proteins to further our understanding of evolution and archeology.
Paleoproteomics	Investigations of how mammals' immune systems may respond to exo-microbes found in space and studies of the prebiotic organic matter found on meteorites.
Astrobiology	



Proteomics studies whose goal is to map out the proteins present in a specific cellular organelle or the structure of protein complexes are known as structural proteomics. Structural analysis can aid in the identification of the functions of newly discovered genes, show where drugs bind to proteins and where proteins interact with each other. Technologies employed in structural proteomics are X-ray crystallography and nuclear magnetic resonance spectroscopy.

The quantitative study of protein expression between samples that differ by a certain variable is known as expression proteomics. This type of proteomics can help identify the main proteins found in a particular sample and proteins differentially expressed in related samples, *e.g.* when comparing diseased and healthy tissue. Technologies such as 2D-PAGE and mass spectrometry are used here.

Functional proteomics represents a wide-ranging term for many specific, directed proteomics methodologies. The characterization of protein-protein interactions is used to determine protein functions and to demonstrate how proteins assemble in larger complexes. In some cases, specific subproteomes are isolated by affinity chromatography for additional analysis.



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOINFORMATICS

UNIT – 2- SBIA1303 – Proteomics and Interactomics

Unit-2

Protein purification

Purity is defined by the general level of protein contaminants and also by the absence of contaminants of special interests such as microbes, toxins etc.

Protein purification is divided into five stages:

Preparation of sources

Knowledge of protein properties

Development of an assay

Primary isolation

Final purification

I. Preparation of sources:

- The raw materials from which proteins can be isolated such as microbial culture or animals or plant sources should be selected.
- The amount of protein can be increased by increasing cultivation volume.

II. Knowledge of protein properties:

- Before employing any procedure, one should know about different properties of proteins such as- *intracellular or extracellular occurrence, denaturation temperature, pH range, ionic stability, molecular weight, charge, iso-electric point, binding partners* etc.

III. Development of an assay:

- An assay developed should be convenient, easy, rapid, precise for purification.

IV. Primary isolation:

This consists of separation of protein from other cellular components.

For this propose there are different methods:

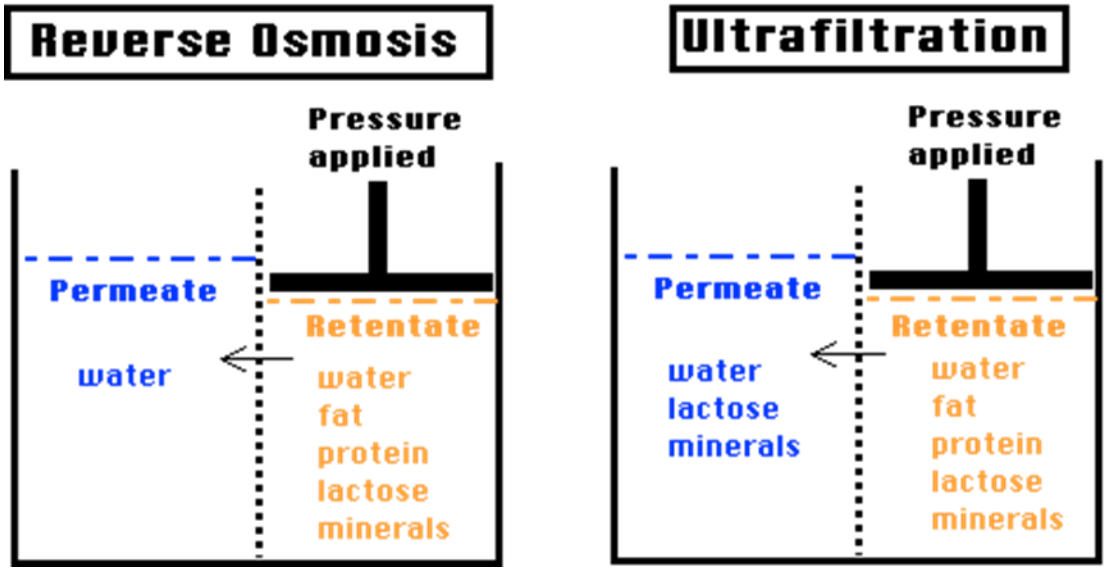
Concentration:

Different methods can be employed for concentration of extracellular protein.

Ultrafiltration are usually used to concentrate extracellular proteins from cell.

Ultrafiltration is a membrane filtration in which hydrostatic pressure is applied which causes movement of solution across the semi-permeable membrane.

- Water and low molecular weight solute pass while other high molecular weight of molecules trapped in membrane.
- The protein molecules are adsorbed in the membrane surface.

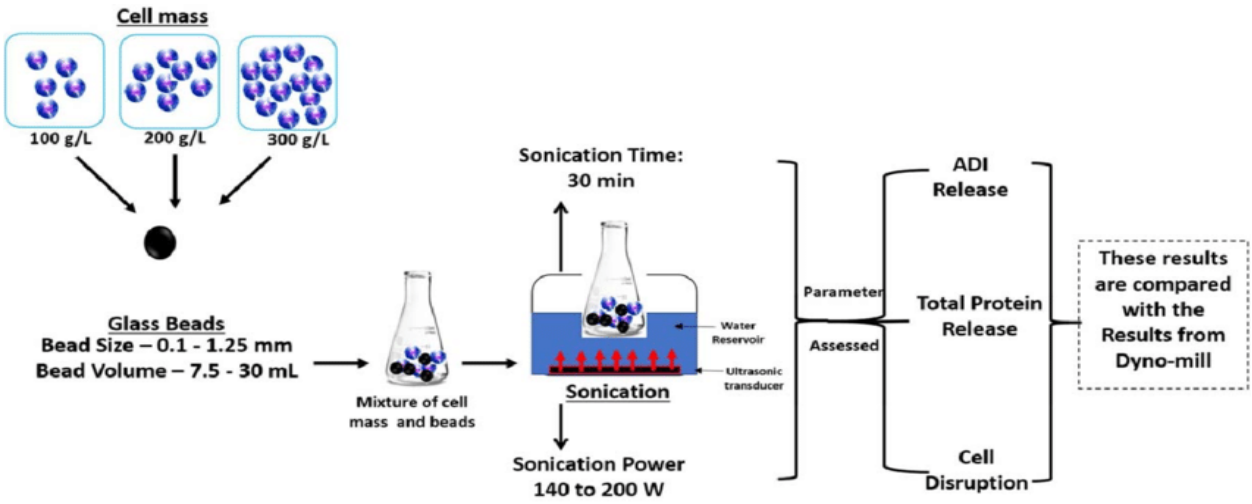


Cell lysis: (For intracellular protein)

- The intracellular proteins are liberated by cell lysis.
- There are different methods for cell lysis.

Physical method:

- Mechanical method: Bead mill, Homogenizer, Microfluidizer, Sonicator, French press/ X-mesh
- Non-mechanical method: Decompression, Osmotic shock, Thermolysis, Freeze thaw, Dessication, Cell bomb



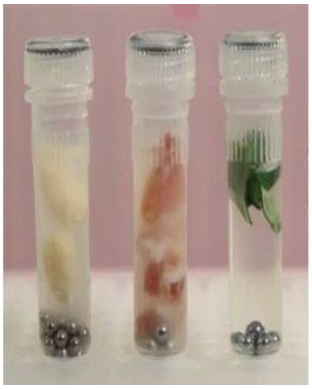
Bead-based homogenization uses plastic or metal beads combined with high-speed shaking to create shearing forces. This technique is well-suited to whole animal, insect (e.g., Drosophila melanogaster) or plant specimens, which requires disruption of sturdy cell walls.

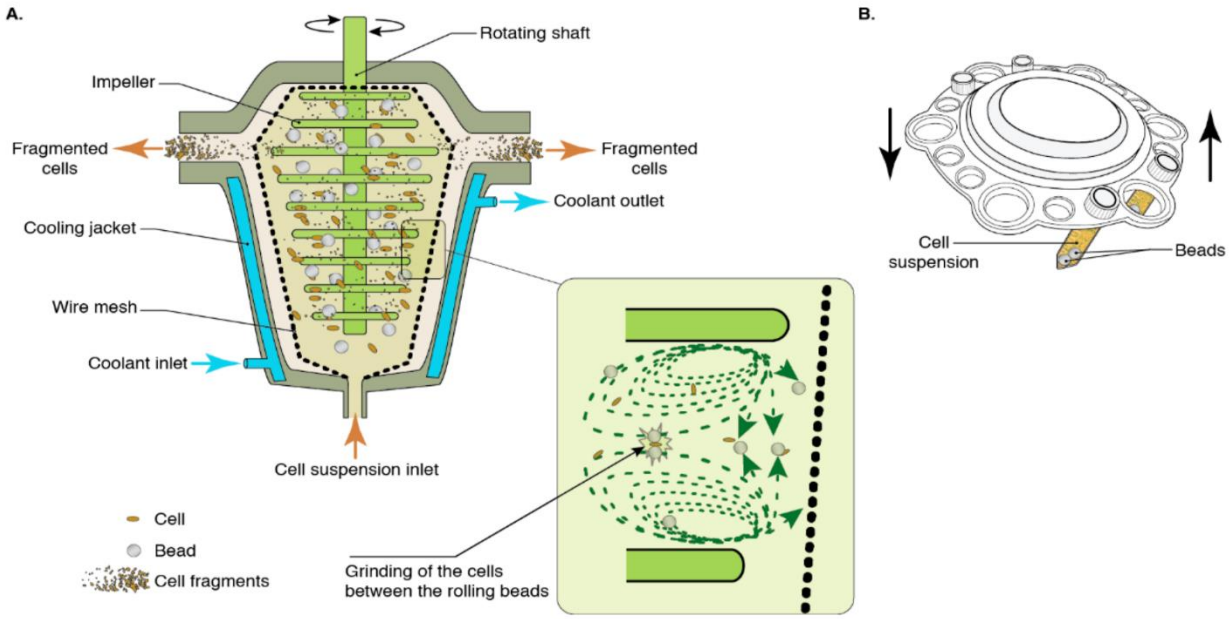
A drawback of bead-based disruption is that it requires the proper selection of bead material and diameter.

For example, some plastic beads will readily bind with DNA, depleting it from solution when they are removed.

Similarly, the diameter of the bead can impact the amount of force created and the degree to which a sample is disrupte

Using a bead size that is either too large or too small can result in incomplete homogenization.



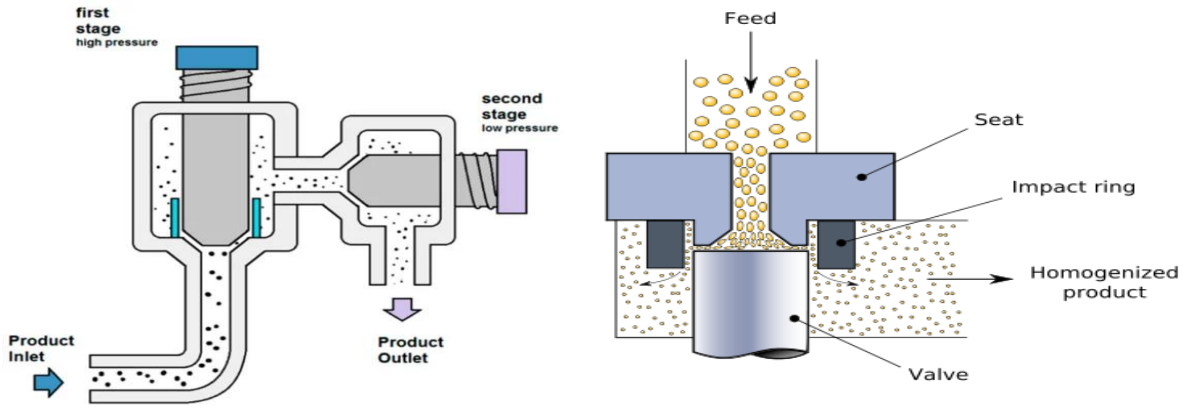


Mechanical Homogenization

- Mechanical homogenization utilizes direct physical force to bring a biological sample in solution to a state of uniform distribution, such that all fractions' molecular composition is consistent.
- Traditionally, mechanical disruption was achieved by freezing tissues and then grinding with a mortar and pestle.
- Presently, there are two other common methods: bead-based disruption and rotor-stator disruption.
- These separation techniques efficiently homogenize samples, but often require further downstream fractionation to obtain the desired concentration or purity of molecules.
- Rotor-stator disruption involves the use of a stationary stator housing a rapidly moving rotor.
- The movement of tissues and cells within the space between the rotor and stator creates a high degree of shearing force. These devices are typically hand-held and the rotor-stator attachment is interchangeable.
- These devices efficiently disrupt tough animal and plant tissues (e.g., muscle).
- One salient benefit of rotor-stator disruption is the ability to progressively increase the degree of homogenization by sequentially processing with increasingly smaller rotor-

stator distance attachments. This process can effectively homogenize even the most resilient specimens.

- Additionally, very large-volume samples can be processed using larger disruption instrumentation.
- Drawbacks of rotor-stator disruption include increased cost and more cumbersome equipment management. Attachments often require washing and sterilization prior to reuse. Disposable attachments are offered, but with a higher operating cost.



Rotor-stator disruption involves the use of a stationary stator housing a rapidly moving rotor. The movement of tissues and cells within the space between the rotor and stator creates a high degree of shearing force. These devices are typically hand-held and the rotor-stator attachment is interchangeable.

These devices efficiently disrupt tough animal and plant tissues (e.g., muscle).

One salient benefit of rotor-stator disruption is the ability to progressively increase the degree of homogenization by sequentially processing with increasingly smaller rotor-stator distance attachments. This process can effectively homogenize even the most resilient specimens.

Additionally, very large-volume samples can be processed using larger disruption instrumentation.

Drawbacks of rotor-stator disruption include increased cost and more cumbersome equipment management. Attachments often require washing and sterilization prior to reuse. Disposable attachments are offered, but with a higher operating cost.

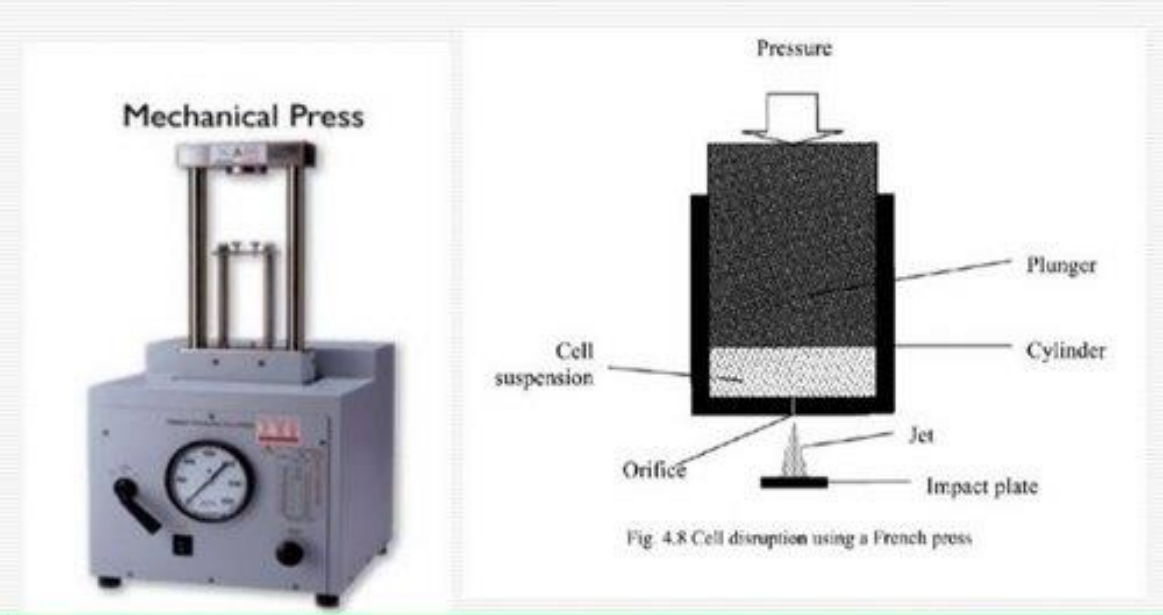
Microfluidizer

- Cell disruption is the method or process for releasing biological molecules from inside a cell.
- Utilizing these intracellular contents (proteins, organelles, DNA/RNA, enzymes and Adeno-Associated Virus (AAVs) Vectors for Gene Therapy Applications) which are found and/or grown inside cells is the next generation for drug development.
- For cells that do not secrete these intracellular contents, it is vital to lyse the cell to liberate these contents. During this process, it is important to prevent denaturing of these intracellular components by the unnecessary elevation of temperature or excessive shear rates.
- Microfluidizer® processors effectively rupture a variety of cell types which require different levels of shear — including bacterial, mammalian, plant, insect, fungi, algae and yeast cells — whilst ensuring high protein recovery. They are tough on cell walls but gentle on the intracellular contents.
- LM20 Microfluidizer® high-pressure homogenizer can handle the toughest of cell types. The LM10 Microfluidizer® homogenizer ensures repeatability and is simple to operate.

French Press

- Generates high pressure in a pressure cell.
- A manually controlled valve releases the pressurized fluid from the pressure cell, resulting in cell rupture.
- Not scalable or repeatable; needs strength to close and open the valve.
- There are numerous hazards involved with French Presses.
- They are difficult and time-consuming to clean, which has to be done for every sample.

Most manufacturers of French Presses have discontinued production but they are still in use, available from small companies and second hand



Sonicator

- Sonication applies sound energy to agitate particles in your sample. The ultrasonic frequency used is usually greater than >20 kHz. In an experimental setting this is usually carried out using an ultrasonic bath or ultrasonic probe generally referred to as sonication.
- In the laboratory sonication is used mainly as a method of cell disruption. Sonication is used to disrupt cellular membranes and release the cells contents, this process is generally referred to as sonoporation.
- Sonication is carried out during the preparation of protein extracts in order to break the cell apart. Although lysis buffer can be used sonication can help break the cell apart.
- Sonication can also be used to fragment/shear DNA, preventing it from interfering with further sample preparation. Other biological uses include the production of nanoparticles, liposomes, extraction of anthocyanins and antioxidants.
- Depending on your cell type (bacterial or eukaryotic) it can be difficult to lyse certain cell types and placing them in a detergent buffer alone wont result in full cell lysis.
- Furthermore you may also require to lyse the cellular organelles and not just lyse the cell wall to release the cytosol.

- Sonication of cells using a titanium probe can help lyse cells fully and help all extract all DNA, RNA and protein contents of your cells. This can help downstream when looking for more homogenous extract for ELISA assays and immunoprecipitation.

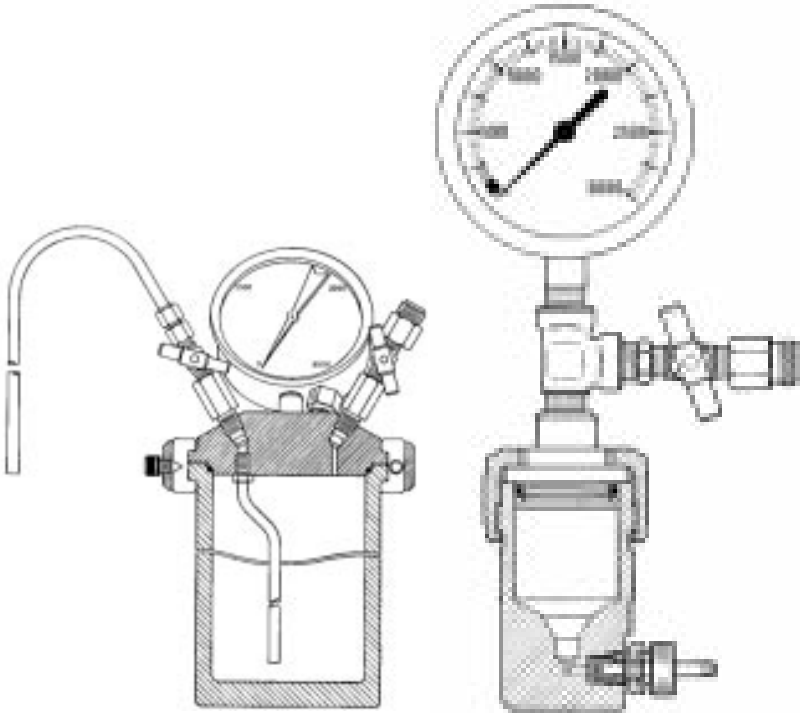
	Microfluidizer	Homogenizer	Bead Mill	French Press	Sonication
Principle	Fixed-geometry interaction chamber and constant pressure pumping system. Uniform and highly precisely controlled shear rates.	Variable valve geometry in combination with less constant pressure profile. Less controlled energy input/ less uniformity.	Cylindrical rotating shell partly filled with beads that fall onto the material to be ground. Forces applied are impact and attrition.	Pressurization and decompression depends on a manually operated valve. Speed of human user's valve turn determines actual applied shear.	Uses cavitation to generate shear – typically much lower than high pressure methods. Increasing shear results in higher processing temperatures.
Continuous	Yes	Yes	No	No	No
Scalable	Yes	Limited	Yes	No	Limited
Optimal Temp Control	Yes *(cooling coils or heat exchangers)*	Yes *(cooling coils or heat exchangers)*	No	No	No
Contamination Free	Yes	Uncertain	No		
Minimum Volume	1ml	10 ml	1 ml	1 ml	< 1ml
Constant Shear Rate	Yes	No	No	No	No

- Non-mechanical method: Decompression, Osmotic shock, Thermolysis, Freeze thaw, Dessication, Cell bomb
- Decompression:

A rapid and effective way to:

- Homogenize cells and tissues
- Release intact organelles
- Prepare cell membranes
- Release labile biochemicals
- Produce uniform and repeatable homogenates without subjecting the sample to extreme chemical or physical stress

- Cell disruption by rapid decompression from a pressure vessel has been used for many years by investigators who wanted to overcome the limitations imposed by other cell disruption procedures.

- Although the technique is not new, interest in the decompression method and many new applications for it have grown rapidly in recent years following the introduction of convenient pressure equipment such as the Parr Cell Disruption Bomb.
- The nitrogen decompression method is particularly well suited for treating mammalian and other membrane bound cells.
- It has also been used successfully for treating plant cells, for releasing virus from fertilized eggs and for treating fragile bacteria.
- It is not recommended for untreated bacterial cells, but this restriction can be eliminated by using various pretreatment procedures to weaken the cell wall.
- Yeast, fungus, spores and other materials with tough walls do not respond well to this method.
- HOW IT WORKS
- 
- The principle of the method is quite simple. Large quantities of nitrogen are first dissolved in the cell under high pressure within a suitable pressure vessel.
- Then, when the gas pressure is suddenly released, the nitrogen comes out of the solution as expanding bubbles that stretch the membranes of each cell until they rupture and release the contents of the cell.
- Osmotic shock

Osmotic shock:

- In this method, either hypotonic or hypertonic solution is used.
- The cell suspension is placed on either of the solution which create osmotic shock.
- Hypotonic solution:
    - Plasmolysis occur, the water enters the cytoplasm of cell and well burst.
- Hypertonic solution
    - Plasmolysis occur, the cell shrinks due to loss of water from cell.
- This is not always effective so it is used in combination with other process.
- Thermolysis

Thermolysis (Heat shock protein):

- This method is very easy and economical for heat stable product.
- High heat inactivates cell by disrupting cell wall and release intracellular products.
- The effect of heat depends upon various factors such as pH, temperature, chelating agent, ionic strength, presence of enzymes (proteolytic and hydrolytic), time etc.
- Disadvantage:
    - Cannot be used for heat labile substances
    - Spore forming bacteria are also resistant to this method.
- Freeze thaw
- This is commonly used method to disrupt animal and bacterial cell.
- In this method, the cell suspension is frozen in dry ice an ethanol or freezer and then thaw the suspension at room temperature (37^oC).
- This suddenly freezing and thawing causes the cells to swell and break.
- The process is repeated for several time for efficient lysis.
- Disadvantages:
    - Very slow process
    - Not always effective

Dessication

- In this method, the cell suspension is dried by air drying or vacuum drying.
- After dessication, the cell shrinks.
- If excess water is provided than cell burst.
- This process is repeated for several time.

- Disadvantages:
    - Slower process and require several repeats.
    - Always used in combination with other process.

Electric lysis (Electroporation):

- When cell suspension is placed in electric field with intensity higher than certain threshold, there is formation of nano scale pore on cell surface.
- This pore maybe reversible or irreversible, depending upon applied field and direction of field.
- Through the pores, the intracellular material releases out.
- Disadvantages:
    - Heat generation
    - Expensive process
- Cell bomb
- Cell disruption by nitrogen decompression from a pressurized vessel is a rapid and effective way to homogenize cells and tissues, to release intact organelles, and to prepare cell membranes.
- Cells are placed in a pressure vessel and large quantities of oxygen-free nitrogen are dissolved in the cells under high pressure (~5500 kilopascals [kPa], equivalent to 800 pounds per square inch [psi]).
- When the pressure is released suddenly, the nitrogen bubbles out of solution, rupturing the cell membrane and releasing the cell contents.
- Nitrogen cavitation is well suited for mammalian and plant cells and fragile bacteria, but is less effective with yeast, fungi, spores, or other cell types with tough cell walls.

- The chemical and physical stresses imposed by nitrogen cavitation on enzymes and subcellular compartments are minimized compared with ultrasonic and mechanical homogenizing methods.
- Unlike lysis methods relying on shear stresses and friction, there is no heat damage to proteins and organelles during nitrogen cavitation.
- Indeed, the method is accompanied by an adiabatic expansion that cools the sample instead.
- Also, labile cell components are protected from oxidation by the inert nitrogen gas. Furthermore, nitrogen does not alter the pH of the suspending medium.
- The process is fast and uniform because the same disruptive forces are applied within each cell and throughout the sample, ensuring reproducible cell-free homogenates.
- Finally, variable sample sizes (e.g., from ~1 mL to 1 L or more) can be accommodated with most commercial systems.

Chemical method:

- Chemical permeabilizer whents
- Antibiotics
- Detergents
- Chartrops
- Chelating agents
- Hydroxides and hypochlorides

Enzymatic method:

- Autolysis

- Lytic enzyme
- Phage mediated lysis
- After cell lysis, the cellular constituents are concentrated by ultrafiltration.

i. Chemical permeabilizer:

- Chemicals such as Toluene, ether, phenylethylalcohol, benzene, methanol, chloroform etc. form a channel in cell membrane.
- These are lipid solubilizer, forming pore through which cellular content releases out.

ii. Antibiotics:

Polymyxin, azoles, Nystatin are cell membrane inhibitor and destroy cell membrane inhibitor and destroy cell membrane formation causing release of cellular content.

iii. Detergents:

- They are also called as surfactant.
- They solubilize lipid and denature protein.
- There are three types of detergents:
- Anionic – e.g. SDS
- Cationic
- Non-ionic- tween-20, tritan X-100, tritan X-400 etc.
    - The non-ionic detergents are commonly used in cell disruption.
    - The process is very fast.
- Disadvantages:
    - Disruption might be incomplete.
    - Repetition may be required.

iv. Chelating agent:

- EDTA is good example of chelating agent.
- It chelates cations (bivalent) and make unavailable for cell causing disruption of cell membrane. (Mg^{2+}, Ca^{2+})

v. Chaotropic agent:

- Urea, Guanidine chloride etc. are chaotropic agent.
- They disrupt the structure of protein and nucleic acids and decrease hydrophobic interaction among the surrounding molecules by disordering water molecules adjacent to protein molecule.

- Therefore, membrane constituent become disformed and cell lyse.

vi. Peroxide and hypochloride:

- HClO and H_2O_2 oxidizes the cellular structure.
- HClO damages the lipid bilayer and inhibit -SH group of protein.
- The oxidization of cell membrane lyses the cell to release cellular constituents.
- i. Autolysis:
- Cell membrane acting chemicals such as Toluene, antibiotics etc. activates lytic enzyme of the cell.
- The lytic enzyme causes self lysis of cell.
- The chance of contamination of product is less.
- ii. lytic enzyme
- Different enzymes such as lysozyme, cellulase, zymogenase, proteases, etc. are used to lyse the wells.
- This method is used in small scale.
- iii. Phage mediated lysis:
- T4-phage, OX174, ssRNA phage, etc. are bacteriophage.
- They penetrate and multiply inside bacterial cell and then causes lysis of cell by producing endolysin and murein hydrolase enzymes,

Refolding:

- The first step in the refolding is the dissolution of the inclusion bodies (obtained from concentration) in a strong chaotropic solution of 6M urea, 2M thiourea.
- Chaotropic agents are desaturating agent.
- Chaotropic agents disrupt the intramolecular force between water molecules and allows protein and other macromolecule to dissolve easily.
- The desaturated protein is then allowed to renature by removing the chaotropic agent by dilution, dialysis or by chromatographic separation.

Final purification:

- Chromatography is the usual method for obtaining pure protein.
- There are different types of chromatographic methods such as:
- Ion-exchange chromatography:

- In case of ion exchange chromatography, cation or anion is attached to resin beads, depending upon the electric property of proteins.
- If the desired protein is -vely charged then +ve charged resin beads are used.
- The resin beads is packed in the column.
- When the sample is poured in the column, the -vely charged protein (desired protein) stick on the beads while other undesired +vely charged protein eluted first.
- The desired protein (-ve) is obtained as elute by changing the pH of the wash buffer or by washing with high salt solution.

- Hydrophobic chromatography:
  - This chromatography was developed to purify proteins by exploiting their surface hydrophobicity.
  - Groups of hydrophobic residues are scattered over the surface of proteins in such a way that it gives characteristic property to each protein.
  - The hydrophobic groups are covered by ordered layer of water in aqueous solution.
  - When salt is added then hydrophobic groups are exposed and interact with each other.
  - In hydrophobic interaction chromatography, the column is packed with hydrophobic beads (-phenyl, -acetyl group).
  - When the sample is poured, the hydrophobic protein interacts with hydrophobic matrix.
  - The salting-out compound such as ammonium sulfate is used from high to low concentration in the column so the protein with low hydrophobicity elute first.
  - The non-ionic detergents such as tween-20, triton-x-100 etc. are used to elute the protein.

Affinity chromatography:

- In affinity chromatography, a compound having specific affinity to desired protein is attached to the resin. For. e.g. Antibody against desired protein is coated on resin.

- The resin is then packed into a column. When mixture of protein is poured, only those proteins having specific affinity with resin (Ab coated) stick on the column.
- All the other protein gets eluted.
- Only the undesired protein gets eluted.
- The protein of interest stuck on the column can be eluted by changing the ionic strength of the solution so that the desired protein no longer binds to resin and get eluted.
- This can also be achieved by adding special compound on elution solution which change the equation state and elute the protein.

- Size-exclusion chromatography:
- This process is also known as gel filtration.
- The method used to separate proteins on the basis of their size or molecular weight.
- The porous matrix is packed in the column.
- The porous matrix retards the rate of elution of proteins.
- The protein with higher molecular weight elutes first since small protein passes through pores and elute last

Protein characterization:

- The methods of protein characterization are:

i. Electrophoresis:

- It is the process of separation of charged particles under the influence of electric field.
- SDS-PAGE is the most widely used method for analysis of protein in the mixture.
- It is useful for monitoring the protein purification.
- SDS-PAGE separates proteins on the basis of molecular weight.
- At first polyacrylamide gel is made and a well is made.
- The protein sample is mixed with beta-mercaptoethane and sodium dodecyl sulphate (SDS) and boiled for 5 minutes.
- During boiling, proteins get denatured.
- Each SDS molecule binds to two amino-acids molecules of denatured protein.
- SDS molecule is highly -vely charged so the protein binds with SDS become -vely charged.
- When electrophoresis is done, the protein moves towards anode (+ve charge).

- The small size protein migrate faster and large size moves slower forming different band.
- The band can be visualized by staining with Coomassie brilliant blue (CBB).

Peptide sequencing:

- This method is developed by Pehr Edman so it is also known as Edman degradation.
- The polypeptide is reacted with phenylisothiocyanate under mild alkaline condition.
- The amino terminal of peptide is converted to phenylthiocarbomyl (PTC).
- The phenylthiocarbomyl (PTC) derivatives is washed thoroughly with organic solvent (e.g. benzene) and dried.
- The dried PTC is treated with anhydrous acid (e.g. heptafluorobutyric acid).
- This results in cleavage of PTC-polypeptide near PTC substitution releasing N-terminal aminoacid as thioazoline derivatives.
- The thioazoline derivative is stable. So, it is converted to thiohydantoin derivative containing aminoacid is identified by high performance liquid chromatography (HPLC).
- If the aminoacid is alanine then the first aminoacid is the polypeptide along N-terminal is alanine.
- The Edman degradation process is repeated for sequencing other amino-acids.

Tryptic mapping:

- Edman degradation method for determining amino-acid sequence from N-terminal require free amino-group at N-terminal of protein.
- However, 50-70% of proteins have N-terminal blocked by Formyl, acetyl or acryl group during post translational modification.
- For such protein, sequencing is not possible so, the protein have been cleaved by endopeptidases to produce peptide which is then sequenced.
- Trypsin enzyme is an example of endopeptidases which cleaves C-terminal of arginine and lysine.
- Similarly, other endopeptidases have their own restricted cleavage site.
- The generated short fragment is sequenced by Edman method followed by HPLC to identify the amino acids.

iv. Analytical ultracentrifugation:

- This method measures variety of properties of protein sample including molecular weight, interaction with other molecules and sample homogeneity.

v. Spectroscopy:

- It is used in analysis of wide range of sample.
- The metal containing protein (co-factors) can be analysed by spectroscopy.
- The different co-factor gives different electromagnetic spectrum.

vi. Biosensors:

- It is a device used for the detection of particular protein in cell.
- The device is coated with specific Ab against the desired protein.
- When sample is added, the particular protein binds the Ab producing signal on device.

vii. Mass spectroscopy:

- Mass spectroscopy is an analytical technique that provide information about molecular structure of organic and inorganic compound.
- The mass of particular protein can be determined by mass spectroscopy.
- It can detect post translational modification or any variation in structure or protein.
- Chromatography
- "Writing Color"
- *Chromatography gets its name from a technique first used in the late 19th century to separate pigments in a complex mixture.*
- If a sheet of paper or cloth contacts a container filled with water or alcohol in which a complex pigment is dissolved, capillary action will carry the mixture up the paper or cloth, but the components of the pigment will not all travel at the same rate.
- The largest molecules of the mixture will travel more slowly while the smallest ones race ahead, causing the stationary phase to develop discrete bands of color corresponding to each component of the mixture. This gives the technique the name "chromatography" or "writing color."

What is Chromatography and How Does it Work?

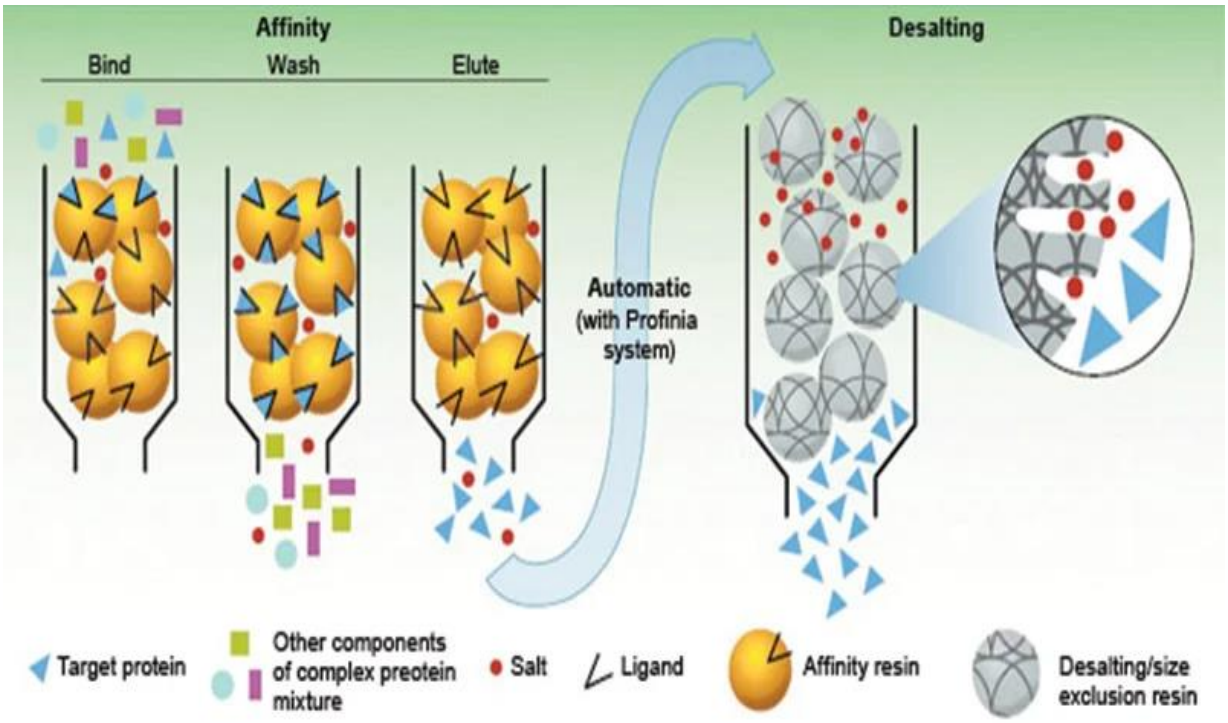
- Chromatography is a process for separating components of a mixture.
- To get the process started, the mixture is dissolved in a substance called the mobile phase, which carries it through a second substance called the stationary phase.
- The different components of the mixture travel through the stationary phase at different speeds, causing them to separate from one another.

- The *nature of the specific mobile and stationary phases* determines which substances travel more quickly or slowly, and is how they are separated. These different travel times are termed retention time.

Some of the most common forms of chromatography are as follows.

- In gas chromatography, the mixture of interest is vaporized and carried through a stationary phase (usually a metal or glass separation column) with an inert gas, usually nitrogen or helium. Larger molecules in the mixture take longer to pass through the column and reach the detector at the far end.
- In liquid chromatography, the mixture of interest is dissolved in a liquid and passed through a solid stationary phase, which is often made of a silica material. Several varieties of liquid chromatography exist, depending on the relative polarities of the mobile and stationary phases (normal-phase versus reverse-phase) and whether the mobile phase is pressurized (high-performance).
- In thin-layer chromatography (TLC), the stationary phase is a thin layer of solid material, usually silica-based, and the mobile phase is a liquid in which the mixture of interest is dissolved. Thin-layer chromatography comes with the advantage of photographing well, making its output easy to digitize.
- Ion exchange chromatography separates the components of a mixture based on their charge, in addition to or instead of their size. In essence, positively (cations) or negatively (anions) charged ions are separated using different stationary phases and different pH mobile phases.
- Chromatography
- Chromatography is an important biophysical technique that enables the separation, identification, and purification of the components of a mixture for qualitative and quantitative analysis.
- It is a separation technique in which a mobile phase carrying a mixture is caused to move in contact with a selectively absorbent stationary phase.
- Affinity chromatography is a type of liquid chromatography for the separation, purification or specific analysis of sample components.

- It utilizes the reversible biological interaction or molecular recognition called affinity which refers to the attracting forced exerted in different degrees between atoms which cause them to remain in combination.

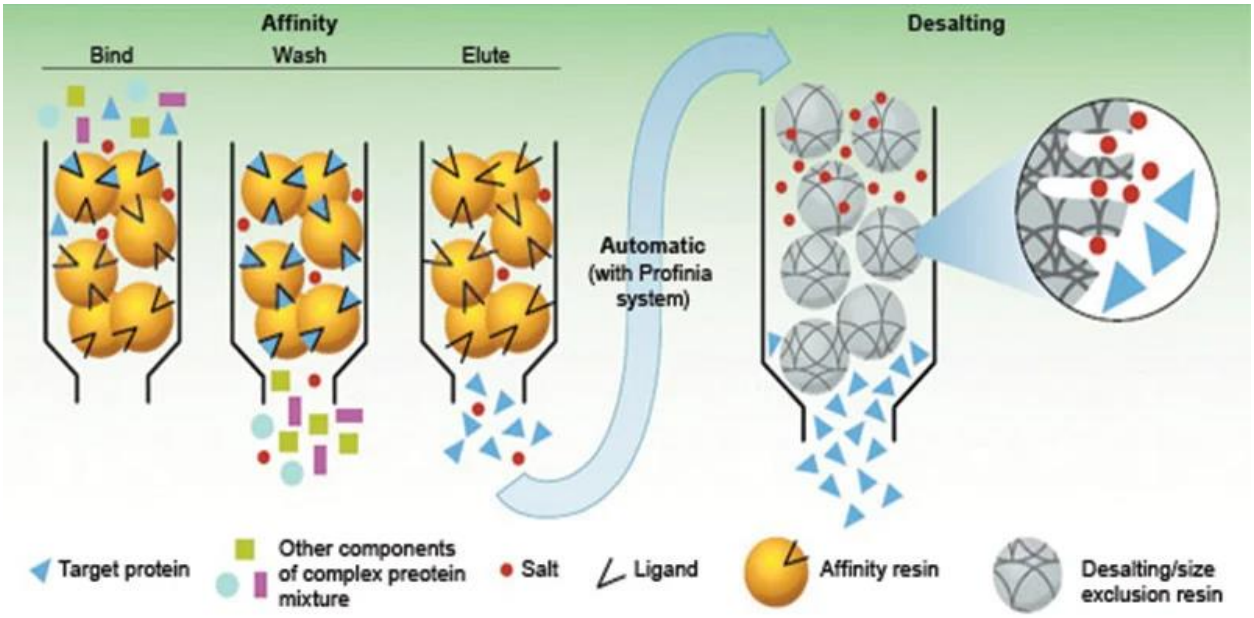


Affinity chromatography

- Affinity chromatography is a separation method based on a specific binding interaction between an immobilized ligand and its binding partner.
- Examples

1. Antibody/antigen
2. Enzyme/substrate
3. Enzyme/inhibitor interactions.

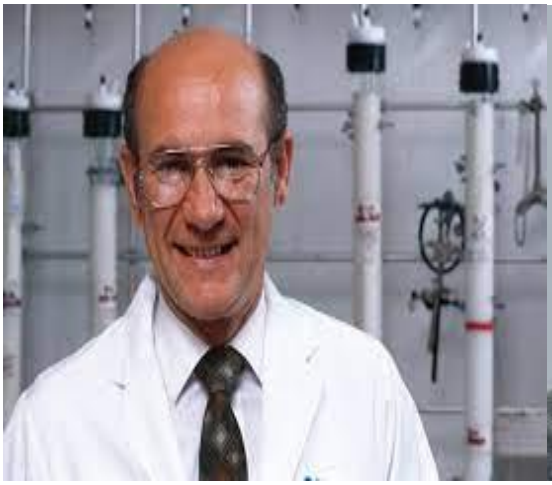
- The degree of purification can be quite high depending on the specificity of the interaction and, consequently, it is generally the first step, if not the only step, in a purification strategy.



Why Use Affinity Chromatography?

- Affinity chromatography offers *high selectivity, resolution, and capacity in most protein purification schemes*.
- It has the advantage of utilizing a protein's biological structure or function for purification.
- As a result, purifications that would otherwise be time consuming and complicated, can often be easily achieved with affinity chromatography.
- It was discovered by

Pedro Cuatrecasas and Meir Wilcheck



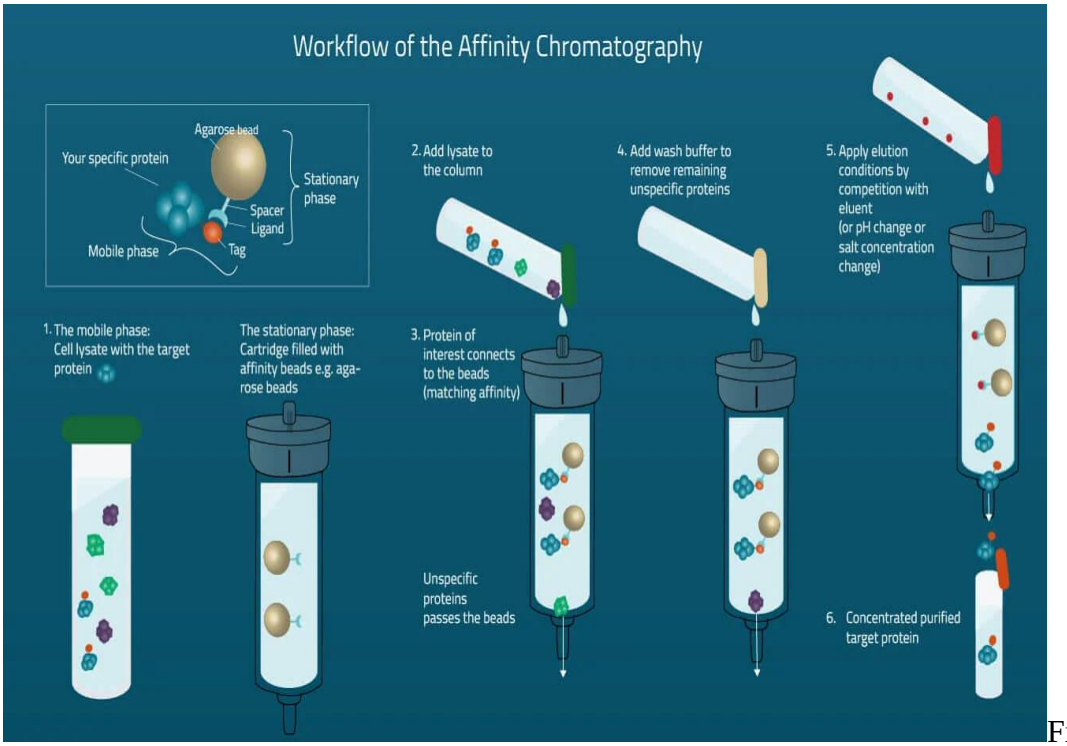


Fig. :

Workflow of an affinity chromatography using a drip-column.

1: The two phases of an affinity chromatography: The mobile and the stationary phase.

2: First step - Add cell lysate to the column.

3: The protein of interest will interact with the beads through matching affinity

4: Add wash buffer and remove remaining unspecific protein and other substances.

5: Elute your protein of interest from the affinity beads through an elution buffer. Either via pH change or salt concentration.

6: Purified & active protein of interest in collected!

Components of Affinity Chromatography

Matrix

- The matrix is an inert support to which a ligand can be directly or indirectly coupled.
- In order to for the matrix to be effective it must have certain characters:
- Matrix should be chemically and physically inert.
- It must be *insoluble in solvents and buffers employed in the process*
- It must be *chemically and mechanically stable*.
- It must be easily coupled to a ligand or spacer arm onto which the ligand can be attached.
- It must exhibit good flow properties and have a relatively large surface area for attachment.
- The most useful matrix materials are agarose and polyacrylamide.

The Core

Table 1: Possible Cores for a stationary (solid) phase

- The preferred solid phase for protein purification are porous agarose gels.
- They have crucial advantages:
- Easily dispensed to fill and "pack" columns with resin beds of any size
- Large enough that biomolecules (proteins, etc.) Can flow as freely into and through the beads
- Ligands are covalently attached to the bead polymer (external and internal surfaces) by various means
- Beaded agarose is good for routine applications as it allows easy setup, making it suitable for gravity-flow, low-speed-centrifugation, and low-pressure procedures

	Features
Agarose	• hydrophilic • almost no unspecific bonds • the gold standard for protein purification
Silica gel	• nanoporous (leads to unspecific bonds) • functionalized via silane • silanes are washed away by alkaline buffer → reduced stability

- applications: bound nucleic acids chaotropic

Aluminium oxide	• acidic surface • bounds amines irreversibly • used to reduce the amount of specific substances
Acrylate	• partially hydrophobic (unspecific bounds possible) • monodisperse particles • used for cell separation
Organic polymers	• partially hydrophobic (unspecific bounds possible) • monodisperse particles • can be used for ligand coupling • not recommended for protein purification because of unspecific bounds

The Stationary Phase

The stationary phase of an affinity chromatography, also called the solid phase, consists of the

- Core
- Spacer
- Ligand

2. Spacer arm

- It is used to improve binding between ligand and target molecule by overcoming any effects of steric hindrance.

3. Ligand

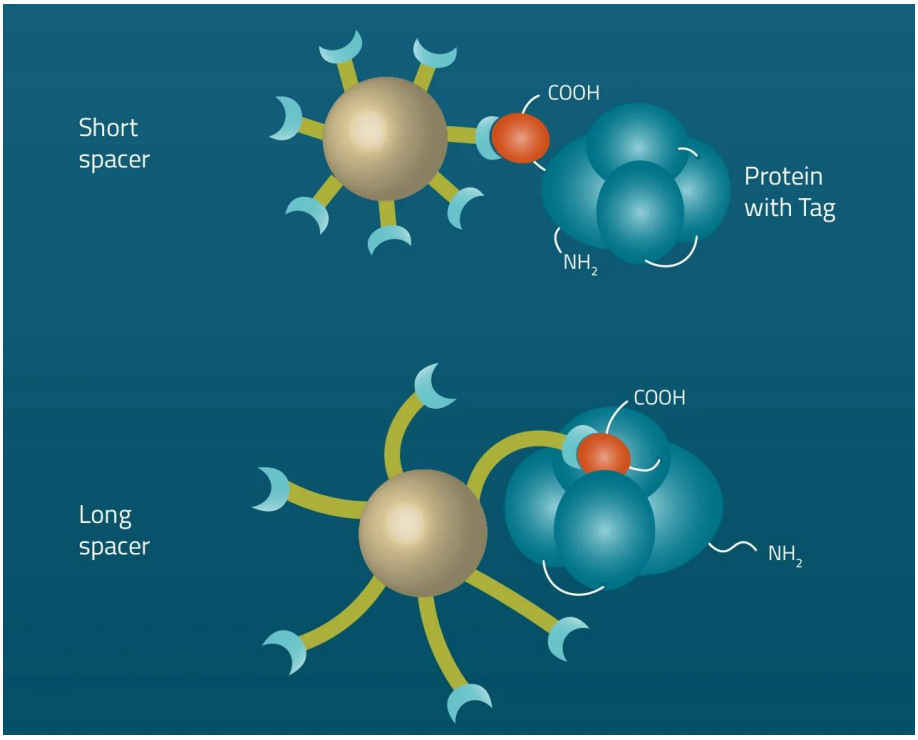
- It refers to the molecule that binds reversibly to a specific target molecule.
- The ligand can be selected only after the nature of the macromolecule to be isolated is known.
- When a hormone receptor protein is to be purified by affinity chromatography, the hormone itself is an ideal candidate for the ligand.
- For antibody isolation, an antigen or hapten may be used as ligand.

- If an enzyme is to be purified,a substrate analog, inhibitor, cofactor, or effector may be used as a the immobilized ligand.

The Spacer

The spacer of an affinity bead can vary in size and length. Longer chain lengths have a higher mobility and are thus crucial for stericaly hindered tags.

Table : Overview of different spacers for affinity beads.



Chemical	Chain length
Cyanogen bromide	C1
Epoxide	C3
Epoxide with C6 acid	C10
Diamin	C10

The Ligand

The Ligand is responsible for the affinity of the bead.

In some cases the affinity of the ligand can be further modified. A noteworthy example here are

His-tag purifications. In this case the ligand is a chelator coupled with a metal ion. The metal ion is responsible for the balance between affinity and specificity of the purification

Table : Typical ligands used in affinity chromatography

Ligand	Target
Antibody	Antigen
Iron-, aluminium-ions	Phosphoproteins
Avidin	Biotin
Glutathione	GST
Chelator + Ni-, Co-ions	His-tagged proteins

Steps in Affinity Chromatography

- Affinity medium is equilibrated in binding buffer.
- Sample is applied under conditions that favor specific binding of the target molecule(s) to a complementary binding substance (the ligand). Target substances bind specifically, but reversibly, to the ligand and unbound material washes through the column.
- Elution is performed specifically, using a competitive ligand, or non-specifically, by changing the pH, ionic strength or polarity. Target protein is collected in a purified, concentrated form.
- Affinity medium is re-equilibrated with binding buffer.

These events can be summarized into the following three major steps:

1. Preparation of Column

- The column is loaded with solid support such as sepharose, agarose, cellulose etc.
- Ligand is selected according to the desired isolate.
- Spacer arm is attached between the ligand and solid support.

2. Loading of Sample

- Solution containing a mixture of substances is poured into the elution column and allowed to run at a controlled rate.

3. Elution of Ligand-Molecule Complex

- Target substance is recovered by changing conditions to favor elution of the bound molecules.

Applications of Affinity Chromatography

- Affinity chromatography is one of the most useful methods for the separation and purification of specific products.
- It is essentially a sample purification technique, used primarily for biological molecules such as proteins.

Its major application includes:

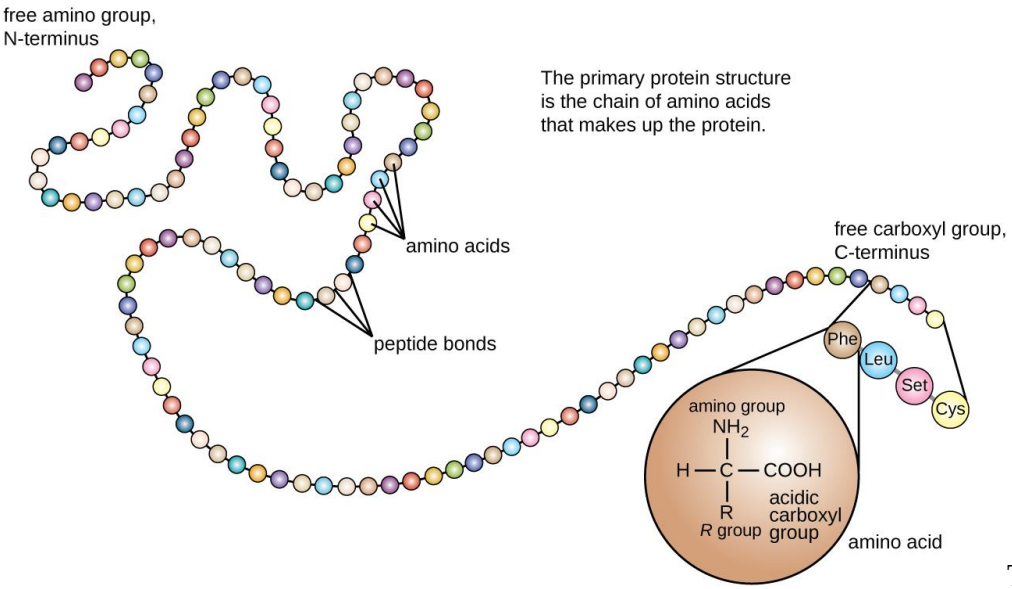
- Separation of mixture of compounds.
- Removal of impurities or in purification process.
- In enzyme assays
- Detection of substrates
- Investigation of binding sites of enzymes
- In in vitro antigen-antibody reactions
- Detection of Single Nuceotide polymorphisms and mutations in nucleic acids

Advantages of Affinity Chromatography

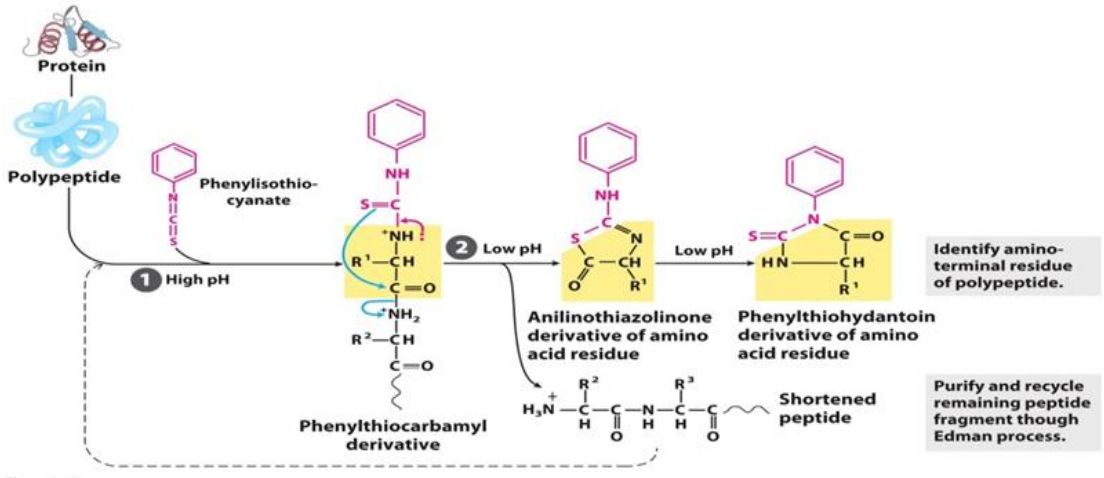
- High specificity
- Target molecules can be obtained in a highly pure state
- Single step purification
- The matrix can be reused rapidly.
- The matrix is a solid, can be easily washed and dried.
- Give purified product with high yield.
- Affinity chromatography can also be used to remove specific contaminants, such as proteases

Limitations of Affinity Chromatography

- Time consuming method.
- More amounts of solvents are required which may be expensive.
- Intense labour
- Non-specific adsorption cannot be totally eliminated, it can only be minimized.
- Limited availability and high cost of immobilized ligands.
- Proteins get denatured if required pH is not adjusted.
- Protein Sequencing - Edman Degradation



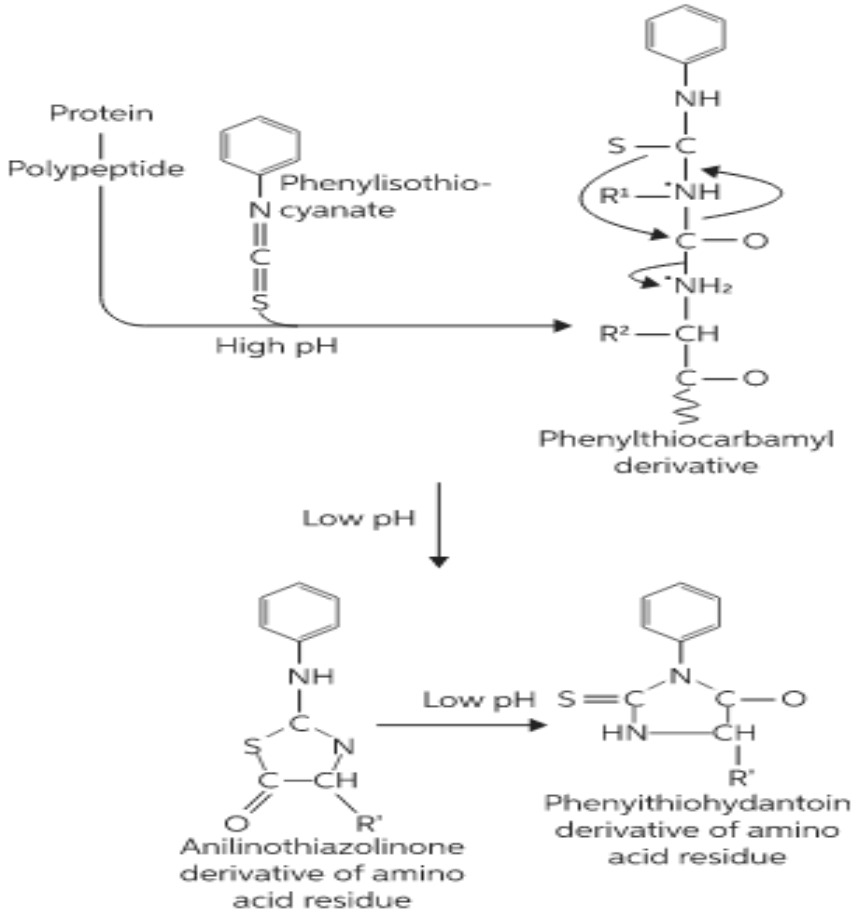
- The sequence of amino acids in a protein or peptide can be identified by Edman degradation, *which was developed by Pehr Edman.*
- This method can label and cleave the peptide from N-terminal without disrupting the peptide bonds between other amino acid residues.
- The Edman degradation reaction was automated in 1967 by Edman and Beggs.
- Nowadays, the automated Edman degradation (the protein sequenator) is used widely, and it can sequence peptides up to 50 amino acids.



Edman's Reagent

- Phenyl isothiocyanate (PITC) also known as Edman's reagent is used in Edman's degradation reaction as it can enable sequential degradation in amino acids.

- Preparation: When aniline is reacted with carbon disulfide and concentrated ammonia, the ammonium dithiocarbamate salt of aniline is obtained, which on further reaction with lead (II) nitrate gives phenyl isothiocyanate.
- PITC enables the sequential degradation of amino acids thus primary structural information in a peptide or protein can be obtained. It can react with amino acids at alkaline pH. PITC is volatile, thus it minimizes the possibility of reagent interference. PITC derivatization can be used for the identification of sulfonated amino acids, methylated, phosphorylated, and halogenated. Phenyl isothiocyanate is an aromatic isothiocyanate. It participates in dehydration reactions of alcohols. It is widely used for the synthesis of various biologically important heterocyclic compounds.
- The chemical name of Edman's reagent is isothiocyanatobenzene and its chemical formula is written as C7H5NS. Its molecular weight is 135.19 gm/mol. It is a clear liquid which is generally colorless but can even look slightly pale yellow in appearance.



- Cyclic degradation of peptides based on the reaction of phenylisothiocyanate with the free amino group of the

N-terminal residue such that amino acids are removed one at a time and identified as their phenylthiohydantoin derivatives.

- Speaking to the specific process, an uncharged peptide is reacted with phenylisothiocyanate (PITC) at the amino terminus under mildly alkaline conditions to give a phenylthiocarbamoyl derivative (PTC-peptide).
- Then, under acidic conditions, the thiocarbonyl sulfur of the derivative attacks the carbonyl carbon of the N-terminal amino acid.
- The first amino acid is cleaved as anilinothiazolinone derivative (ATZ-amino acid) and the remainder of the peptide can be isolated and subjected to the next degradation cycle.
- Once formed, this thiazolone derivative is more stable than phenylthiocarbamyl derivative.
- The ATZ amino acid is then removed by extraction with ethyl acerate and converted to a phenylthiohydantoin derivative (PTH-amino acid).
- And the chromatography can be used to identify the PTH residue generated by each cycle.
- As to the automated Edman degradation, proteins can be analyzed by applying them in-solution onto a TFA(trifluoroacetic acid) filter and then loaded onto the Edman sequencing instrument.
- Proteins in mixtures are first separated by 1D or 2D gels and then blotted onto a PVDF(Polyvinylidene fluoride) membrane.
- The proteins are detected by Coomassie blue, Amido black or Poncau S staining and the proteins of interest cut out and the PVDF membrane piece loaded onto the Edman sequencer.

Analytical Procedures-Automated

- In 1967, Edman developed an automated processing device for Edman degradation in a liquid phase2).
- In the 1980s, the first Japanese gas-phase protein sequencer PSQ-1 was developed to automatically perform Edman degradation in the gas phase and analyze amino acids via isocratic elution with high-speed liquid chromatography (HPLC)3).
- Subsequently, the PPSQ (protein sequencer) series have evolved through the development of analyzing unit and data processing software.

- The advantages of Edman degradation sequencing
- The Edman degradation method has been widely used as the gold standard for N-terminal sequence testing of existing protein samples. It is a valuable research tool for N-terminal sequence analysis of the entire purified protein and the most reliable sequencing method.
- The disadvantages of Edman degradation sequencing
- Edman degradation method is subject to many restrictions, such as the protein or peptide used for sequence analysis must be of high purity, and is not suitable for high-throughput analysis, and the sensitivity is not enough.

Applications:

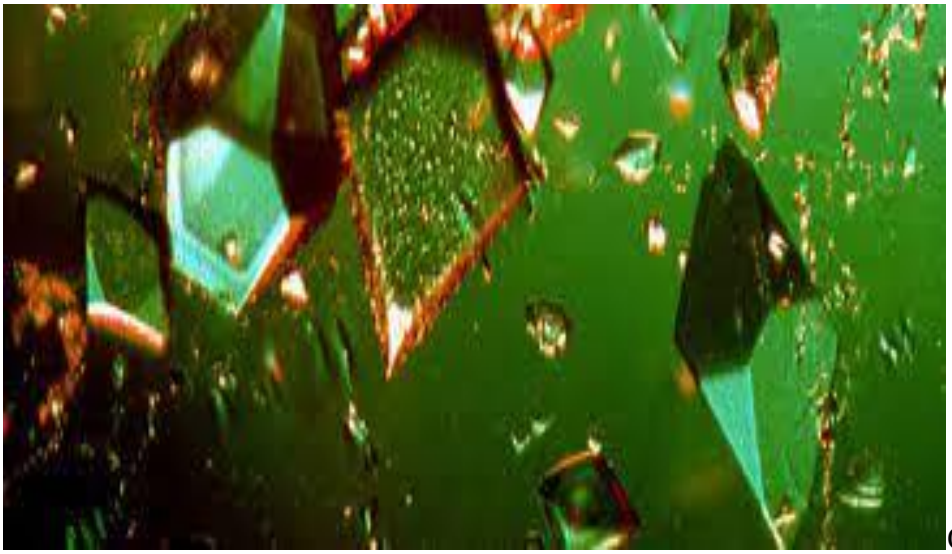
- By using the Edman degradation reaction, the proteolytic cleavage sites and the newly formed N-terminal in protein fragments can be identified.
- Edman degradation can also provide help with new and unknown peptides and proteins, whose sequences are not known and cannot be found in the sequences database.
- It can precisely produce up to 30 amino alkanoic acids with present-day machines fit for over 99% productivity per amino acid.
- Just 10 - 100 pico-moles of peptides are required for the sequencing procedure.
- When the protein or domain is >40 to 80 kDa or cannot be readily purified, the recombinant proteins' N-terminal boundary or the protease-resistant domains' N-terminal can still be verified by using Edman degradation.

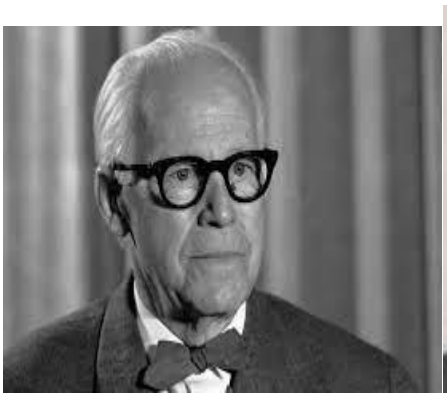
Limitations:

- Proteins with High molecular weight are difficult to be analyzed (over 75 kDa).
- Blank cycles are obtained for unmodified Cys and glycosylated residues.
- For Edman's degradation to work effectively, there should be no blocked N-terminal amino groups.
- Even slightly impure samples will result in poor results, hence sample quality becomes critical.
- Any chemical modification in the N-terminal of protein will result in the failure of Edman's degradation.
- In presence of a non-α-amino acid, the 5-membered ring intermediate cannot be produced.
- It also requires a lot of guesswork in the overall synthesis.

- Disulfide bridges cannot be determined.
- There should at least be 1 picomole of peptide for obtaining positive results.

Protein Structure – Basics of X ray Crystallography

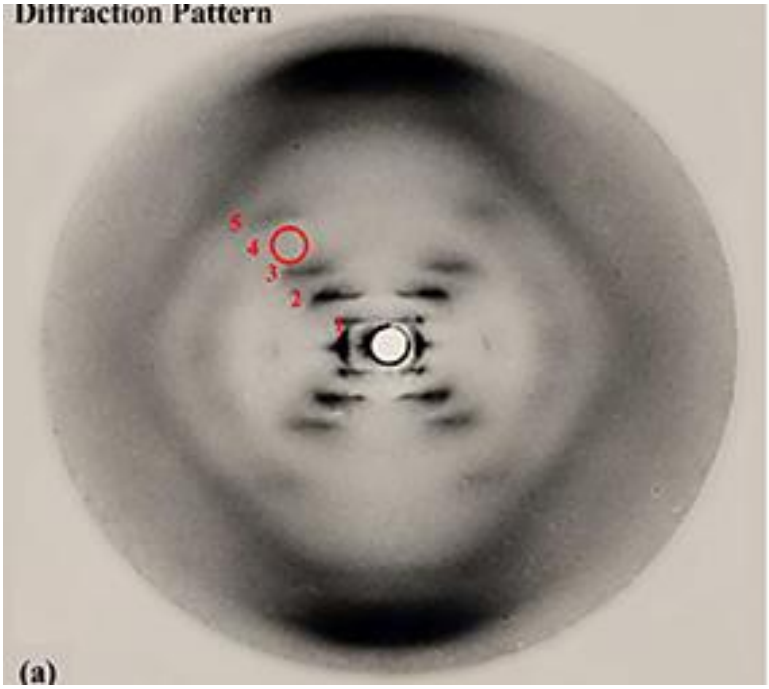
- *So, what is crystallography? Put simply, it is the study of crystals.*
- *This is a serious scientific subject, with around 26 or so Nobel prizes to its credit. And yet, despite this, it remains a largely hidden subject, at least in the public mind.*
-  Crystallography as a science has a long and venerable history going back to the 17th century when the sheer beauty of the symmetry of crystals suggested an underlying order of some kind.
- For the next three centuries, our knowledge of what crystals actually were was based on conjecture and argument, with a few simple experiments thrown in.
- From their symmetry and shapes it was argued that crystals must consist of ordered arrangements of minute particles: today we know them as atoms and molecules.
- But it was the discovery of X-rays in 1895 that changed all that, for a few years later in 1912 in Germany, Max Laue, Walter Friedrich, and Paul Knipping showed that an X-ray beam incident on a crystal was scattered to form a regular pattern of spots on a film (we call this diffraction).
- Thus it was proved that X-rays consisted of waves and furthermore this gave direct evidence of the underlying order of atoms in the crystal.
- Hence Nobel Prize number 1 went to Laue in 1914.

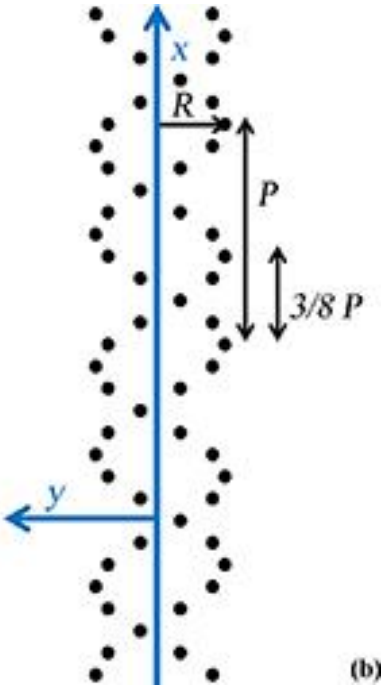


- However, it was William Lawrence Bragg (WLB) who in 1912 at the age of 22 showed how the observed diffraction pattern could be used to determine the positions of atoms in the crystal, thus launching a completely new scientific discipline, X-ray crystallography.
- Working with his father, William Henry Bragg (WHB), they quickly determined the crystal structures of several materials starting with that of common salt and diamond.
- Both father and son shared Nobel prize number 2 in 1915. William Henry Bragg and William Lawrence Bragg went on to create world-class research groups working on a huge range of solid materials and incidentally they were active in encouraging women into science.

- 

  Since then X-ray crystallography, which today is used throughout the world, has been the method of choice for determining the crystal structures of organic and inorganic solids, pharmaceuticals, biological substances such as proteins and viruses, and indeed all kinds of solid substances.
- Crick and Watson's determination of the double helix of DNA is probably the most well-known example of the use of crystallography, incidentally a discovery made in William Lawrence Bragg's laboratory in Cambridge.
- Had it not been for X-ray (and later neutron and electron) crystallography we probably would not have today much of an electronics industry, computer technology, new pharmaceuticals, new materials of all sorts, nor the modern field of genetics.
- The Braggs left a huge legacy which today continues to make astonishing progress.



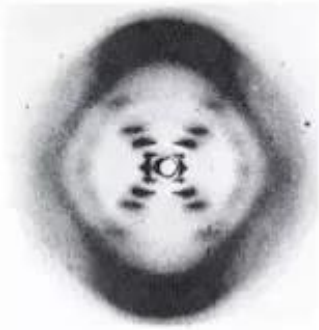


7.1 A.1 Rosalind Franklin's and Maurice Wilkins' investigation of DNA structure by X-ray diffraction

Rosalind Franklin (1950's)

- Worked with Maurice Wilkins
- X-ray crystallography = images of DNA
- Provided measurements on chemistry of DNA

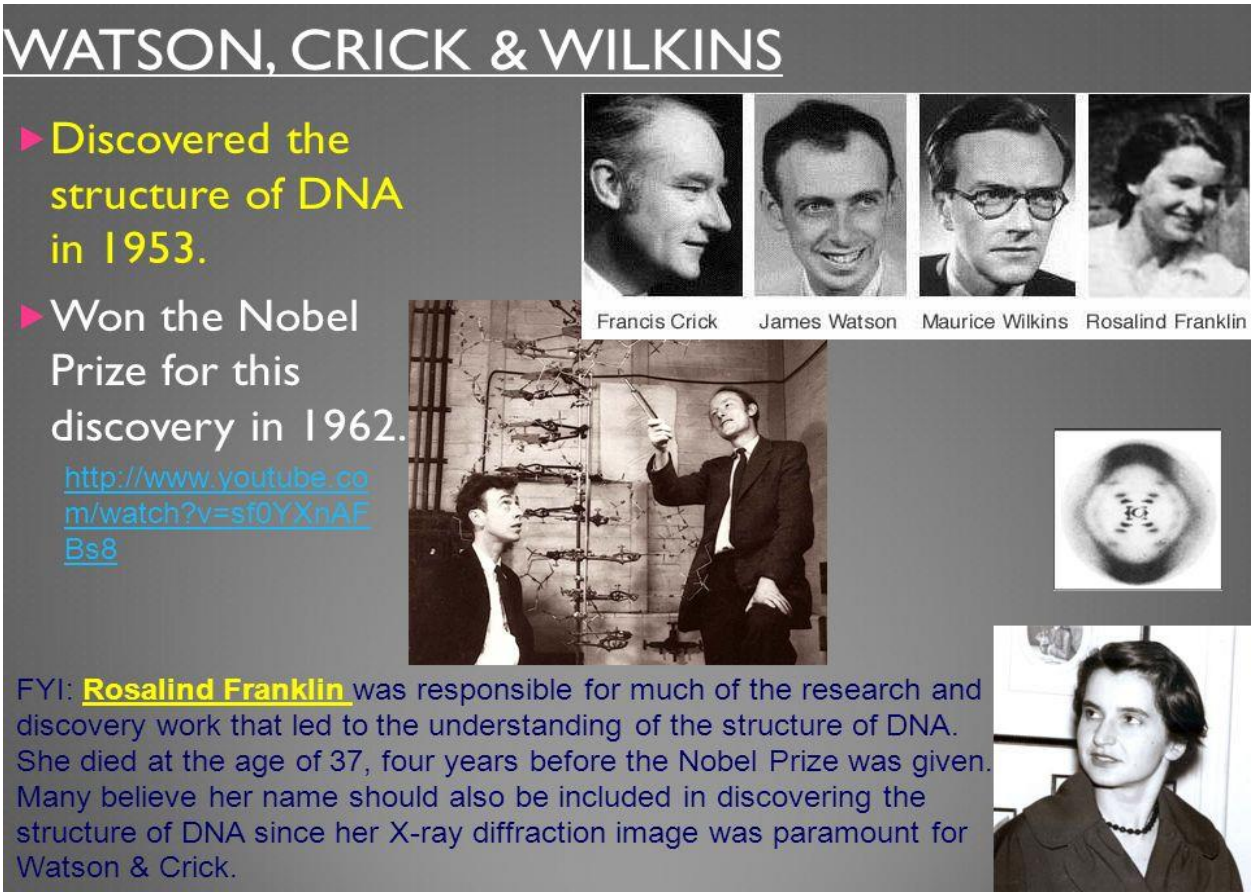
(a) Rosalind Franklin



(b) Franklin's X-ray diffraction photograph of DNA

James Watson & Francis Crick (1953)

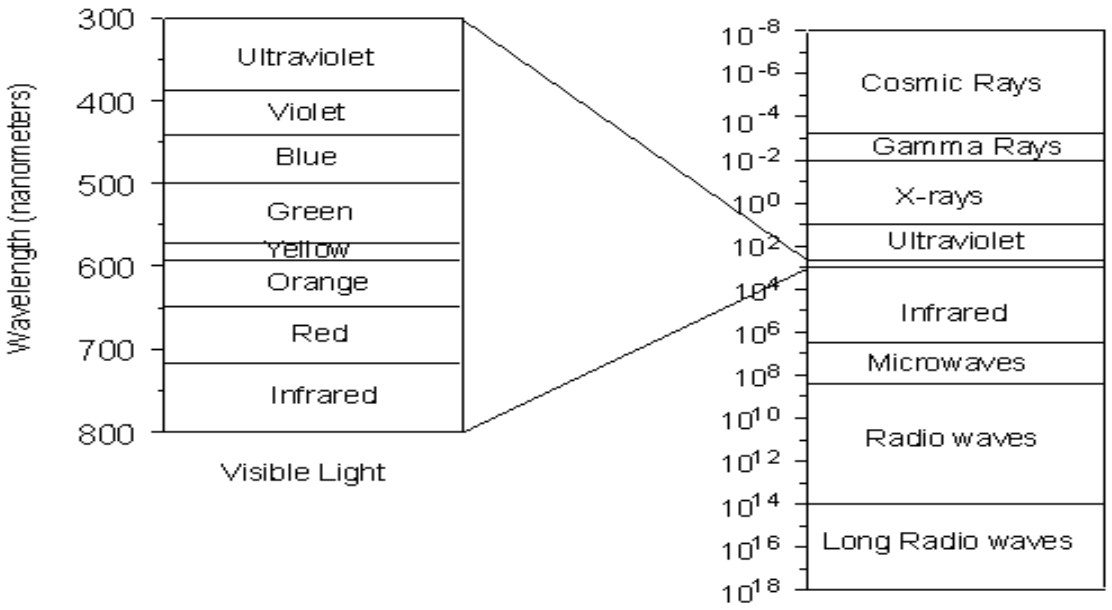
- Discovered the double helix by building models to conform to Franklin's X-ray data and Chargaff's Rules.



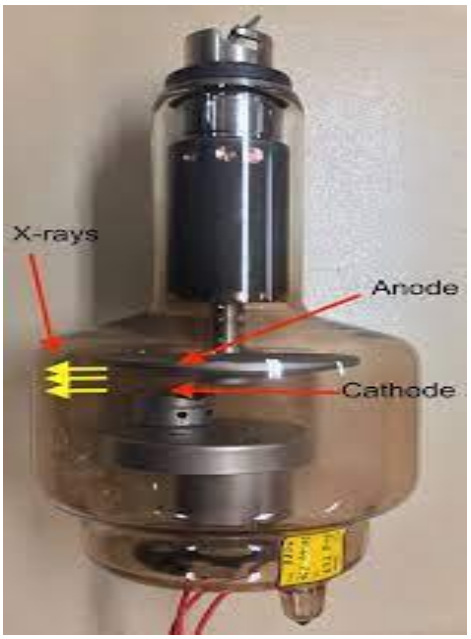
X-rays and the Production of X-rays

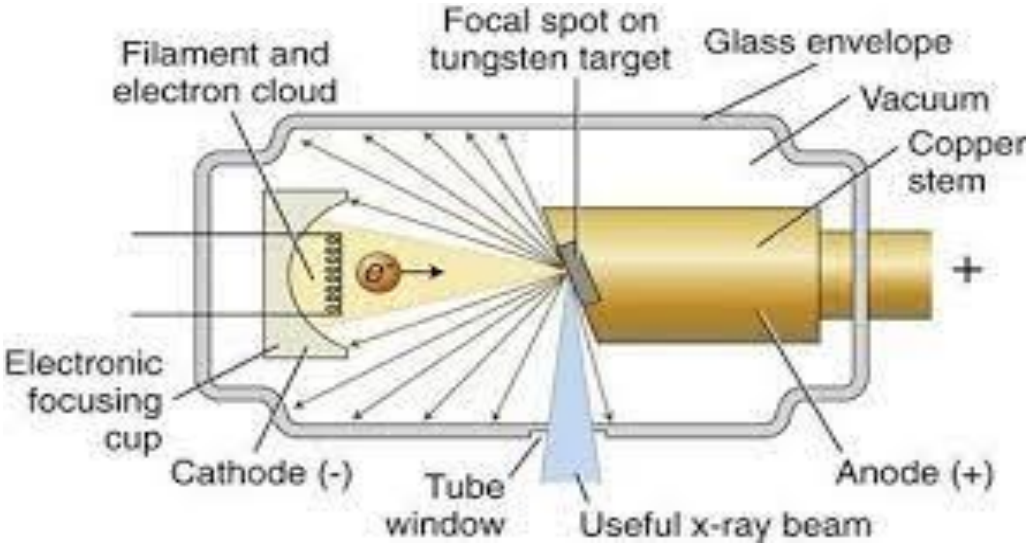
- X-rays are electromagnetic radiation with wavelengths between about 0.02 Å and 100 Å (1Å = 10^{-10} meters).
- They are part of the electromagnetic spectrum that includes wavelengths of electromagnetic radiation called visible light which our eyes are sensitive to (different wavelengths of visible light appear to us as different colors).

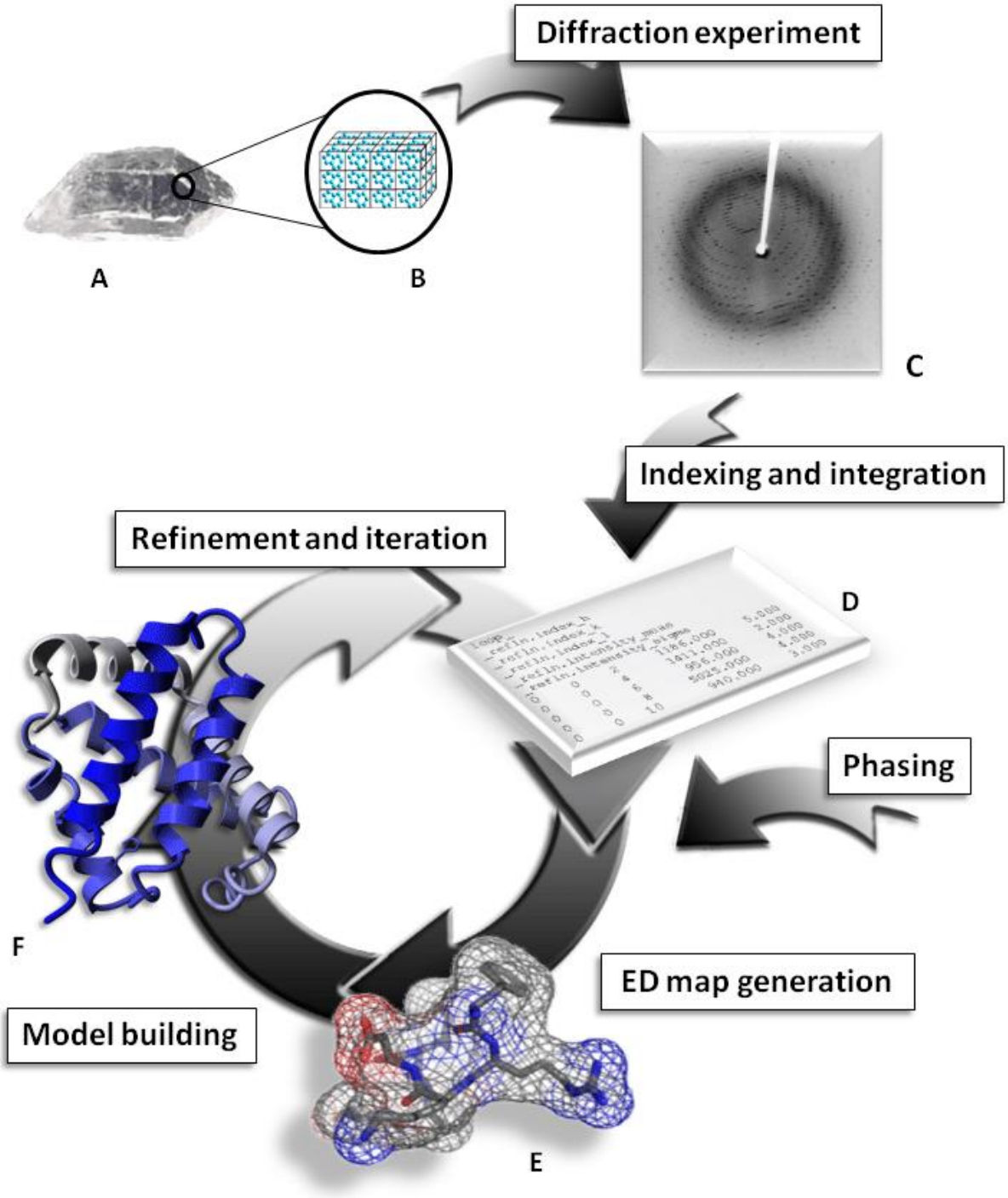
- Because X-rays have wavelengths similar to the size of atoms, they are useful to explore within crystals.



- Thus, since X-rays have a smaller wavelength than visible light, they have higher energy.
- With their higher energy, X-rays can penetrate matter more easily than can visible light.
- Their ability to penetrate matter depends on the density of the matter, and thus X-rays provide a powerful tool in medicine for mapping internal structures of the human body (bones have higher density than tissue, and thus are harder for X-rays to penetrate, fractures in bones have a different density than the bone, thus fractures can be seen in X-ray pictures).
- X-rays are produced in a device called an *X-ray tube*
- It consists of an evacuated chamber with a tungsten filament at one end of the tube, called the cathode, and a metal target at the other end, called an anode.
- Electrical current is run through the tungsten filament, causing it to glow and emit electrons.
- A large voltage difference (measured in kilovolts) is placed between the cathode and the anode, causing the electrons to move at high velocity from the filament to the anode target.
- Upon striking the atoms in the target, the electrons dislodge inner shell electrons resulting in outer shell electrons having to jump to a lower energy shell to replace the dislodged electrons. These electronic transitions results in the generation of X-rays.

- The X-rays then move through a window in the X-ray tube and can be used to provide information on the internal arrangement of atoms in crystals or the structure of internal body parts.
- 

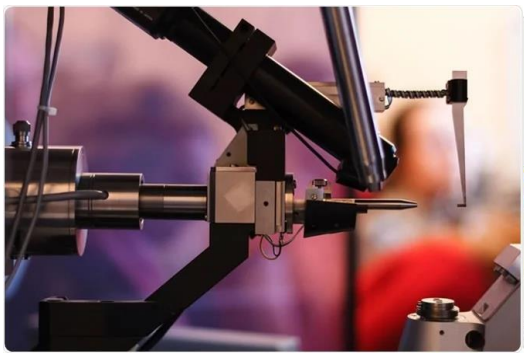


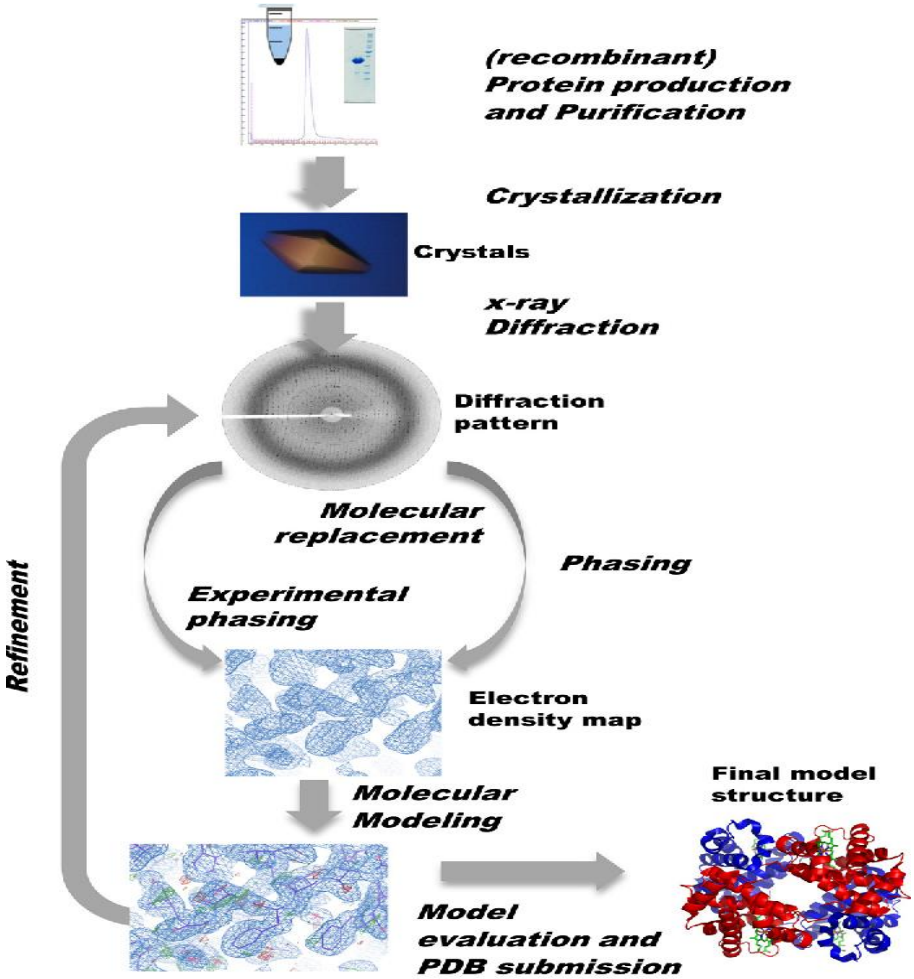


The process of tertiary structure determination consists of an X-ray diffraction experiment on a protein crystal (A) which is a 3-dimensional grid of stacked proteins.

- The Xray beam is diffracted into smaller X-rays of varying intensity which are recorded as spots on an X-ray film or a digital sensor (C).

- The spots (reflections) are processed to a computer file with indices and intensities for each reflection (D).
- Together with phases from an outside source (either another diffraction experiment or a molecular replacement calculation) this data is converted to an electron density map (ED map; E) which is a 3-dimension representation of the content of the protein crystal.
- A tertiary structure model (F) is built inside this ED map. The initial model is usually suboptimal, so the model has to be refined.
- The cycle of ED map generation, model building, and refinement is iterated until the model is complete and the tertiary structure can be considered 'solved'.



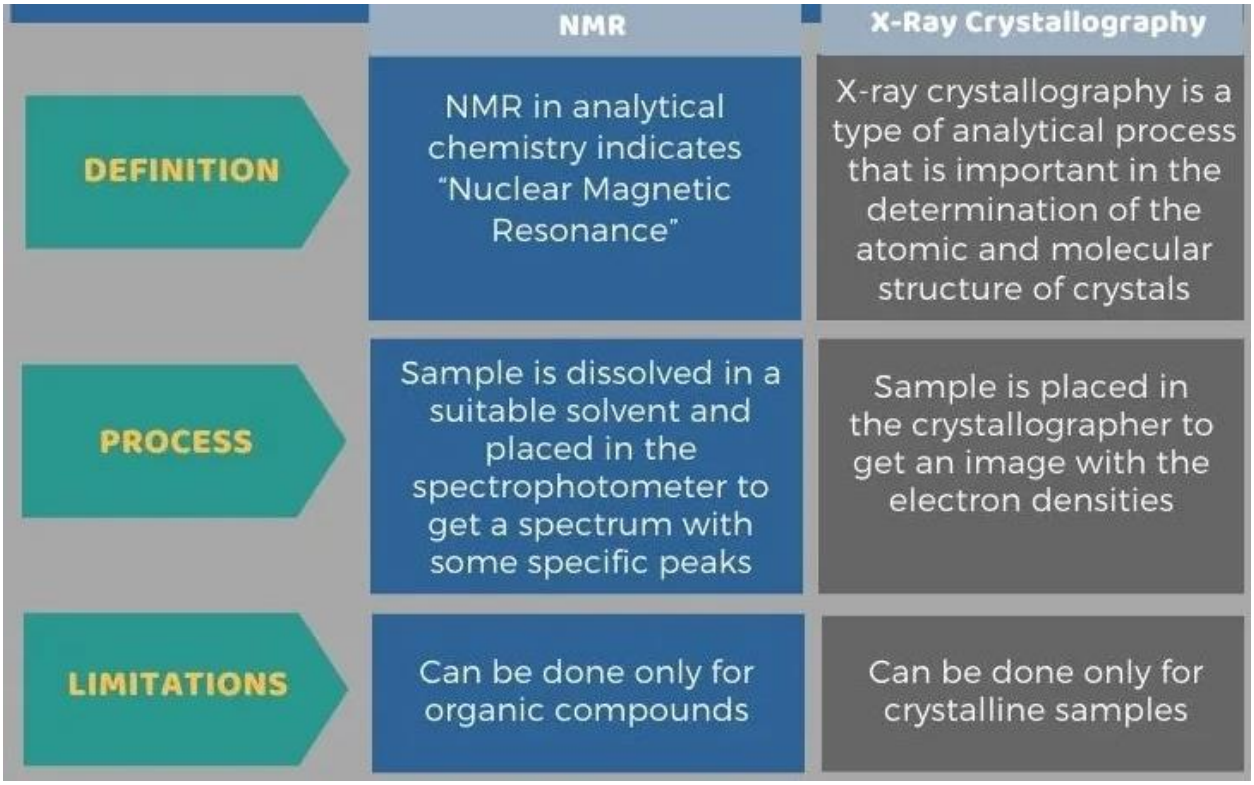

(recombinant)
Protein production
and Purification
Crystallization
Crystals
x-ray
Diffraction
Diffraction
pattern
Molecular
replacement
Phasing
Experimental
phasing
Refinement
Electron
density map
Molecular
Modeling
Final model
structure
Model
evaluation and
PDB submission

Advantages of X-ray crystallography include:

1. X-Ray crystallography provides a two-dimensional view that gives an indication of the three-dimensional structure of a material
2. Relatively cheap and simple
3. Useful for large structures: Not limited by size or atomic weight.
4. Can yield high atomic resolution.

Advantages of NMR spectroscopy include:

1. Dynamic technique
2. Non-destructive and non-invasive
3. Three-dimensional structures in their natural state can be measured directly in solution
4. Can provide unique insights into dynamics and intramolecular interactions.
5. Macromolecular three-dimensional structure resolution can be as low as sub nanometer.



Disadvantages of X-ray crystallography include:

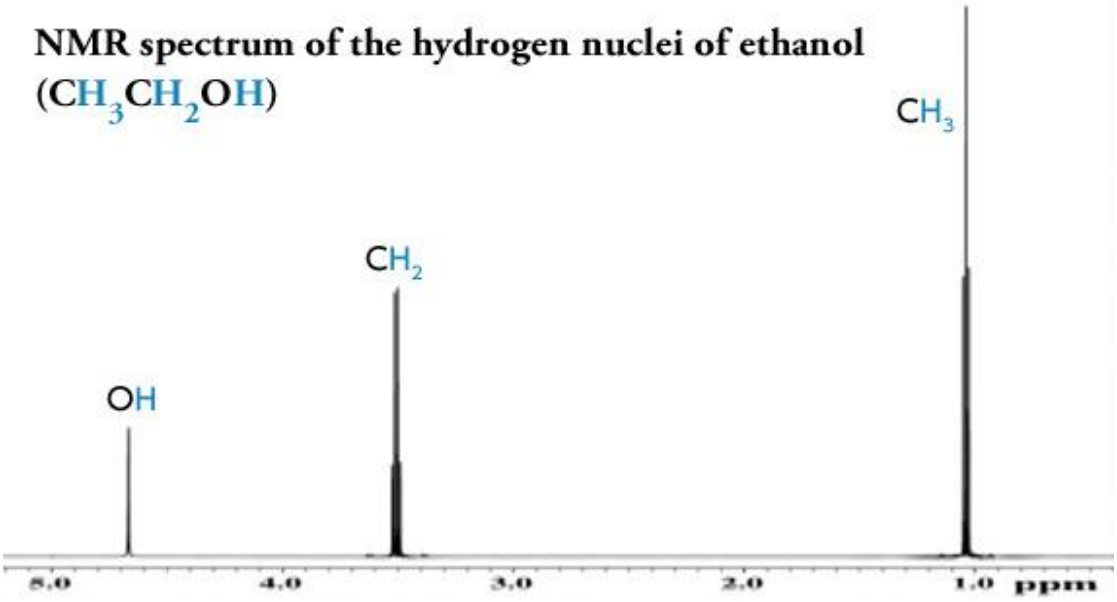
1. The sample must be crystallizable
2. The types of sample that can be analyzed are limited. In particular, membrane proteins and large molecules are difficult to crystallize, due to their large molecular weight and relatively poor solubility
3. An organized single crystal must be obtained to produce the desired diffraction
4. Non-dynamic method due to preparation of samples and crystallization. Only a static three-dimensional analysis is produced
5. X-Ray crystallography has limited applications for studies of biological samples due to the aforementioned issues.

Disadvantages of NMR Spectroscopy include:

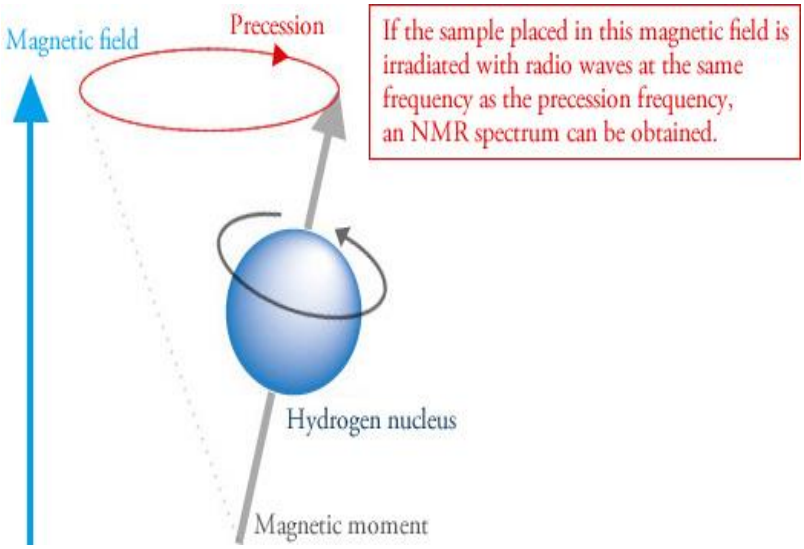
1. The application of NMR in large biomolecule analysis is limited by the complication and difficulty of interpretation of biomolecules with large molecular weight
2. Large amounts of pure samples are needed to achieve an acceptable signal to noise level
3. Highly sensitive to motion. This can lead to signal distortions in artifacts
4. The high-magnetic field can cause problems with other equipment in a laboratory. Therefore, extra precautions may need to be taken, especially if working space is limited.

5. Protein Structure – Basics of NMR
6. NMR is an abbreviation for Nuclear Magnetic Resonance.
7. An NMR instrument allows the molecular structure of a material to be analyzed by observing and measuring the interaction of nuclear spins when placed in a powerful magnetic field.

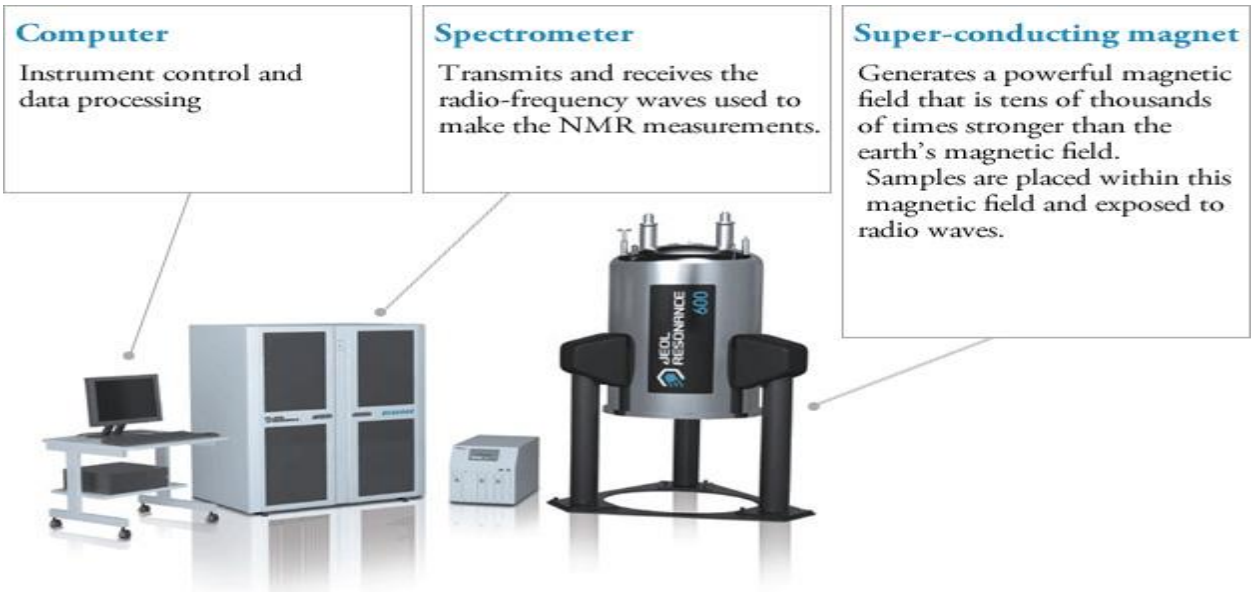
Nuclear Magnetic Resonance (NMR) spectroscopy is an analytical chemistry technique used in quality control and research for determining the content and purity of a sample as well as its molecular structure.

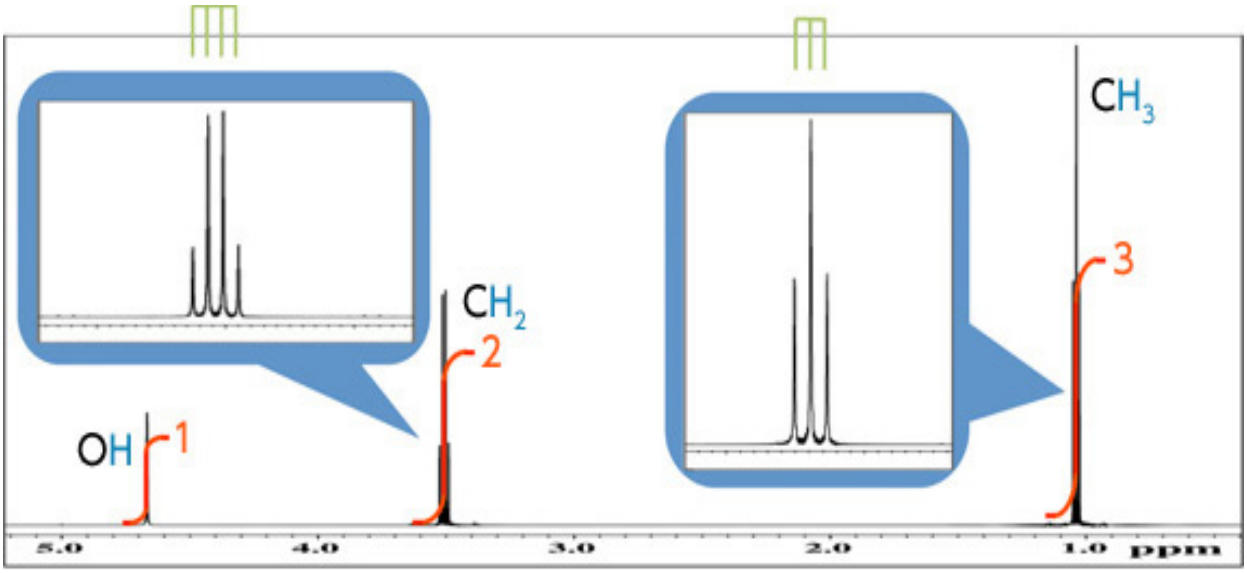
- For example, NMR can quantitatively analyze mixtures containing known compounds.
- For unknown compounds, NMR can either be used to match against spectral libraries or to infer the basic structure directly.
- Once the basic structure is known, NMR can be used to determine molecular conformation in solution as well as studying physical properties at the molecular level such as conformational exchange, phase changes, solubility, and diffusion.
- In order to achieve the desired results, a variety of NMR techniques are available.
- Principles of nuclear magnetic resonance (NMR)
- 

-
-
-



-
-
- The principle behind NMR is that *many nuclei have spin and all nuclei are electrically charged*.
- If an external magnetic field is applied, an energy transfer is possible between the base energy to a higher energy level (generally a single energy gap).
- The energy transfer takes place at a wavelength that corresponds to radio frequencies and when the spin returns to its base level, energy is emitted at the same frequency.
- The signal that matches this transfer is measured in many ways and processed in order to yield an NMR spectrum for the nucleus concerned.
- NMR instrument composition





NMR spectrum of the hydrogen nuclei of ethanol (CH_3CH_2OH)

What we can learn from NMR spectra

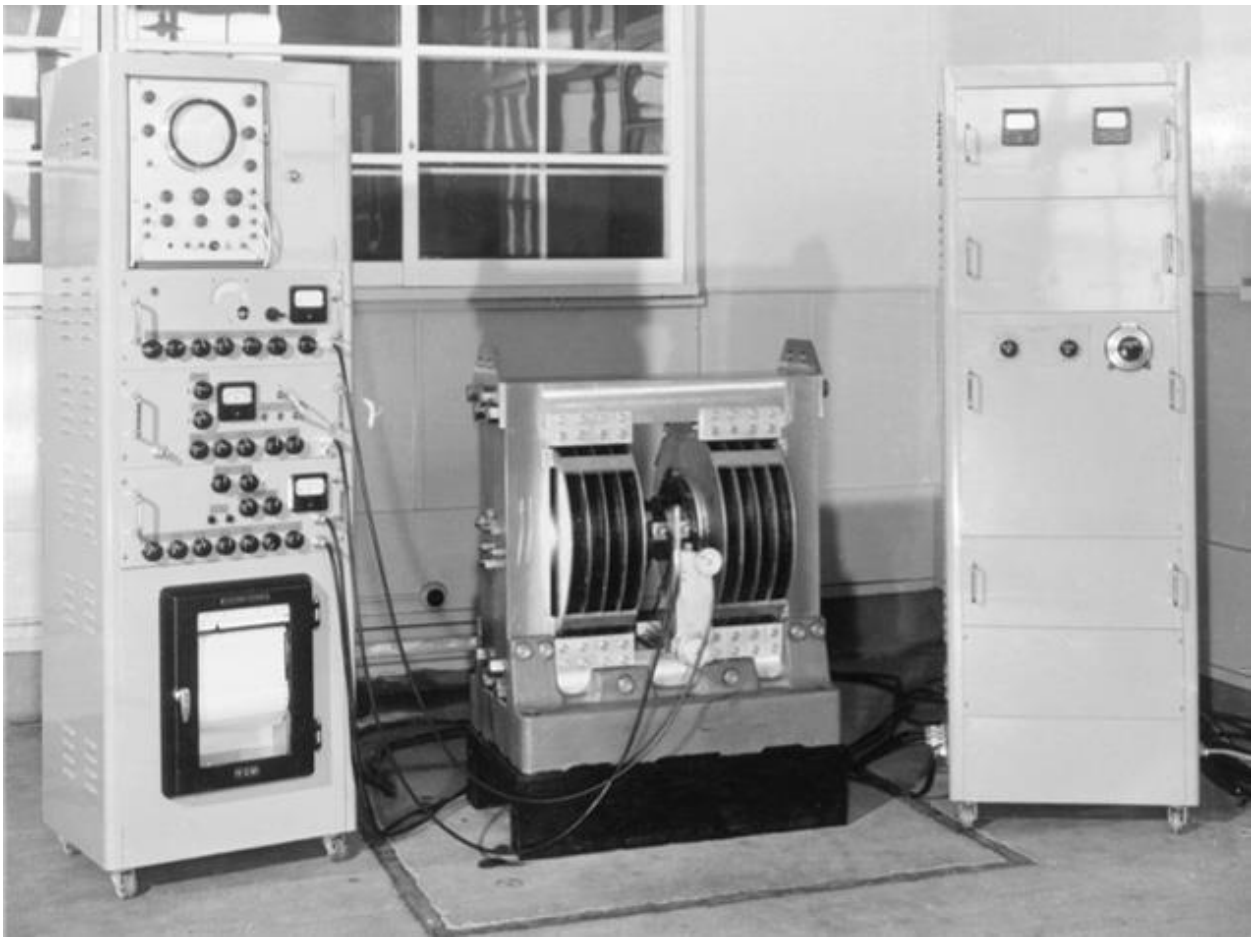
Chemical shift:

Information about the composition of atomic groups within the molecule.

- Spin-Spin coupling constant: Information about adjacent atoms.
- Relaxation time: Information on molecular dynamics.
- Signal intensity: Quantitative information, e.g. atomic ratios within a molecule that can be helpful in determining the molecular structure, and proportions of different compounds in a mixture.
- History of NMR
- In 1938, I. I. Rabi of Columbia University successfully made accurate measurements of nuclear magnetic moments using magnetic resonance absorption of molecular beams (and was awarded the Nobel Prize in Physics in 1944 for his work). In 1946, F. Block and E. M. Purcell successfully demonstrated nuclear magnetic resonance (NMR) for condensed matter (and shared the Nobel Prize in Physics in 1952), marking the start of NMR.
- In 1950, it was discovered that there were slight changes in the atomic nucleus Larmor frequencies due to the chemical bonding state of the atoms (chemical shift and spin coupling). This led to the idea of using nuclear magnetic resonance as a means to analyze and identify materials, marking the beginning of NMR spectroscopy.

History of JEOL RESONANCE NMR

- In 1956 the first JNM-1 NMR instrument was released by JEOL (at that time, Japan Electron Optics Laboratory Co. Ltd.).
- 1H resonance freuqency of the CW NMR system was abe to choose among 4, 12 and 32 MHz.
- A spectrum was displayed on an oscilloscope, and a strip chart recorder (recording onto paper tape) was used to print the results.
- Three separated 1H signals of ethyl alcohol were observed at 32 MHz (7,680 gauss).



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOINFORMATICS

UNIT – 3- SBIA1303 – Proteomics and Interactomics

Unit III Interaction Proteomics

Interaction Proteomics: Protein - Protein Interactions – Characteristics and methods to study – Yeast two hybrid analysis, Phage display. Protein DNA interactions – DNA binding motifs

Unit III Interaction Proteomics

Interaction Proteomics: Protein - Protein Interactions – Characteristics and methods to study – Yeast two hybrid analysis, Phage display. Protein DNA interactions – DNA binding motifs

- Proteins are the basic building blocks of life that are made by amino acids. The amino acids are coded by genes and form the peptides, peptides further form various proteins, and the proteins form the living tissues.
-
- *Besides, proteins also have a central role in biological processes such as catalyze reactions, transport molecules, immune reactions to the various pathogens, and signal transduction between cells.*
- What is more, the critical biological processes in the cells that directly associate with our health like DNA replication, transcription, translation, and transmembrane signal transduction all rely on the functional specific proteins. The aforementioned biological activities are regulated through protein complexes, which are typically controlled via protein–protein interactions (PPIs).
- PPIs in cells form a complicated network which has a term named "interactome".
- The interactome has a significant role in physiological and pathological processes, including signal transduction, cell proliferation, growth, differentiation, and apoptosis, etc.
- Therefore, the aberrant PPIs are associated with many human diseases such as cancer, infectious diseases, and neurodegenerative diseases.
- Since the classic drug targets are usually enzymes, ion channels, or receptors, the PPIs indicate new potential therapeutic targets.

PPIs are crucial to the formation of macromolecular structures and enzymatic complexes that form the basis of nearly every cellular process ranging from signal transduction and cellular transport to catalyzing metabolic reactions, activating or inhibiting other proteins and biomolecular synthesis.

In recent years, the PPIs have received increasing attention and became attractive targets. Recent studies indicate that the PPIs have great potential as an intervention target for novel treatment of refractory diseases, and its regulation is widely regarded as a promising strategy in drug discovery.

Table 1 Summary of some PPI modulators in clinical trials

From: Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials

PPI	Related disease	Drug	Developer	Status	NCT number	Refs.
Small molecules						
MDM2/p53	Acute myeloid leukemia	Idasanutlin	Roche	Phase III	NCT02545283	245
MDM2/p53	Metastatic melanoma	AMG232	Amgen	Phase I/II	NCT02110355	246
MDM2/p53	Solid tumor with p53 wild type status	CGM097	Novartis	Phase I	NCT01760525	247
MDM2/p53	Advanced solid tumor, lymphoma	DS-3032b	Daiichi Sankyo	Phase I	NCT01877382	248
MDM2/p53	Neoplasm malignant	SAR405838	Sanofi	Phase I	NCT01636479	249
Bcl-2/Bax	Chronic lymphocytic leukemia	ABT-199	AbbVie	Approved in 2016	–	250
XIAP/caspase-9	Relapsed or refractory multiple myeloma	LCL-161	Novartis	Phase II	NCT01955434	251
XIAP/caspase-9	Relapsed or refractory multiple myeloma	LCL-161	Novartis	Phase II	NCT01955434	251
XIAP/caspase-9	Recurrent head and neck squamous cell carcinoma	TL32711	National Cancer Institute	Phase I	NCT03803774	252
XIAP/caspase-9	Solid tumors, lymphoma	ASTX-660	Astex	Phase I/II	NCT02503423	253
XIAP/caspase-9	Solid cancers	GDC-0917	Genentech	Phase I	NCT01226277	254
PD-1/PD-L1	Prostatic neoplasms	CA-170	Astellas	Phase II	NCT01288911	255
Gp120/CCR5	HIV	Maraviroc	Pfizer	Approved in 2007	–	256
LFA-1/ICAM-1	Dry eye	Lifitegrast	Lifelong Vision Foundation	Phase IV	NCT03451396	257
B-catenin/CBP	Liver cirrhosis	RPI-724	Komagome Hospital	Phase I/II	NCT03620474	258
Bromodomain/histone	Cardiovascular diseases	RVX-208	Resverlogix	Phase III	NCT02586155	259
Bromodomain/histone	NUT midline carcinoma	GSK525762	GSK	Phase I	NCT01587703	260
Peptides						
MDM2/p53	Advanced solid tumors, lymphomas	ALRN-6924	Aileron	Phase I/II	NCT02264613	261

Advances in Protein Chemistry and Structural Biology
Volume 111, 2018, Pages 243-261

Chapter Eight - Investigating the Influence of Hotspot Mutations in Protein–Protein Interaction of IDH1 Homodimer Protein: A Computational Approach

Protein–protein interaction plays a crucial role in the biological system. In some cases,

the disruption of PPI due to the mutation in conserved amino acids positioned at the interface might lead to both structural and functional loss. In the case of homodimer IDH1, the position 132 which is found to be conserved amino acid at the interface of PPI is prone to five mutations (R132C, R132G, R132H, R132L, and R132S). In this study, we conclude that mutants R132H and R132L disrupt the PPI in higher range when compared to the other three (R132S, R132G and R132C) mutations and contribute to the destabilization of the protein. We believe that this type of computational approach could highly help in understanding the mutations in the complex PPI of the biological system.

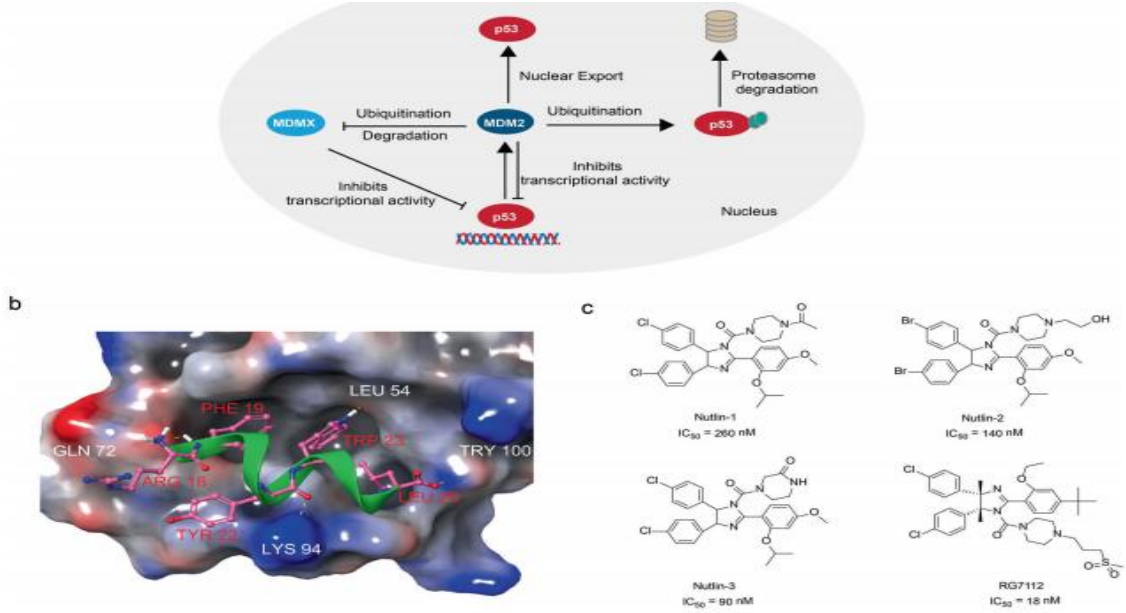
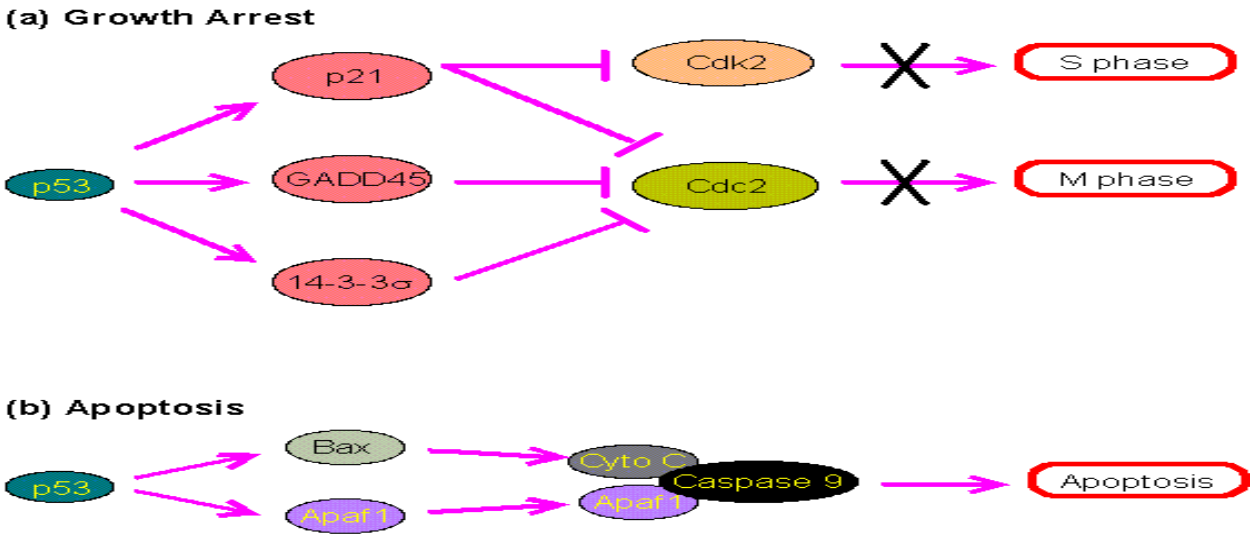


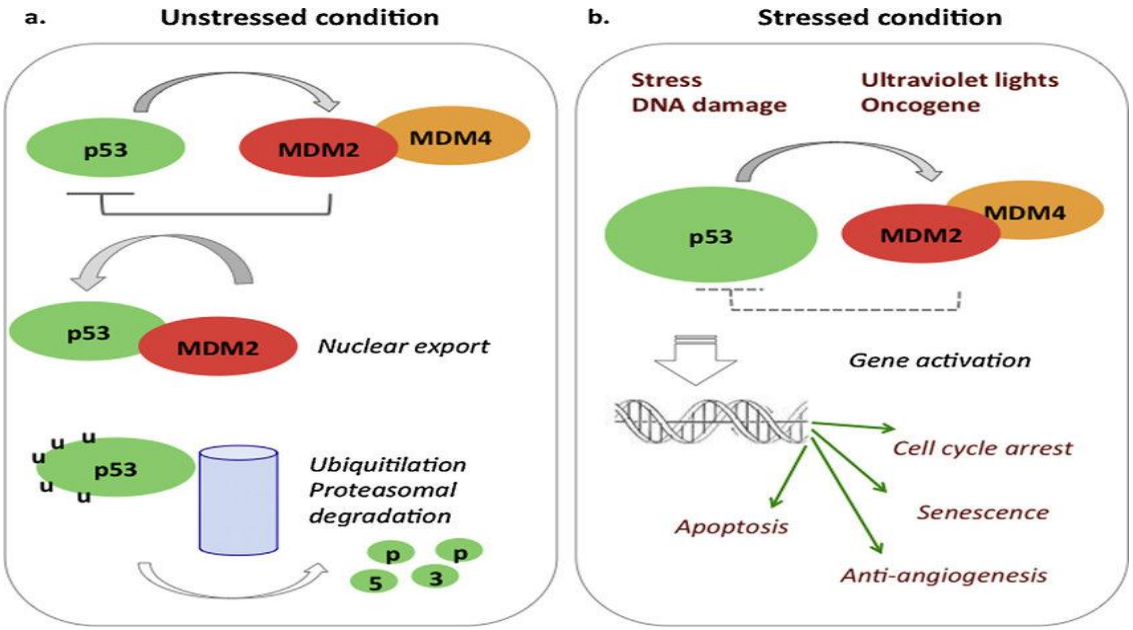
Fig. 2 The p53/MDM2 interactions and inhibitors. **a** The p53/MDM2 signaling pathway: MDM2 directly binds to p53 and inhibits its transcriptional activity, causes ubiquitination and proteasomal degradation of p53, and exports p53 out of the nucleus which promotes p53 degradation. **b** MDM2 (surface)-p53 peptide (green) complex (PDB:1T4F). **c** The chemical structures of inhibitors of MDM2/p53

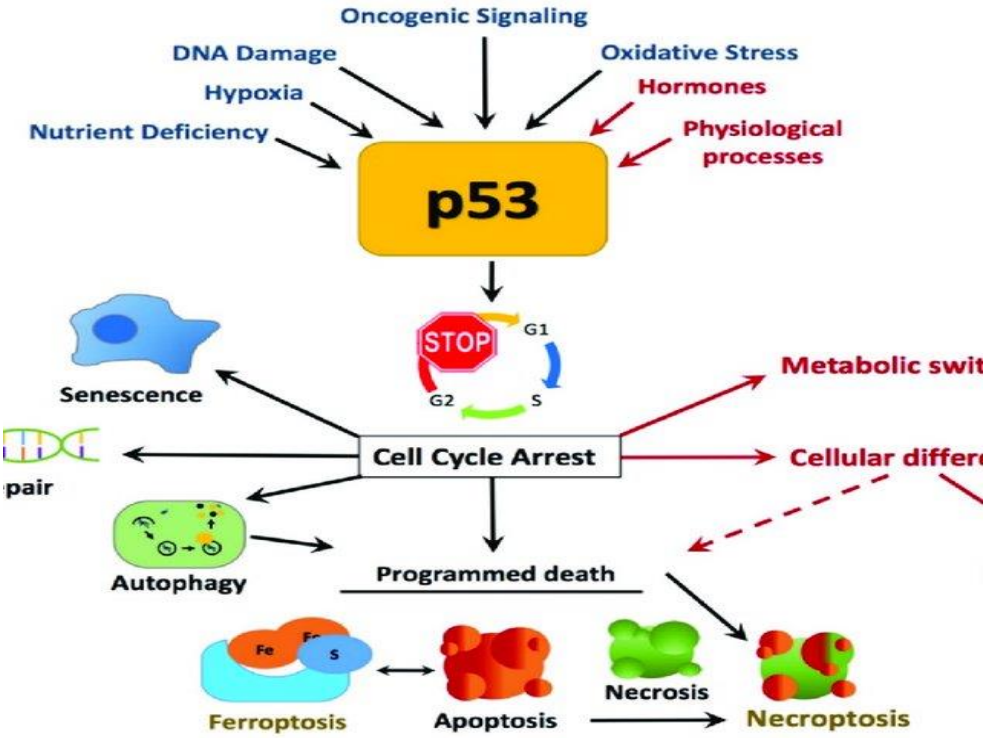
The tumour suppressor p53 induces cell death by apoptosis in response to various stress conditions, such as oncogene activation or DNA damage.

- The loss of p53 tumour-suppressor activity — either by mutation/deletion of the *TP53* gene or by inhibition of the p53 protein — favours the development of cancer.



- The MDM2 protein encoded by the mouse double minute 2 (*MDM2*) gene is the primary negative regulatory factor of the p53 protein. MDM2 can ligate the p53 protein via its E3 ubiquitin ligase, and the ubiquitinated p53 can be transferred to the cytoplasm and degraded by proteasomes.
- The MDM2 protein is a negative regulator of p53. After binding to p53, it inhibits its transcriptional activity, favours its nuclear export and stimulates its degradation
- The overexpression of MDM2 in various tumours inhibits p53, therefore favouring uncontrolled cell proliferation.
- The inhibition of the p53–MDM2 interaction is an attractive strategy to activate p53-mediated apoptosis in tumours with overexpressed MDM2, but wild-type p53.
- Several low-molecular-weight compounds and peptides that inhibit the p53–MDM2 interaction have been obtained. The peptidic inhibitors show an antiproliferative effect in tumour cells overexpressing MDM2.





Inhibitors of Bcl-2/Bax interaction (small molecules)

The Bcl-2 family is a key regulator of apoptosis, and it has over twenty members. According to their role in apoptosis, the Bcl-2 family members can be divided into two categories including the anti-apoptotic proteins and the pro-apoptotic proteins (Fig).

The anti-apoptotic proteins include Bcl-2, Bcl-w, Mcl-1, and Bcl-A1. The pro-apoptotic proteins include Bax, Bok and Bak, Bid, Bad, Bmf, Noxa, Puma, Hrk (among them, Bid, Bad, Bmf, Noxa, Puma, and Hrk are BH3-only protein).

Both antiapoptotic and pro-apoptotic members usually synergize in the form of dimers, having the role of apoptotic switch.

Proapoptotic proteins such as Bax and Bad have critical roles in the apoptosis.

The functions of these pro-apoptotic proteins are blocked when they bind to the anti-apoptotic proteins like Bcl2.

Therefore, inhibiting the interaction between the pro- and anti-apoptotic proteins prevents the tumor cells from escaping apoptosis

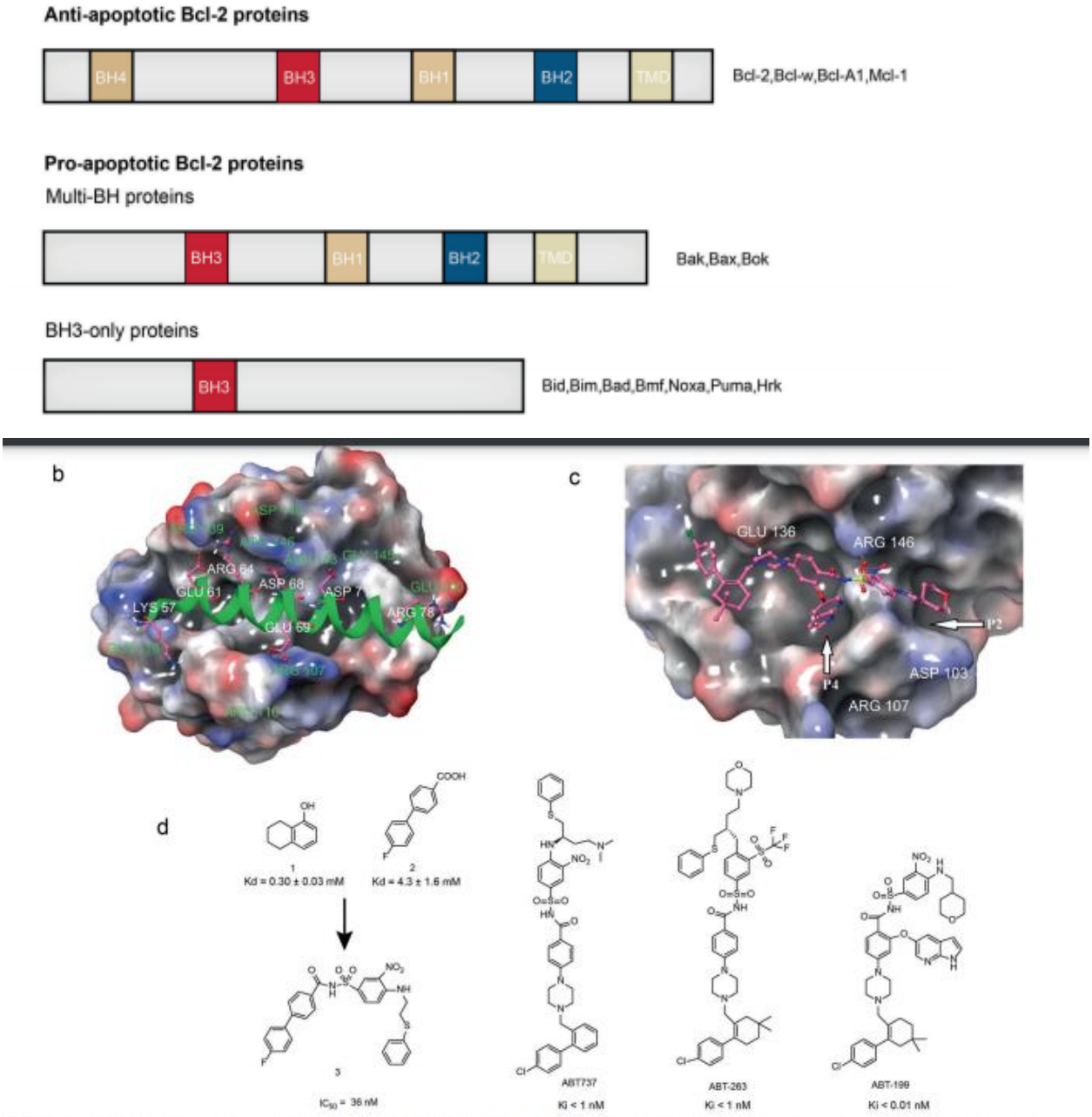
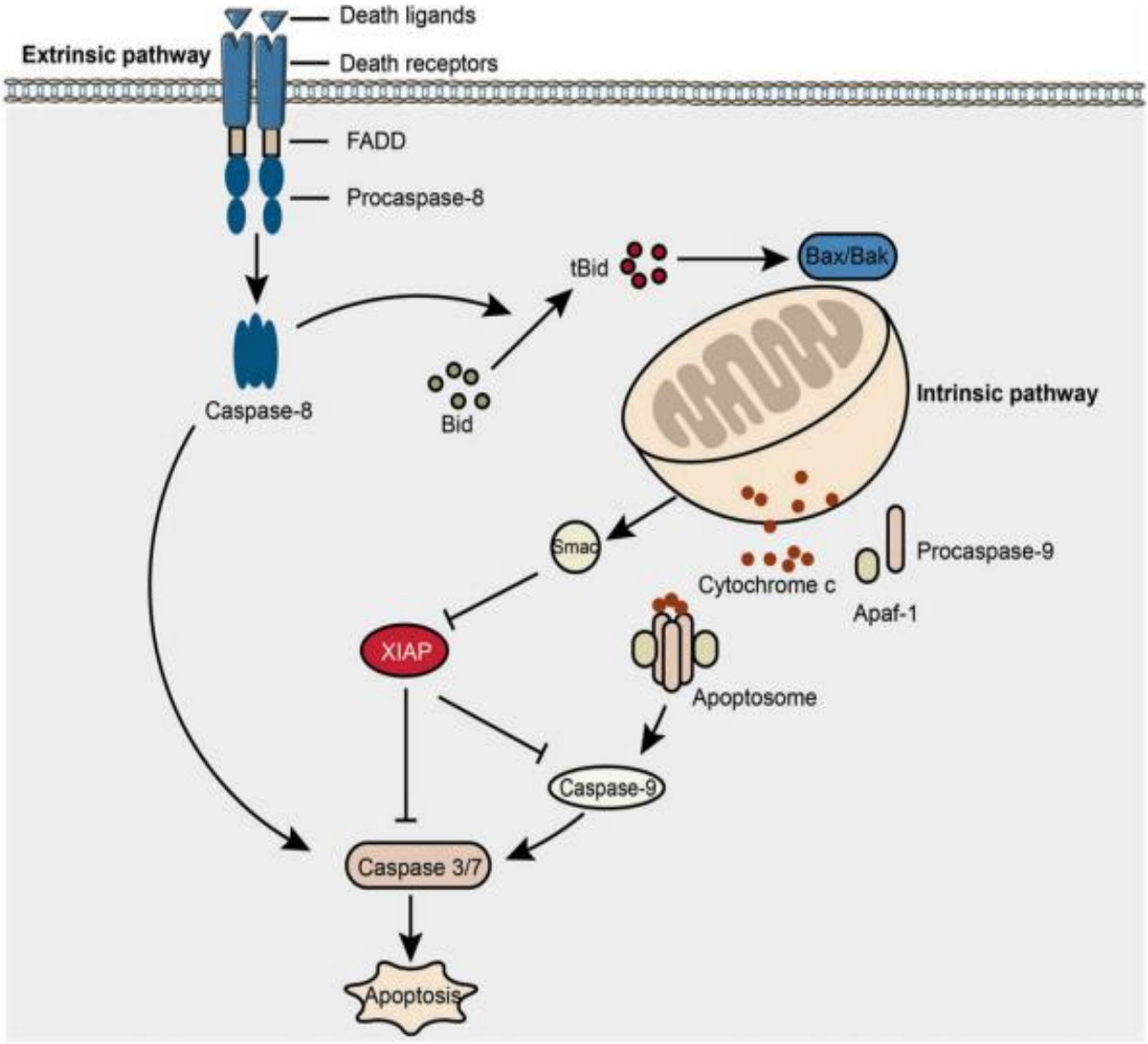


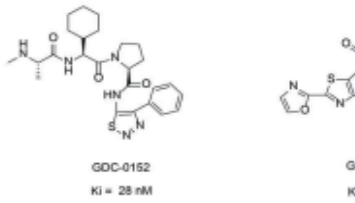
Fig. 3 The Bcl-2/Bax interactions and inhibitors. **a** The Bcl-2 family can be classified into two categories: the anti-apoptosis proteins and pro-apoptosis proteins. The pro-apoptosis proteins can be divided into multi-BH proteins and BH3-only proteins. **b** The crystal structure of Bcl-2 in complex with Bax BH3 peptide (PDB:2XA0). **c** The binding modes of ABT-199 binds to Bcl-2 (PDB:6GL8). **d** The chemical structures of inhibitors of Bcl-2/Bax

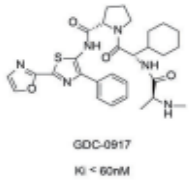


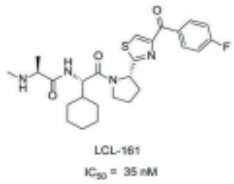
Inhibitors of apoptosis proteins (IAPs) are an important class of endogenous anti-apoptotic proteins.

- They bind to the caspase or other pro-apoptotic proteins, results in the inhibition of the proapoptotic proteins functions and promotes their degradation, thereby regulates the apoptosis
- The IAPs has eight family members: *XIAP, c-IAP1, c-IAP2, ML-IAP/Livin, ILP2, NAIP, Bruce/ Apollon, and surviving*.
- The caspase, a cysteine-containing aspartate proteolytic enzyme, is the main implementer of apoptosis, which induces apoptosis through two pathways.

- One of which is the death receptor pathway (extrinsic pathway) that mediated through caspase-8. The other one is the mitochondrial pathway (intrinsic pathway), which mediated via cytochrome C/ caspase-9
- The BIR3 domain of the XIAP binds to and inhibits pro-apoptotic caspase-9, thus suspends the apoptosis.
- Interestingly, the endogenous protein inhibitor of the XIAP–caspase-9 interaction exists in the form of Smac (second mitochondria-derived activator of caspase).
- When the Smac released from the mitochondria, its N-terminal amino acids, alanine–valine–proline–isoleucine (AVPI) bind to the BIR3 domain of XIAP, which makes the XIAP lose the ability to combine with caspase, so as to promote apoptosis
- The four amino groups of the AVPI at the N terminus of the Smac protein have a very important role in the binding of XIAP to caspase-9, which competes with the caspase-9 protein for binding to XIAP.

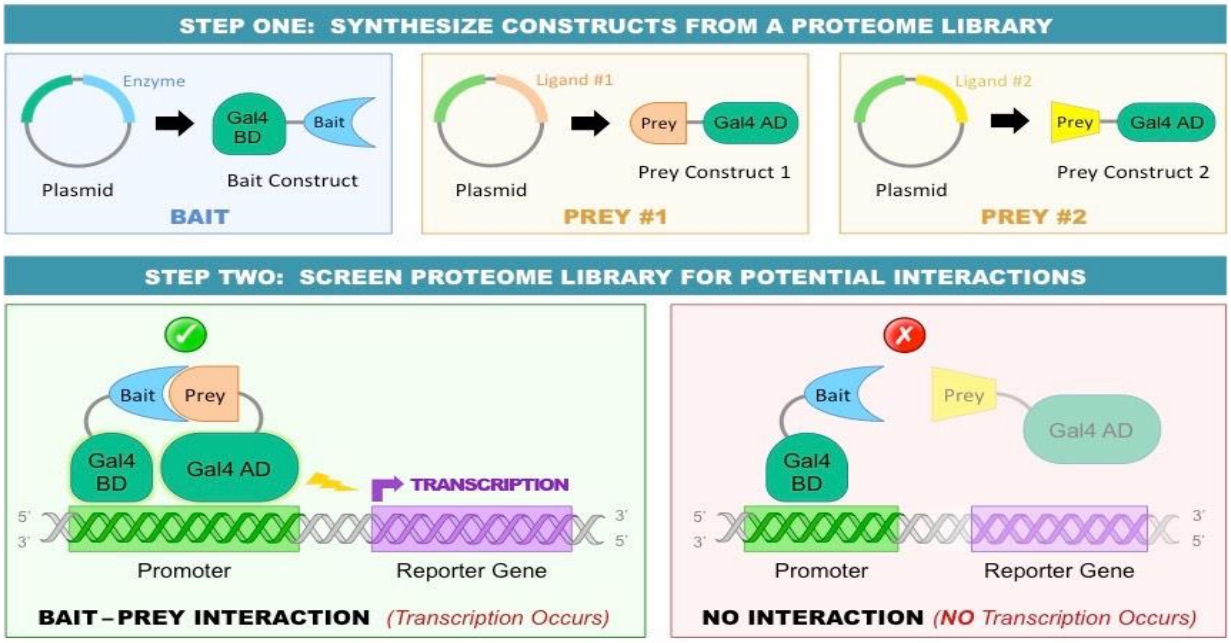






- Therefore, the interaction between XIAP and caspase9 can be inhibited by the Smac protein mimetics that exhibit the similar affinity to XIAP.
- The crystal structure of Smac and XIAP-BIR3 domain revealed that the Val of P2 position and the Ile of P4 position in the Smac formed three hydrogen bonds with the Gly306 and the Thr308 of XIAP-BIR3 domain.
- The 3-position Pro ring bind with the hydrophobic region formed by the Trp323 and Tyr324 of XIAP-BIR3 domain through van der Waals force.
- Moreover, the Pro ring is essential for maintaining the conformation of AVPI peptide chain, so proline is relatively stable and usually not replaced by other amino acids.

- Flygare et al discovered the first Smac simulator GDC-0152 through a combination of peptide-like design strategies and high throughput screening.
- The GDC-0152 binds to XIAP-BIR domain with high affinity by mimicking the structure of the Smac AVPI peptide.
- Another Smac mimetic GDC-0917 (CUDC-427) has entered phase I clinical trials for the safety evaluation of patients with advanced solid tumors and lymphomas.
- Novartis LCL-161 , which is currently progressing rapidly, has entered phase II clinical trials for triple negative breast cancer.
- Yeast-2-hybrid system
- The yeast two-hybrid (Y2H) method was first described over 20 years ago.
- It soon appeared as a major methodological breakthrough in the discovery and analysis of protein interactions, which play a pivotal role in all biological phenomena.
- Since its inception the Y2H method has constantly evolved and has inspired various assays that have found multiple applications of interest for drug discovery.
- Y2H methods are used to identify and validate therapeutic targets, discover protein interaction modulators, identify drug targets, and select combinatorial recognition molecules, which themselves find a wide range of applications.



The yeast-2-hybrid system is a simple scientific technique used to screen a library of proteins for potential interactions

- Firstly, a transcription factor is broken into two parts – a DNA-binding domain (BD) and a catalytic activation domain (AD)
- The DNA-binding domain is fused to a protein of interest called the bait (e.g. an enzyme)
- The activation domain is fused to a number of potential binding partners – called the prey (e.g. different ligands)
- If the bait and prey interact, the two parts of the transcription factor are reconstituted and activate transcription of a gene
- If the bait and prey do not interact, the two parts of the transcription factor remain separate and transcription doesn't occur
- The yeast-2-hybrid system detects protein-protein interactions according to the activation of a reporter gene
- The reporter gene may encode for the production of a protein that causes a visible colour change (e.g. ß-galactosidase)
- Alternatively, the reporter gene may encode for the production of an essential amino acid that is required for the yeast to grow on a deficient media (hence yeast growth would indicate successful interaction between bait and prey)
- Yeast-2-hybrid screens are a simple technique and hence have a relatively high rate of false positives (partial interactions)
- Consequently, the yeast-2-hybrid system is typically only used as an initial test to identify possible protein interactions

Yeast two-hybrid methods and their applications in drug discovery

Review

Cell PRESS

Feature review

Yeast two-hybrid methods and their applications in drug discovery

Amel Hamdi and Pierre Colas

Centre National de la Recherche Scientifique (CNRS) USR 3151, P_2I_2 Group, Station Biologique, Place Georges Teissier, 29680 Roscoff, France

The Y2H system

The Y2H system was first described in a seminal article published in 1989 by Fields and Song [1]. The principle of the assay relied on major discoveries on transcription initiation accumulated in the mid-1980s (Box 1). One of the key reasons for the tremendous success of the Y2H methodology is its timeliness [2]. In the early 1990s, unveiling a protein interaction was an arduous endeavor, mostly accessible to hard-core, cold-resistant protein biochemists. The Y2H thus filled a spectacular methodological gap in democratizing the discovery of protein interactions. Moreover, having gone through an exponential growth curve of published uses at the end of the last century [3], the Y2H became the undisputed reference method to discover protein interactions when genomics produced the first whole-genome sequences and created the need for high-throughput experimental biology (then termed functional genomics). Another obvious reason for the great success of the methodology lies in its remarkable versatility, which has allowed the development of several related assays, many of which have applications for drug discovery.

Y2H methods and their applications to detect protein–protein or protein–small-molecule interactions and to discover small-molecule inhibitors.

(a–d) The principle of different Y2H methods is depicted.

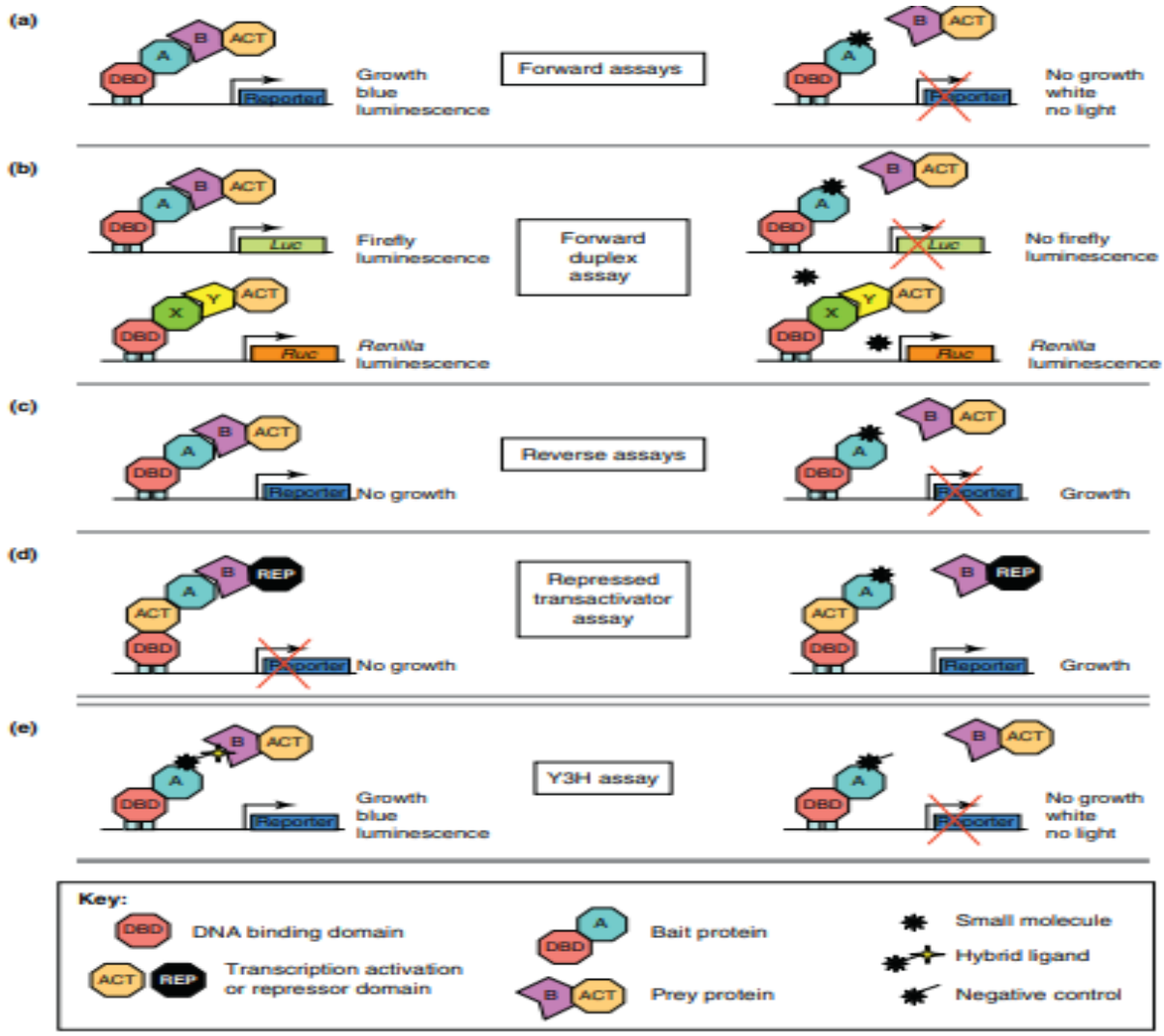
Left panels: detection of protein–protein interactions; Right panels: detection of inhibition of protein interactions by small molecules.

Y2H phenotypes are indicated next to the reporter genes.

(e) The principle of the Y3H method is depicted.

Left panel: detection of an interaction between a small molecule (yellow star) and a protein (labeled B).

Right panel:negative control setting, where a truncated hybrid molecule is used.



- Few proteins that are directly associated with a disease offer tractable therapeutic targets.
- Some present loss-of function mutations or are downregulated in disease states, which makes it difficult, although not always impossible,to restore their function by small-molecule drugs.
- A good example concerns the tumor suppressor p53, found mutated in half of cancers, and whose pathway is at least partially inactivated in the other half part.
- Other disease associated proteins, such as the oncoprotein Ras, present gain-of-function mutations or an upregulation but do not harbor druggable sites or instead harbor sites that cannot be targeted specifically because they are substantially conserved in other related proteins.
- The discovery of interacting partners of such intractable disease-associated proteins can offer excellent opportunities to identify promising therapeutic targets.

- In the case of p53, the discovery of its interaction with the MDM2 oncoprotein, which triggers its ubiquitin-mediated degradation, revealed that the interaction itself was a pertinent target, now actively pursued in clinical trials .
- In the case of Ras, early use of the Y2H method unveiled its long-suspected interaction with the Raf oncogenic protein kinase, which has become a very attractive target in a number of cancers .
- High-throughput Y2H studies have been performed on many pathogenic infectious agents.
- Studies have focused, for example, on the *SARS, Epstein–Barr, Varicella zoster viruses, the Campylobacter jejuni and Treponema pallidum bacteria, and Trypanosoma brucei.*
- The protein interaction maps generated by these efforts enhance the general knowledge on these infectious agents and suggest new potential therapeutic targets for future anti-infective drugs.
- Another very promising approach is to use Y2H screens to build host–pathogen protein interaction maps, which has been achieved for the Epstein–Barr, hepatitis C, influenza and dengue viruses.
- These efforts allow the investigation of biological questions that are very relevant for drug discovery, such as virulence, species barrier, chronicity and derived pathologies (such as cancer).
- They also considerably extend the target exploration field to the human proteins appearing on these maps. Targeting human proteins to treat infectious diseases holds great promise for at least three reasons.
- First, infectious agents (especially viruses) offer a very limited pool of tractable therapeutic targets, owing to the small size of their proteome.
- Second, treatments targeting viral or bacterial proteins often elicit drug resistance. Third, human proteins involved in interactions with viral proteins may constitute pertinent targets for broad-spectrum drugs, because different viruses seem to share common interactions with host proteins

Table 1. Small-molecule modulators of protein interactions discovered by Y2H screening assays

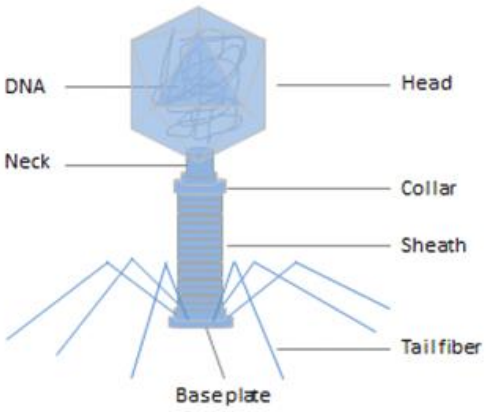
Y2H assays	Reporters	Targets	Biological pathways/ therapeutic fields	Chemical libraries	Hit number	Hit biological effects	Refs
Forward	*HIS3*	Ras/Raf1	Ras–MAPK/ cancer	10 000 microbial extracts	≥1	n.d.	[92]
Forward	*lacZ*	Ras/Raf1	Ras–MAPK/ cancer	73 400 diverse cpds	38	Inhibition of MAPK activation and tumor cell growth *in vitro*	[52]
Forward	*HIS3*	CFTR NBD1 ΔF508 dimerization	Chloride channel/ cystic fibrosis	600 plant leaf extracts	1	Restoration of dimerization; increase of chloride permeability in cultured cells	[61]
Forward	*lacZ*	Myc/Max (Id2/E47)[a]	Transcription/ cancer	10 000 diverse cpds	7 (10)[a]	Inhibition of tumor cell growth *in vitro* and in animals	[53]
Forward duplex	*luc* *ruc*	GαZ/RGS-Z1 (Kv4.3/KchIP1)[a]	G-protein signaling/ psychiatric disorders (potassium Channel)[a]	360 000 diverse cpds	75 (65)[a]	Inhibition of GTP hydrolysis of GαZ *in vitro*	[54]
Reverse	*CYH2*	N-type Ca^{2+} channel subunits	Neurotransmitter release/ stroke, traumas	156 000 diverse cpds	10	Inhibition of N-type Ca^{2+} channel activity *in vitro*	[56]
Reverse	*CYH2*	Kv4.3/KchIP1	Potassium channel/ epilepsy, stroke	n.d.	≥1	Modulation of current kinetics and amplitude in cultured cells	[93]
Reverse	*CYH2*	Kv1.1/Kvβ1	Potassium channel/ epilepsy, neuropathic pain	500 000 diverse cpds	18	Inhibition of Kv1.1 inactivation by Kvβ1 in cells; anticonvulsant activity in animals	[57]
Reverse	*GAL1*	Myc/Max	Transcription/ cancer	n.d.	n.d.	n.d.	[59]
RTA	*HIS3*	TGFβR/FKBP12 (TRα/Ncor1)[a] (ATF4/CBP)[a] (Nrf2/CBP)[a]	T cell activation/ immunosuppression (various biological pathways)[a]	23 247 diverse cpds	6 (86)[a] (1)[a] (2)[a]	Inhibition of IL-2 production by T cells; inhibition of SMAD-responsive reporter gene	[60]

Phage display

Bacteriophage (phage) is a virus that infects and replicates within bacteria and archaea, for following the injection of their genome into its cytoplasm.

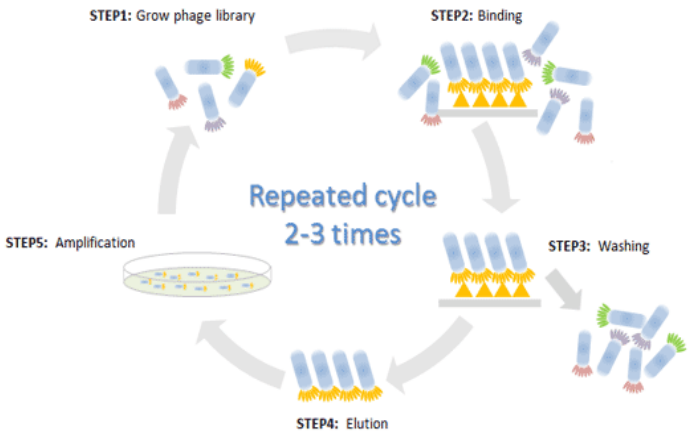
Phage is composed of proteins that encapsulate a DNA or RNA genome, and may have relatively simple or elaborate structures.

- Phage display was *first described by George P Smith in 1985*, who deployed it as a method to identify a gene against which he had raised antibodies.
- The technique was taken further by Greg Winter and John McCafferty at the Laboratory of Molecular Biology in Cambridge, UK, and Richard Lerner and Carlos F. Barbas at The Scripps Research institute, US, who independently used phage display to build large libraries of fully human antibody sequences. This work laid the foundation for the development of human antibody based drugs.
- Today there are several types of phage display libraries, including peptide libraries, protein libraries and antibody libraries.

- 




In phage display technique, a gene encoding a protein of interest is inserted into a phage coat protein gene, causing the phage to display the protein on the outside. And containing the gene for the protein inside, resulting in a connection between genotype and phenotype.
- Phage display is a laboratory platform that allows scientists to study protein interactions on a large-scale and select proteins with the highest affinity for specific targets.



STEP2: Binding

These libraries are exposed to selected targets and only some phages will interact with targets. The target is for which specific ligands planned to be identified such as immobilized protein, cell surface protein or vascular endothelium.

STEP3: Washing

Unbound phages can be washed away, and only those which showing affinity for the receptors was left.

STEP4: Elution

Recovery of the target bound phage by elution.

STEP5: Amplification

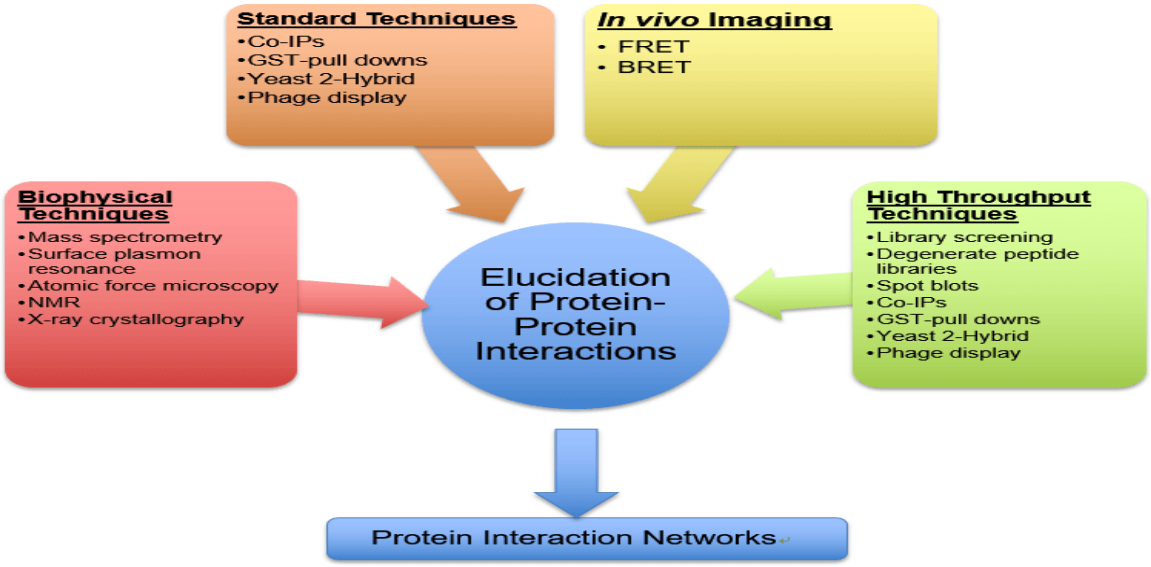
Eluted phages showing specificity are used to infect new host cells for amplification, or direct bacterial infection and amplification of the recovered phage.

Back to step 1, repeated cycle 2-3 times for stepwise selection of best binding sequence. After that, you can Enrichment and purification the phage repertoire by precipitation methods to increasing the phage titer.

Phage display is the technology that looks simple but hard to do. With years of experience, we can offer phage display service such as high-quality phage display library construction and custom phage display library screening services to meet your various demands precisely.

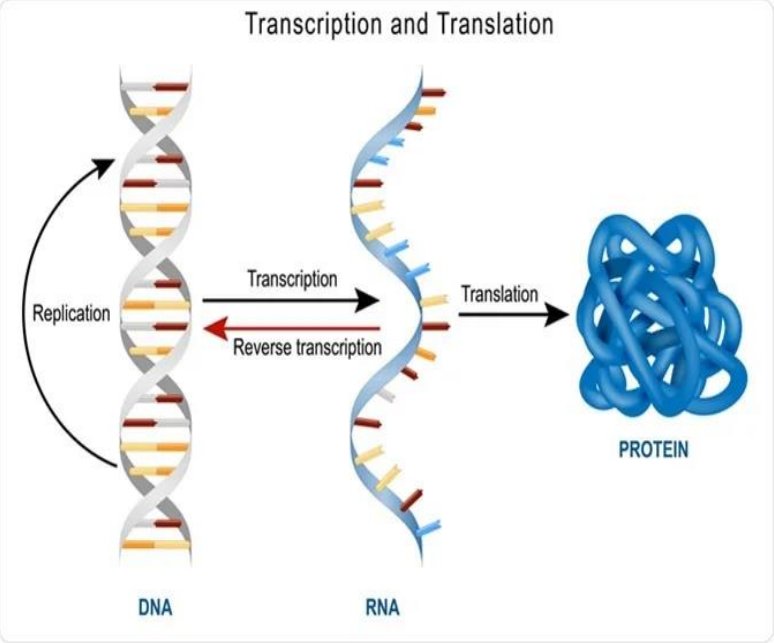
Classification of Phage Display Systems

The most common bacteriophages used in phage display are E.coli filamentous bacteriophages (f1, fd, M13) , though T4, T7, and λ phage have also been used



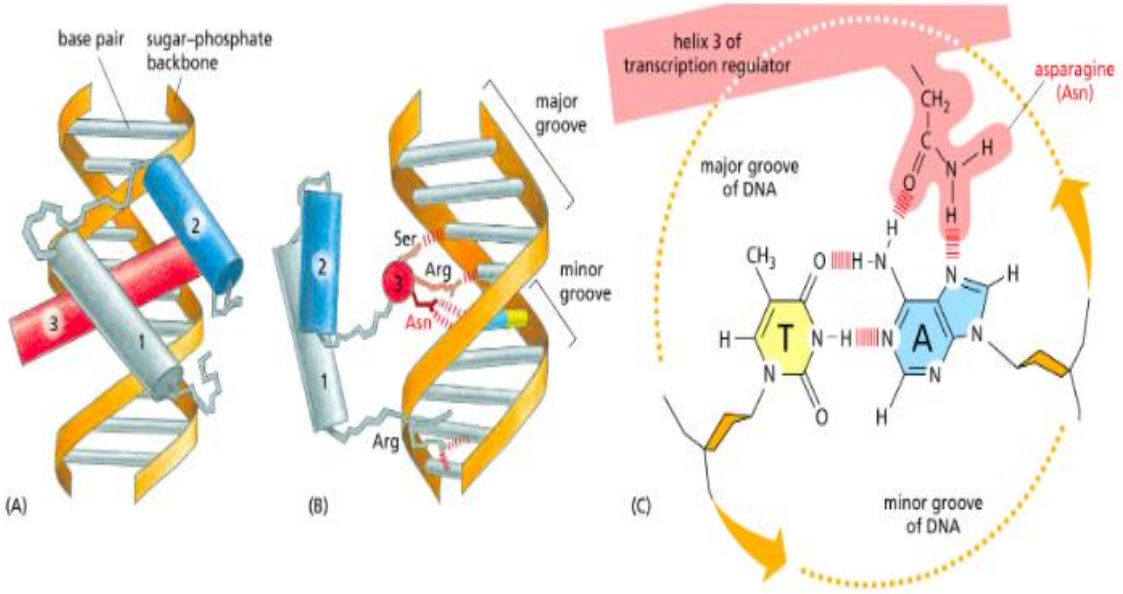
Protein DNA interactions – DNA binding motif

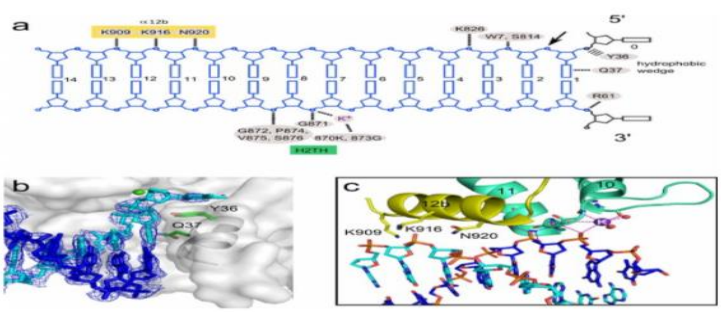
- DNA-Protein interactions play an essential role in the regulation of transcription.
- Understanding how DNA binds to specific residues to mediate binding specificity alongside determining the recognition mechanism has presented a molecular and computational challenge to scientists.
- Nonetheless, the basis of the interaction of DNA and proteins has neem demonstrated through several computational and experimental methods.



Protein–DNA interactions are when a protein binds a molecule of DNA, often to regulate the biological function of DNA, usually the expression of a gene.

Among the proteins that bind to DNA are transcription factors that activate or repress gene expression by binding to **DNA motifs** and **histones** that form part of the structure of DNA and bind to it less specifically.





Figure

Caption

Figure 2. Substrate binding. (a) Diagram of protein–DNA interactions. The preferred site of cleavage is indicated with an arrow. (b) Interactions with the hydrophobic wedge. The protein is shown in transparent surface representation, with helix _2 shown in a gray cartoon. The substrate is shown in cyan for the cleaved strand and blue for the non-cleaved strand. Calcium ion at the active site is shown as a green sphere, and the two residues involved in substrate binding are shown as green sticks. A 2Fo-Fc simulated annealing composite omit map contoured at 1.2 ⊠ is overlaid on the DNA. (c) H2TH motif. The DNA is colored as in (b). The potassium ion is shown as a purple sphere.

DNA-Protein interactions are mediated by one of two means;

(i) Direct contact between the base pairs of DNA and specific amino acids in the protein structure. This type of contact is intermolecular and termed a direct readout mechanism;

(ii) Indirect contact between the protein and DNA, mediated predominantly by water molecules and conformational changes in the structure of DNA. This type of contact is intramolecular and classified as an indirect readout mechanism.

- Direct readout mechanism is both redundant and flexible; this suggests the interaction is not based on a simplistic code.
- Furthermore, mutational analysis of specific bases on DNA that not in direct contact with amino acids often affect binding affinity. These changes are thought to arise from changes in the structure of intramolecular water molecules and bridging amino acids and bases, conformational shifts in the DNA structure and/or its flexibility.
- This further supports the role of indirect readout and the significance of intramolecular interactions for DNA-Protein binding.
- *Structural analysis of DNA-Protein complexes*

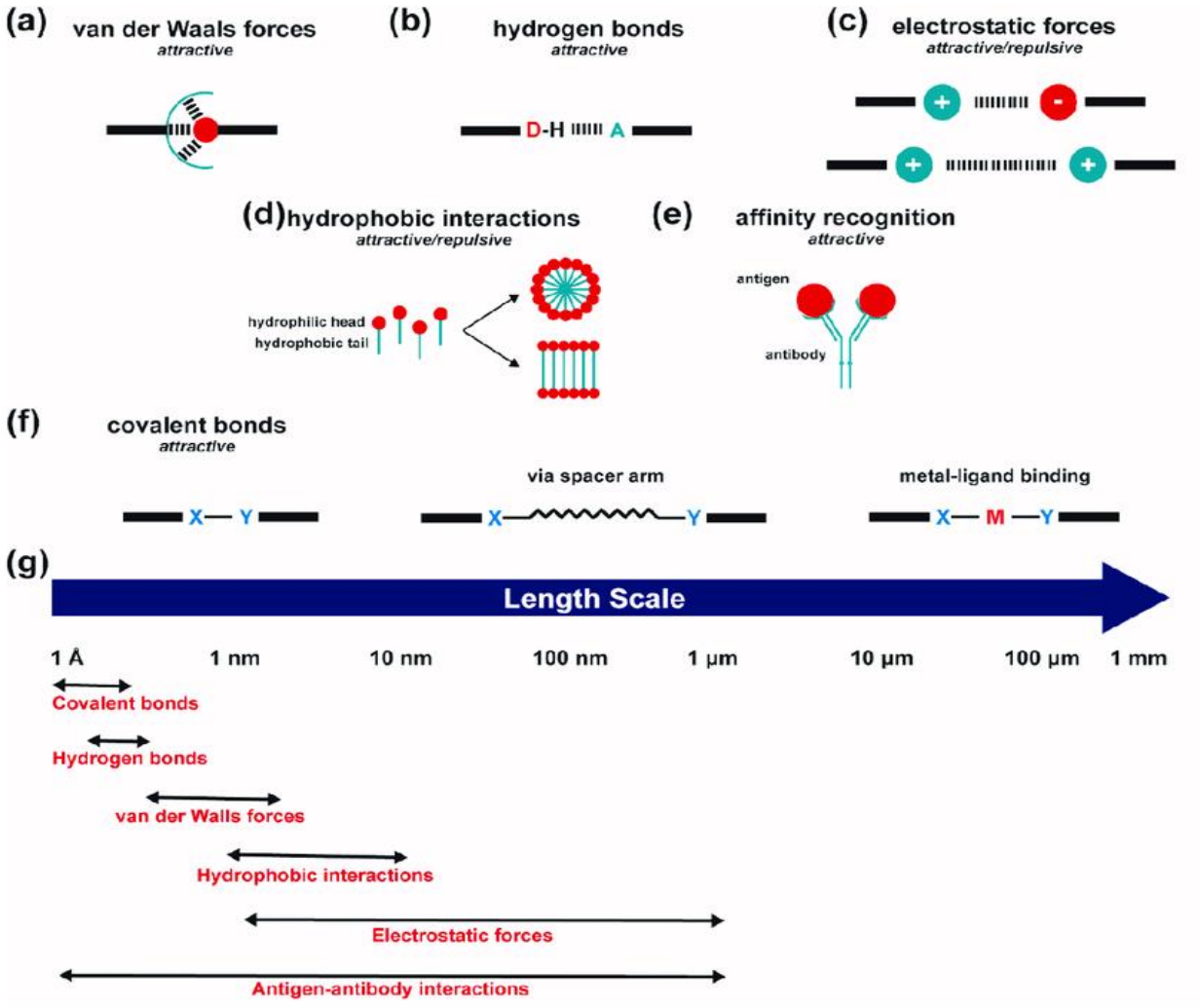
- Three-dimensional structures have been used to understand the variables that affect and mediate the formation of DNA-Protein complexes.
- *One such important factor are hydrogen bonds.*
- Several studies have demonstrated how hydrogen bonds aid in recognition of cognate DNA-Protein interactions; analyses have demonstrated that the C5 of cytosine and C5-Met of thymine form weak hydrogen bonds with the amino acids Aspartate, Asparagine, Glutamine, Glutamate, Serine and Threonine.
- Furthermore, analysis of DNA-binding sites has revealed discontinuous sequence segments form hydrophilic surfaces which are prime sites of hydrogen binding interactions.

Energetic contributions

- Analysis of the free energies of several forms of interactions *including electrostatics, hydrogen bonds, Van der Waals* and packing has demonstrated that Van der Waals

contributions mediate the formation of DNA-Protein complexes whilst electrostatic forces are unfavorable.

- Despite this, basic residues, which carry net positive charges are found to mediate favorable contribution to binding despite a desolvation penalty which arises. This refers to the removal of ordered water molecules from the basic residue which shield the charge in order to allow direct electrostatic interaction with an oppositely charged species.
- Acidic and neutral residues were poor electrostatic contributors; since DNA is negatively charged due to the sugar-phosphate backbone, this concurs with the finding.
- Acidic residues carry a negative charge, which would result in energetically unfavorable interactions with DNA. Similarly, neutral bases offer no charge stabilization.

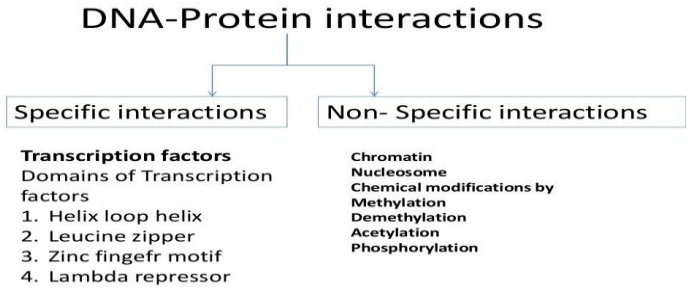


- Another important mediator of DNA-Protein binding is cation–π interactions.
- These are a form of noncovalent interaction in which the face of an electron rich π system, typically a heterocyclic ring system such as those seen in the benzene or indole rings common to the aromatic amino acids, and an adjacent positively charged ion (cation).
- 73% of protein–DNA complexes were found to involve such interactions, and these were found to function over long ranges.
- Furthermore, of the six possible pairs of residues Arginine–Tyrosine was found to be the strongest.

The residues on the binding protein occur largely as a group of conserved residues which display high packing density; the solvent interface area ranges between 1120 and 5800 Å2 and the binding site is populated with positively charged groups predominantly contributed by Lysine and Arginine side chains and the phosphate groups of DNA molecules.

Water-mediated contacts and conformational changes of DNA

- The conformational changes of DNA allow structural rearrangements to take place that are essential in mediating complex formation.
- Numerous studies have demonstrated that this conformational switching of DNA has also been found to mediate specificity.
- The conformational changes are evaluated by measuring six base step parameters that include shift, slide, rise, tilt, roll, and twist.
- In general, local variations particularly in the minor groove of the DNA alongside electrostatic interactions mediate the general mechanism for DNA binding specificity.
- Furthermore, water-mediated contacts control binding specificity.



SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

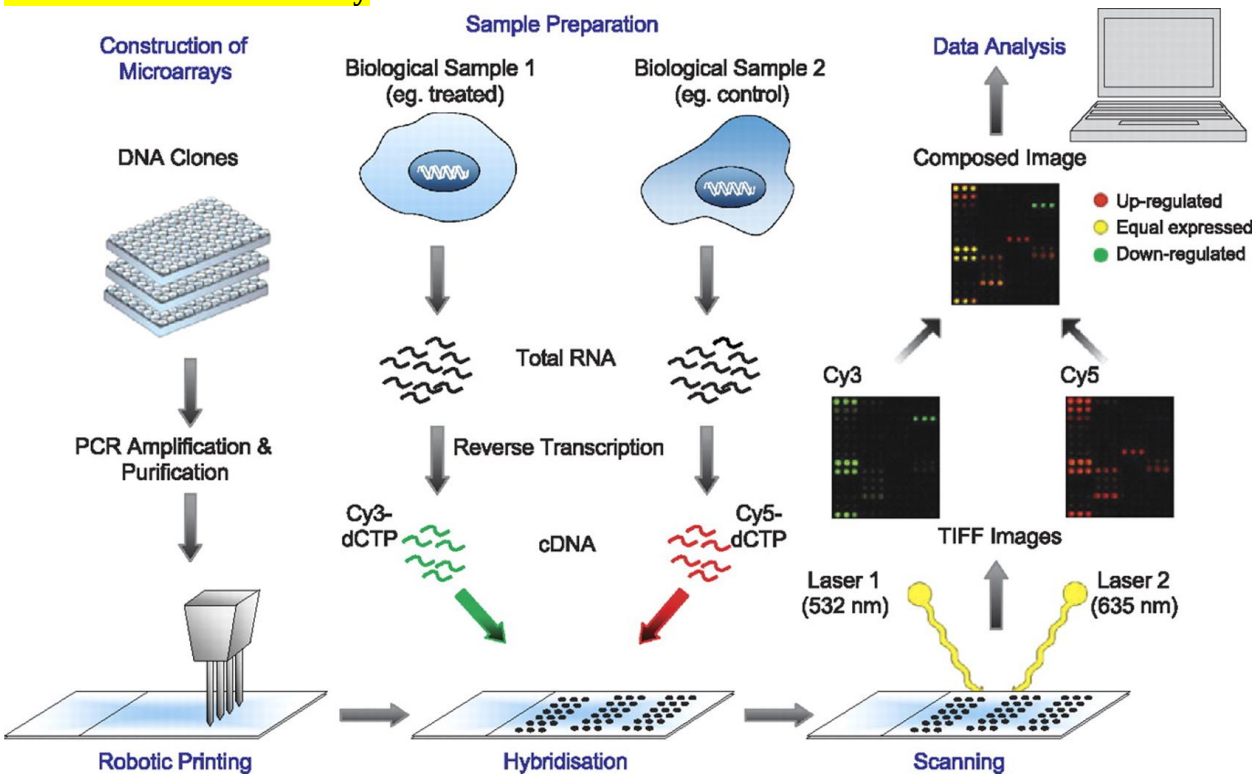
SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOINFORMATICS

UNIT – 4- SBIA1303 – Proteomics and Interactomics

- Microarray technology refers to the miniaturization of thousands of assays on one small plate.
- This approach was developed from an earlier concept called ambient analyte immunoassay, which was first introduced by Roger Ekins in 1989 (Ekins, 1989).
- In the decade that followed, this concept was successfully transformed into the DNA microarray, a technology that determines the mRNA expression levels of thousands of genes in parallel.

Basics of DNA Microarray

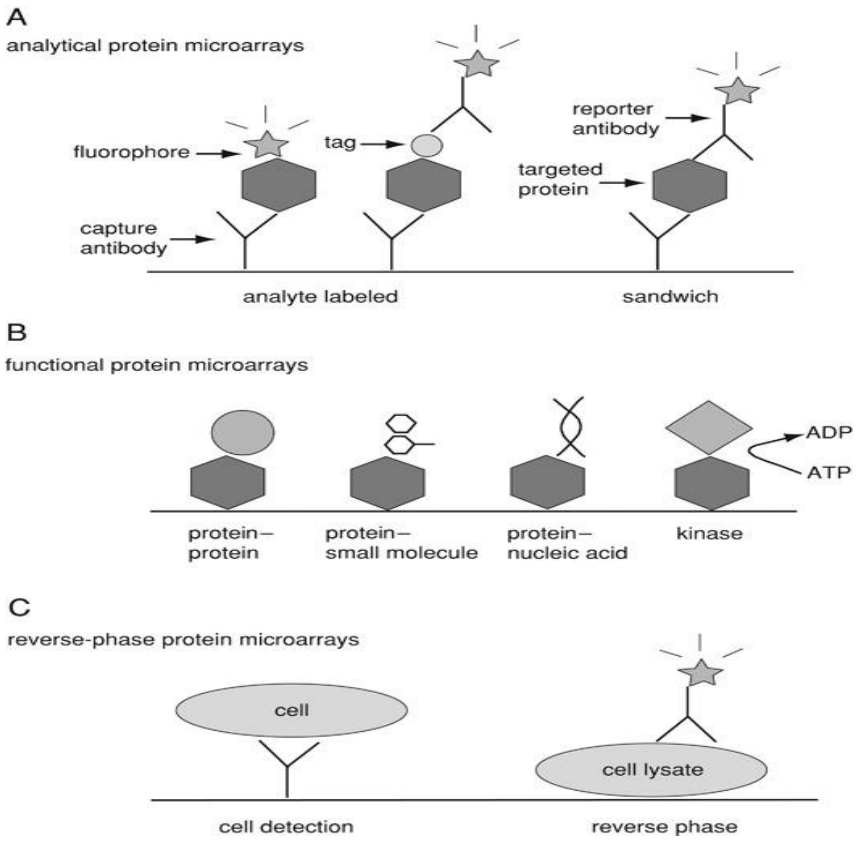


- However, DNA microarray technology possesses some limitations because mRNA profiles do not always correlate with protein expression.
- More importantly, proteins are the major driving force in almost all cellular processes.
- Therefore, protein microarrays were developed as a high-throughput tool to overcome the limitations of DNA microarrays and to provide a direct platform for protein function analyses.

- **Immunoassays**, the first form of protein microarrays, take advantage of highly specific **antigen-antibody recognition** to build a protein detection system.
- The expansion of the capability of conventional immunoassays into antibody array applications enabled a parallel and multiple detection system using a small amount of sample.
- Moreover, this technology has high sensitivity and good reproducibility in quantitative assays. The sensitive and reliable performance of antibody arrays is a valuable advantage when studying complex biological samples.

- Around the same time, another type of protein microarray was developed via the immobilization of purified proteins on glass slides.
- To distinguish this type of array from antibody arrays, they are divided into two classes:
- **Analytical and functional** (Chen and Zhu, 2006).

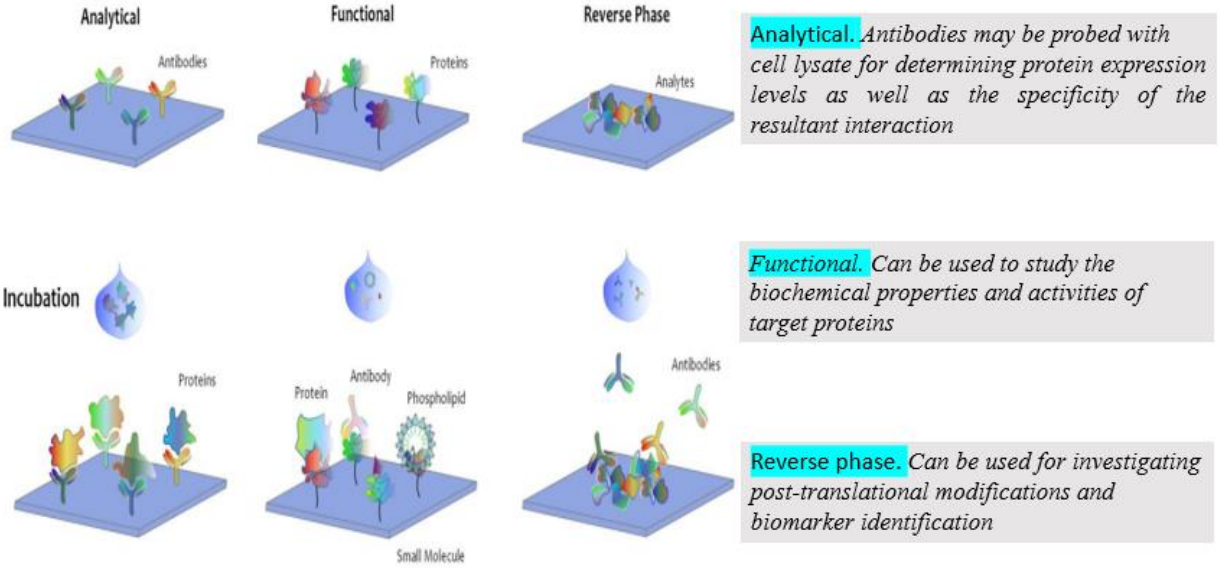
❑ Unlike antibody arrays (analytical microarrays), functional protein microarrays are made by spotting all of the proteins encoded by an organism and therefore are useful for the characterization of protein functions, such as protein-protein binding, biochemical activity, enzyme-substrate relationships, and immune responses



Three categories of protein microarrays.

- **Analytical protein microarrays** are mostly represented by antibody arrays and focus on protein detection. In this class of microarrays, targeted proteins can be detected either by direct labeling or using a reporter antibody in sandwich assay format.

- **Functional protein microarrays** have broad applications in studying protein interactions, including protein binding and enzyme-substrate reactions.
- **Reverse-phase protein microarrays** provide a different array format by immobilizing many different lysate samples on the same chip.



Analytical. *Antibodies may be probed with cell lysate for determining protein expression levels as well as the specificity of the resultant interaction*

Functional. Can be used to study the biochemical properties and activities of target proteins

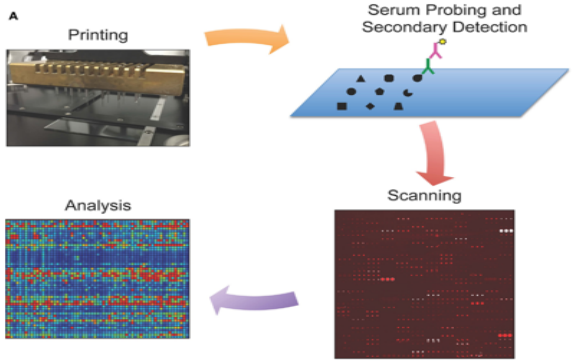
Reverse phase. *Can be used for investigating post-translational modifications and biomarker identification*

Protein microarrays: a new tool for the study of autoantibodies in immunodeficiency

- Autoimmunity is highly coincident with immunodeficiency.
- In a small but growing number of primary immunodeficiencies, autoantibodies are diagnostic of a given disease and implicated in disease pathogenesis.

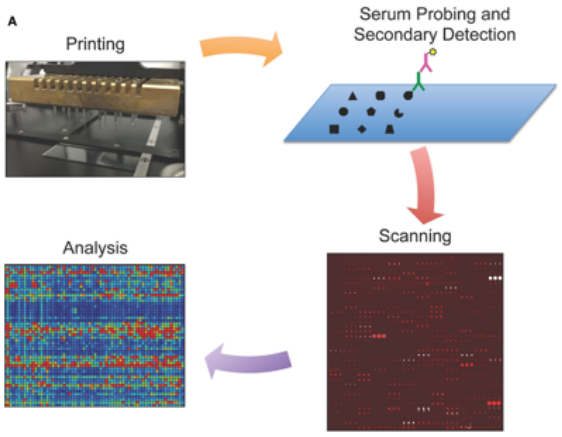
- In order to improve our understanding of the role of autoantibodies in immunodeficiencies and to discover novel autoantibodies, new proteomic tools are needed.

- Protein microarrays have the ability to screen for reactivity to hundreds to many thousands of unique autoantigens simultaneously on a single chip using minimal serum input.



- Antigens are printed onto a specially coated microscope slide surface, and serum antibodies (green) are detected by a fluorescently conjugated secondary antibody (purple).

- Microarrays are then scanned, and images are analyzed using microarray software.
- Values are calculated for each antigen based on mean fluorescent intensity and a statistical analysis is performed. Data can be visualized in a heat map representation



Detection and analysis of autoantibody reactivity by protein microarray have three key steps: (i) array design and fabrication; (ii) array probing, detection, and scanning; and (iii) image processing and data analysis

- Protein microarrays possess the greatest potential for providing fundamental information on protein, ligand, analyte, receptor, and antibody Affinity-based interactions, binding partners and high-throughput analysis.
- Microarrays have been used to develop tools for drug screening, disease diagnosis, biochemical pathway mapping, protein-protein Interaction analysis, vaccine development, enzyme–substrate profiling, and immuno-profiling

Critical issues include:

(1) Inconsistent printed microspot morphologies and uniformities,

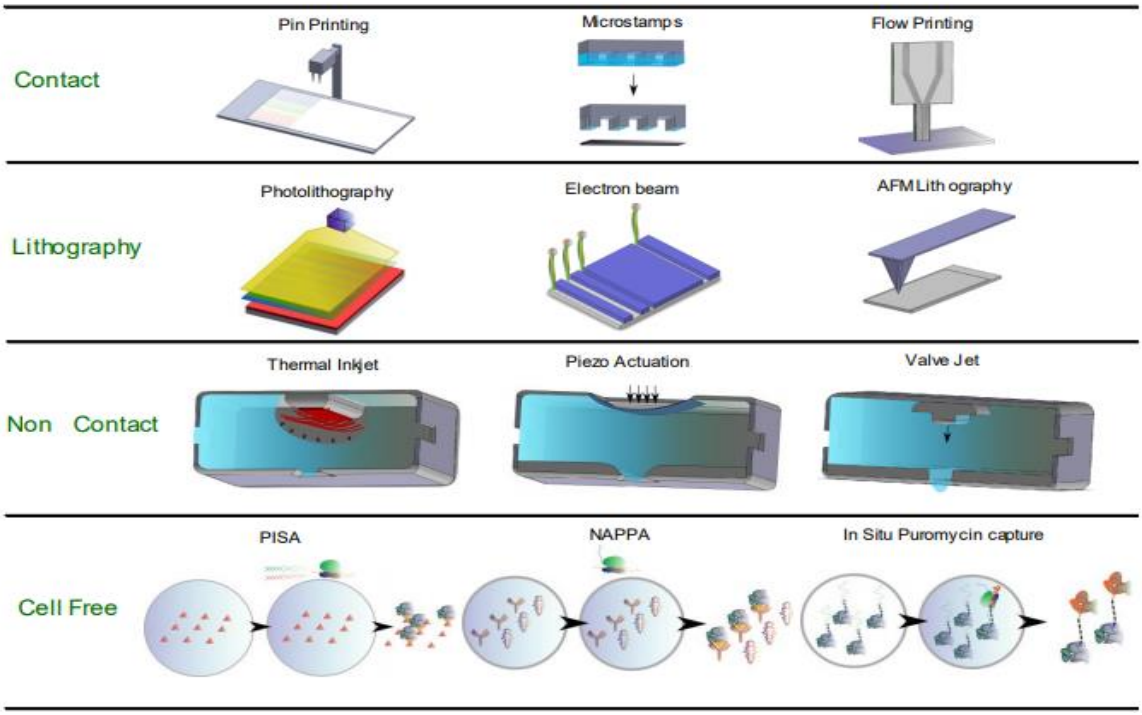
(2) Low signal-to-noise ratios due to factors such as complex surface capture protocols, contamination, and static or no-flow mass transport conditions

(3) Inconsistent quantification of captured signal due to spot uniformity issues

(4) non-optimal protocol conditions such as pH, temperature, drying that promote variability in assay kinetics

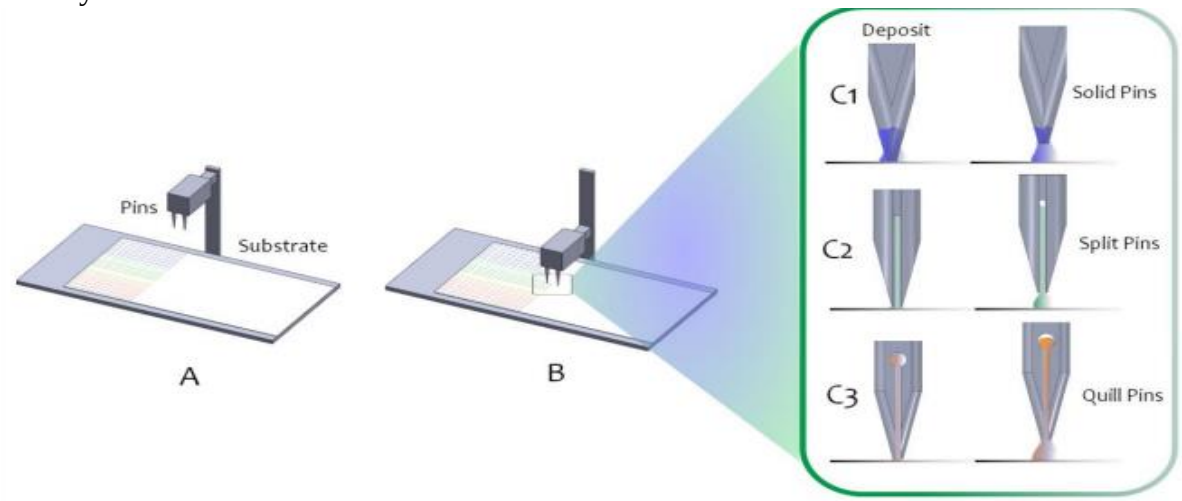
(5) Poor protein (e.G., Antibody) printing, storage, or shelf-life compatibility with common microarray assay fabrication methods, directly related to microarray protocols.

Conventional printing approaches, including contact (e.g., quill and solid pin), non-contact (e.g., piezo and inkjet), microfluidics-based, microstamping, lithography, and cell-free protein expression microarrays, have all been used with varying degrees of success rate.



Microarray Design and Fabrication

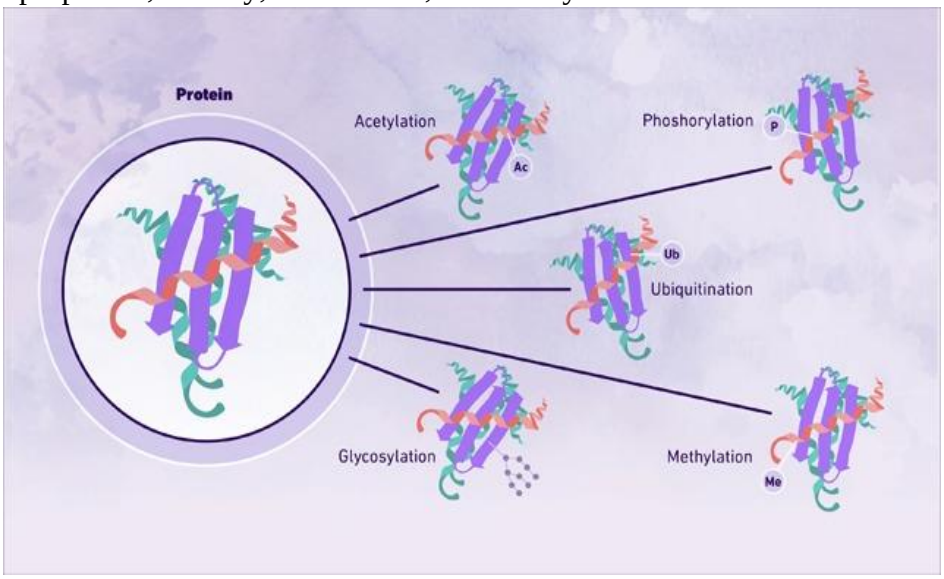
- Protein microarrays can be designed and fabricated independently or purchased commercially.
- Array fabrication requires a microarray printer, purified antigens of interest (either expressed in the laboratory or purchased commercially), and a microarray surface on which to print, typically a specially coated microscope slide.
- Antigens are loaded into one or multiple 384 well plates at either a single concentration or a series of concentrations (our typical protein printing concentration is 200 µg/ml).
- A typical microarray printer can print on the scale of 100 microarrays over the course of 1 day.



Schematic of pin printing.
(A) A robotic print head with multiple printing pins is loaded with print solutions from a source plate and then contacts the substrate surface to deposit protein
solution in (B).
Various types of pins: C1 is a solid pin. C2 is a slotted pin. C3 is a quill pin, distinguished from the split pin by the inclusion of a reservoir.

Concepts of post transitional modifications

- Posttranslational modifications (PTMs) of proteins perform crucial roles in regulating the biology of the cell.
- PTMs are enzymatic, covalent chemical modifications of proteins that typically occur after the translation of mRNAs.
- These modifications are relevant because they can potentially change a protein's physical or chemical properties, activity, localization, or stability



Types of post-translational modification

There are many types of protein modification, which are mostly catalyzed by enzymes that recognize specific target sequences in proteins. These modifications regulate protein folding by targeting specific subcellular compartments, interacting with ligands or other proteins, or by bringing about a change in their functional state including catalytic activity or signaling. The most common PTMs are:

Based on the addition of chemical groups

1. Phosphorylation
2. Acetylation
3. Hydroxylation
4. Methylation

Based on the addition of complex groups

1. Glycosylation
2. AMPylation
3. Lipidation

Based on the addition of polypeptides

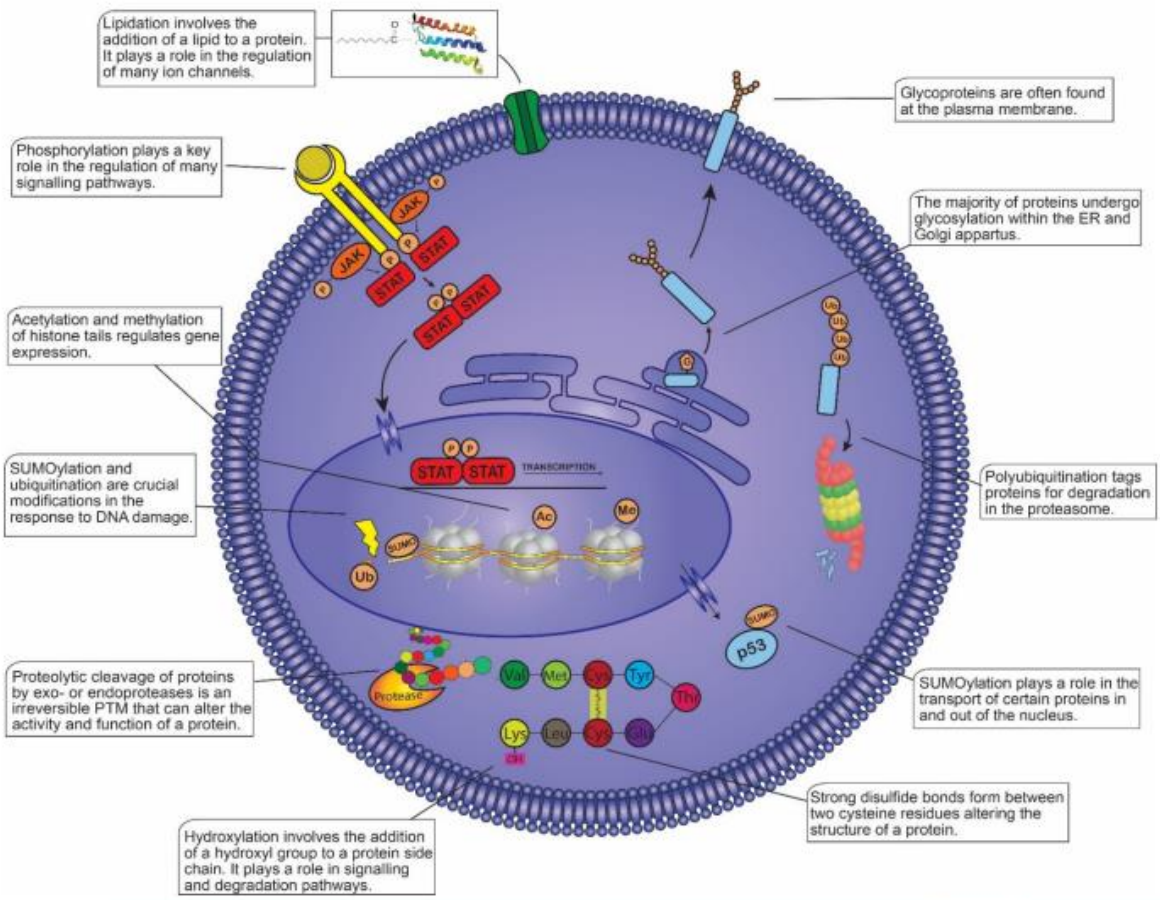
Ubiquitination

Based on the cleavage of proteins

Proteolysis

Based on the amino acid modification

Deamidation



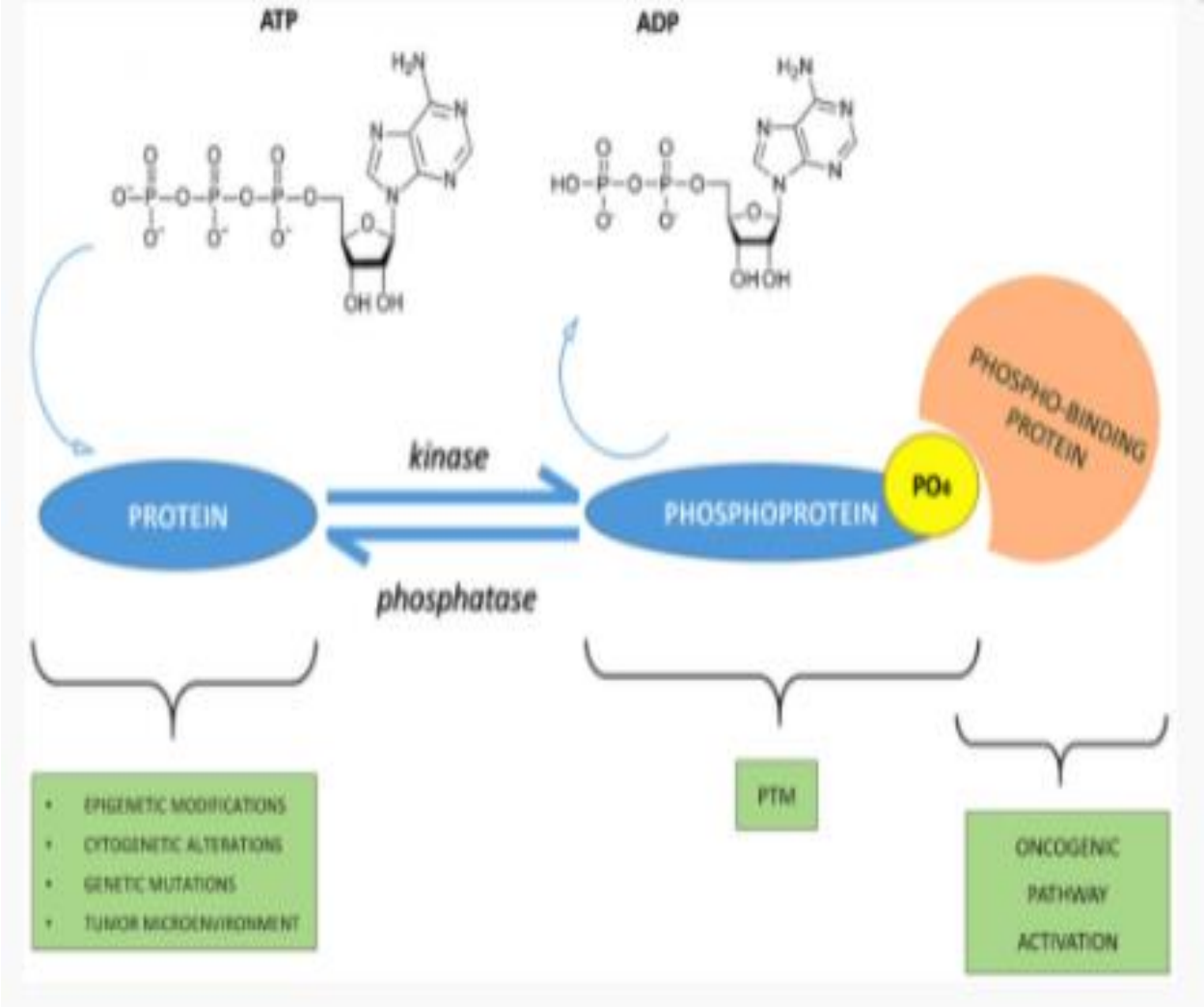
Post-translational modifications (PTMs) within the mammalian cell.

- PTMs are found throughout the cell from the plasma membrane to the nucleosomes present within the nucleus.
- PTMs play crucial roles in almost all cellular processes including the cell cycle,
- degradation, apoptosis, cell signaling, transcription, etc.
- Different proteins modified by the same PTM will not always yield the same response, demonstrating the diverse functions of PTMs within the cell.
- JAK, Janus kinase; STAT, signal transducer

- and activator of transcription; P, phosphate group; G, glycosyl group; Ub, ubiquitin; SUMO, small ubiquitin-like modifier;
- Ac, acetyl group; Me, methyl group; Val, valine; Met, methionine; Cys, cysteine; Tyr, tyrosine; Glu, glutamic acid; Lys, lysine.

Phosphorylation

- Reversible phosphorylation of proteins involves addition of a phosphate group on serine, threonine, or tyrosine residues and is one of the important and extensively studied PTM in both prokaryotes and eukaryotes.
- Several enzymes or signaling proteins are switched 'on' or 'off' by phosphorylation or dephosphorylation. Phosphorylation is performed by enzymes called 'kinases', while dephosphorylation is performed by 'phosphatases'.
- Addition of a phosphate group can convert a previously uncharged pocket of protein into a negatively charged and hydrophilic protein thereby inducing conformational changes in the protein.
- Phosphorylation has implications in several cellular processes, including cell cycle, growth, apoptosis and signal transduction pathways.
- One example is the activation of **p53, a tumor suppressor protein. p53 is used in cancer therapeutics and is activated by phosphorylation of its N-terminal by several kinases**

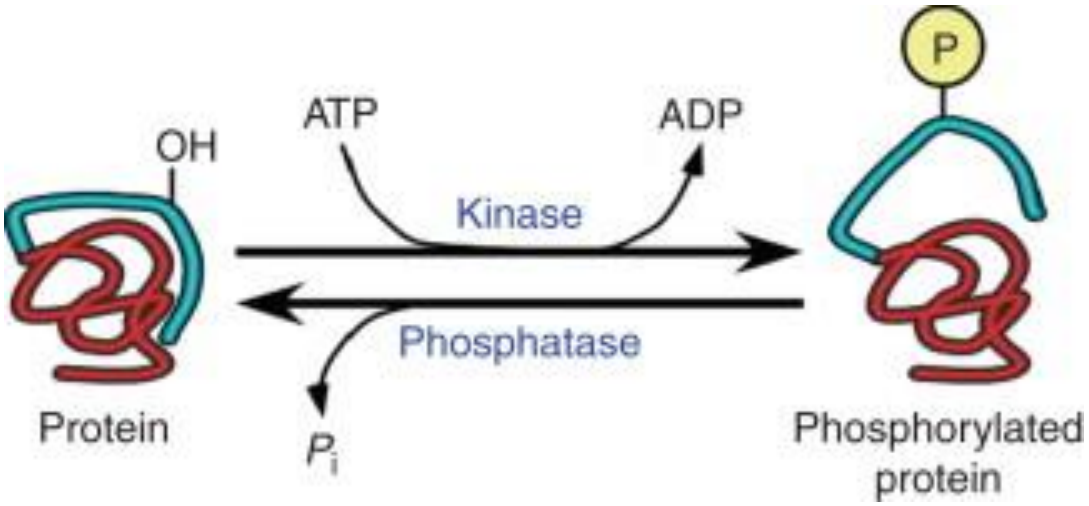


Phospho-signaling networks.

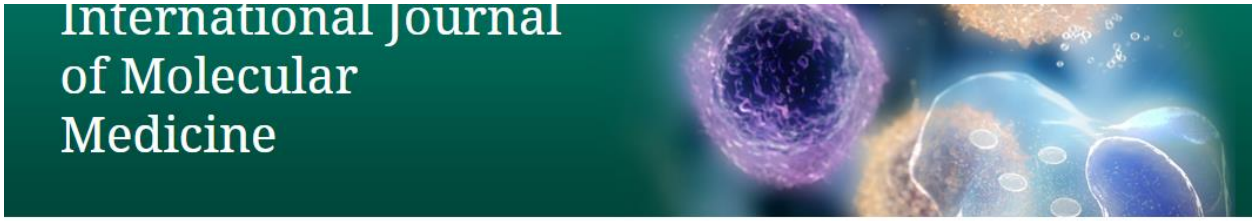
- The mechanism of phosphorylation regulation consists of kinases, phosphatases and their substrates phospho-binding proteins.
- For example, phosphorylation is activated by stimuli such as epigenetic modifications, cytogenetic alterations, genetic mutations or the tumor micro-environment.
- Consequently, the protein receives a phosphate group by adenosine triphosphate (ATP) hydrolysis and due to enzymatic activity of kinase. This is the mechanism for the basis of post-translational modification (PTM) formation.
- In addition, phosphorylation is a reversible process due to activity of phosphatase.

Phosphorylation and dephosphorylation are a molecular switch and, in particular, a PTM can cause oncogenic pathway activation by a phospho-binding protein that bind to the phosphate group of a phosphoprotein

The human genome, in fact, includes approximately 568 protein kinases and 156 protein phosphatases that regulate phosphorylation events and, therefore, play an important role in the control of biological processes such as proliferation, differentiation and apoptosis.



Types of protein kinases based on amino acid residue



Types of protein kinases based on amino acid residue

Types	Examples	Functions
Serine/threonine-specific protein kinases	Calcium/calmodulin-dependent protein kinase II (CaMKII)	Phosphorylate serine or threonine's –OH (hydroxyl) functional group.
Tyrosine-specific protein kinases	Platelet derived growth factor (PDGF) receptor Epidermal growth factor (EGF) receptor1 Insulin growth factor (IGF1) receptor Stem cell factor (scf) receptor	Processes alzheimer's amyloid precursor protein, epithelial cell migration and carcinoma invasion, spermatogonia osmoregulation, and antiaging survival factor
Histidine-specific protein kinases	Histidine kinase	The histidine kinase family is structurally similar to the pyruvate dehydrogenase family of kinases in animals.
Mixed kinases	Muscle action potential kinase (MAPK)	Involved in the cascade of muscle action potential kinase

Classes of protein kinases

Types	Functions			
Protein kinase A (PKA)	Acts as a catalyst, allowing intracellular proteins to be catalysed. Glycogen, sugar, and lipid metabolism are all regulated by this protein. Phosphorylate acetyl-CoA carboxylase and pyruvate dehydrogenase in adipocytes, myocytes, and hepatocytes, resulting in acetyl-CoA-lipogenesis. The dopamine signal is translated into cells in the nucleus acumens neurons.			
Protein kinase B (PKB)	Akt	Akt1	Akt2	Akt3
	Cell proliferation, apoptosis, transcription, and cell migration are all involved in glucose metabolism. Cell survival - both actively and indirectly facilitate growth factor-mediated cell survival. Glycogen synthesis is aided by metabolism. Angiogenesis is linked to tumour growth and angiogenesis.	Is implicated in: cellular survival pathways, by inhibiting apoptosis processes. skeletal muscle hypertrophy and general tissue growth. the transforming retrovirus as the oncogene	The insulin signalling pathway has been linked to this protein. Glucose transport induction is a term used to describe the process of causing glucose	It tends to be expressed primarily in the brain. The brains of mice lacking Akt 3 are weak.
Protein kinase C (PKC)	Phosphorylation of intracellular proteins is catalysed by this enzyme, which changes their activities. Controls cellular growth and differentiation.			

Advances in Protein Chemistry and Structural Biology
Volume 124, 2021, Pages 47-85

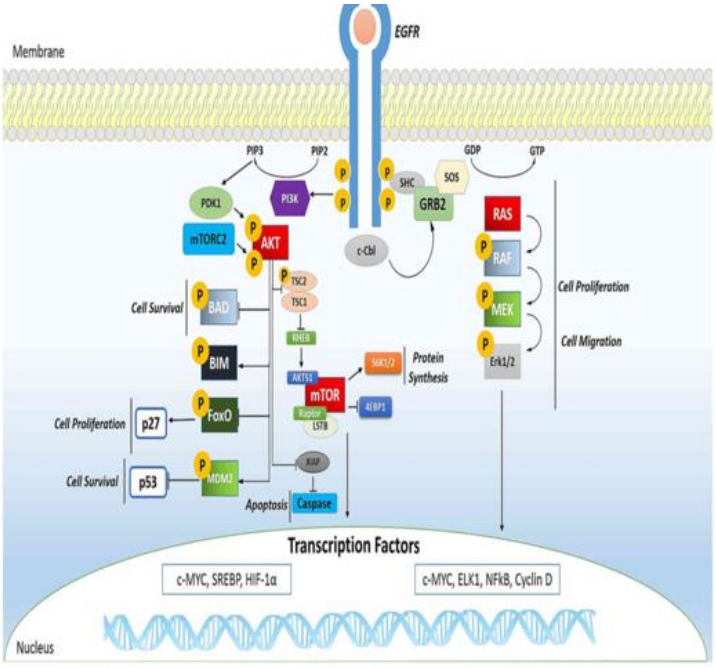
Chapter Three - Emerging role of protein kinases in diabetes mellitus: From mechanism to therapy

Preeti Gupta, Aaliya Taiyab, Md Imtaiyaz Hassan

Show more

- Diabetes mellitus has emerged as a severe burden on the medical health system across the globe. Presently, around 422 million people are suffering from diabetes which is speculated to be expanded to about 600 million by 2035.
- Patients with type 2 diabetes are at increased risk of developing detrimental metabolic and cardiovascular complications.
- The scientific understanding of this chronic disease and its underlying root cause is not yet fully unraveled.

- Protein kinases are well known to regulate almost every cellular process through phosphorylation of target protein in diverse signaling pathways.
- The important role of several protein kinases including AMP-activated protein kinase, IκB kinase and protein kinase C have been well demonstrated in various animal models. They modulate glucose tolerance, inflammation and insulin resistance in the cells via acting on diverse downstream targets and signaling pathways.
- Thus, modulating the activity of potential human kinases which are significantly involved in diabetes by targeting with small molecule inhibitors could be an attractive therapeutic strategy to tackle diabetes.

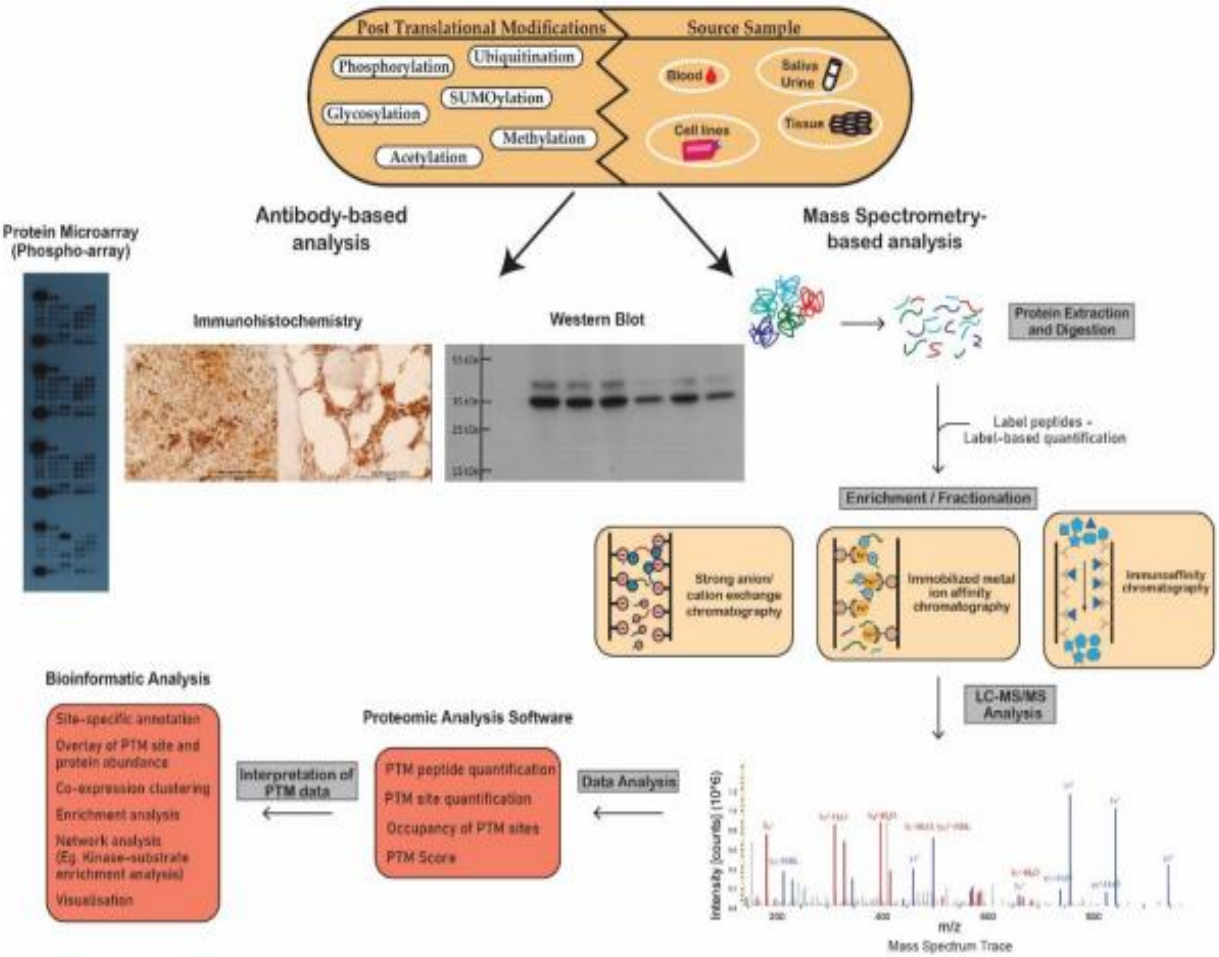


Endothelial growth factor receptor (EGFR) signaling

- Ligand binding to the EGFR activates its intrinsic tyrosine kinase activity.
- The autophosphorylation of EGFR causes the activation of several signaling pathways such as PI3K/AKT and RAS/mitogen-activated protein kinase (MAPK).
- In both, the phosphorylation is a predominant event and plays an important role in cell survival (AKT activation phosphorylates BAD and MDM2), cell proliferation (AKT activation phosphorylates FoxO, RAS phosphorylates RAF and it phosphorylates MEK that phosphorylates Erk1/2), cell migration (phosphorylation cascade of RAS), apoptosis (AKT phosphorylation is able to activate caspase).
- However, the phosphorylation of AKT causes mTOR activation, important in protein synthesis.

- Post-translational modifications (PTMs) are one of the fastest-growing areas of molecular biological research.
- Detecting post-translational modifications, knowing how they work, influence the proteome, and regulate the genome will greatly improve our understanding of both genetics and epigenetics.

- Currently, more than 300 types of PTMs have been discovered, but the number of those that have been examined at the proteome level is much smaller.



Schematic workflow illustrating analytical techniques used in the analysis of post-translational modifications.

- Proteomic techniques used for PTM analysis can be divided into antibody-based analysis or mass spectrometry (MS)-based analysis.
- Protein microarrays, immunohistochemistry and Western blot analysis demonstrate widespread popularity as efficient tools for antibody-based PTM analysis.
- MS-based analysis is a powerful technique in the study of PTMs.
- This figure outlines the steps involved in the preparation of crude samples for MS and the transformation of MS data into interpretable results.

Based on the addition of complex groups

1. Glycosylation
2. AMPylation
3. Lipidation

Glycosylation

- Many eukaryotic and prokaryotic proteins also have carbohydrate molecules attached to them in a process called glycosylation, which can promote protein folding and improve stability as well as serving regulatory functions.
- Glycosylation is an extremely diverse, enzymatic process resulting in the covalent attachment of sugars (glycans and monosaccharides) to proteins. About half of all cellular proteins are glycosylated. Most, such as transmembrane receptors, secreted proteins, organelle residents and surface ligands become glycosylated either coupled to or shortly after synthesis of the protein in the rough endoplasmic reticulum (ER).
- In addition to proteins, lipids and proteoglycans can also be glycosylated.

- Glycosylation has many functions.

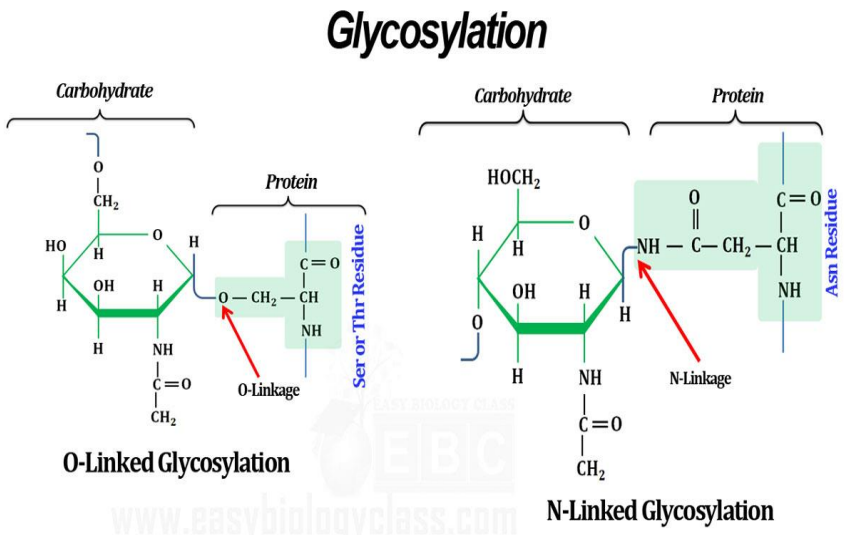
- Glycosylation aids in protein folding and stability and serves as a quality-control checkpoint for properly folded proteins.
- Glycosylation alters protein: protein interactions.
- Glycans can serve as recognition domains to aid in proper trafficking of proteins within the cell as well as facilitate ligand-receptor interactions resulting in signal transduction pathways.
- Glycans help mediate cell-cell adhesion and immune responses.
- In addition, glycosylation can alter the solubility of a protein.
- Changes in glycosylation patterns are linked to diseases, highlighting the importance of glycosylation to protein function.

- Glycosylation is a dynamic event that requires the action of multiple enzymes.
- Glycosyltransferases link sugars to proteins while glycosidases remove sugars from proteins.
- Glycosylation enzymes are segregated to different cellular organelles and act on proteins as they traverse the secretory pathway.
- These step-wise interactions add, remove and trim sugar moieties resulting in a diverse range of glycosylated products.
- Different cell types contain different glycosylation enzymes altering the types of glycans that can be added

There are five types of glycosylation:

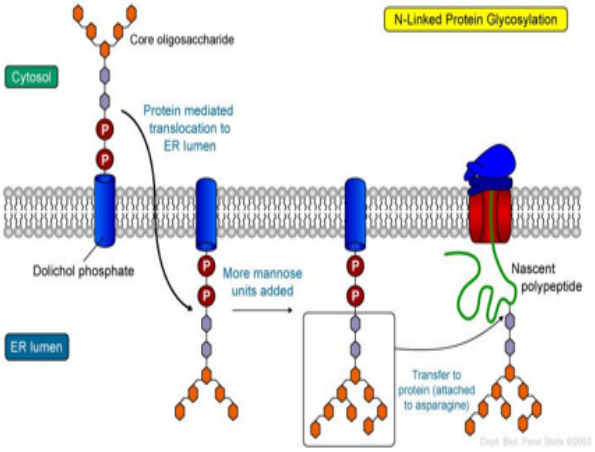
1. N-linked,
2. O-linked,
3. glypiation,
4. C-linked
5. phosphoglycosylation.

This article focuses on the two most frequent types of glycosylation: N-linked and O-linked.

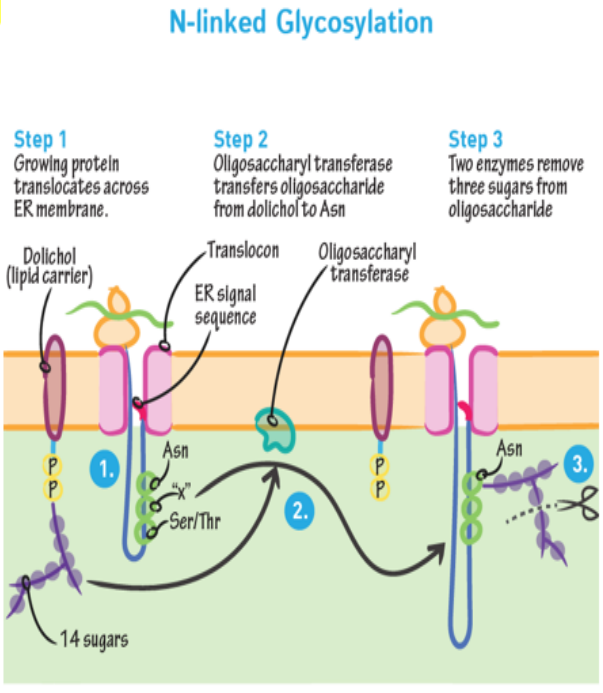


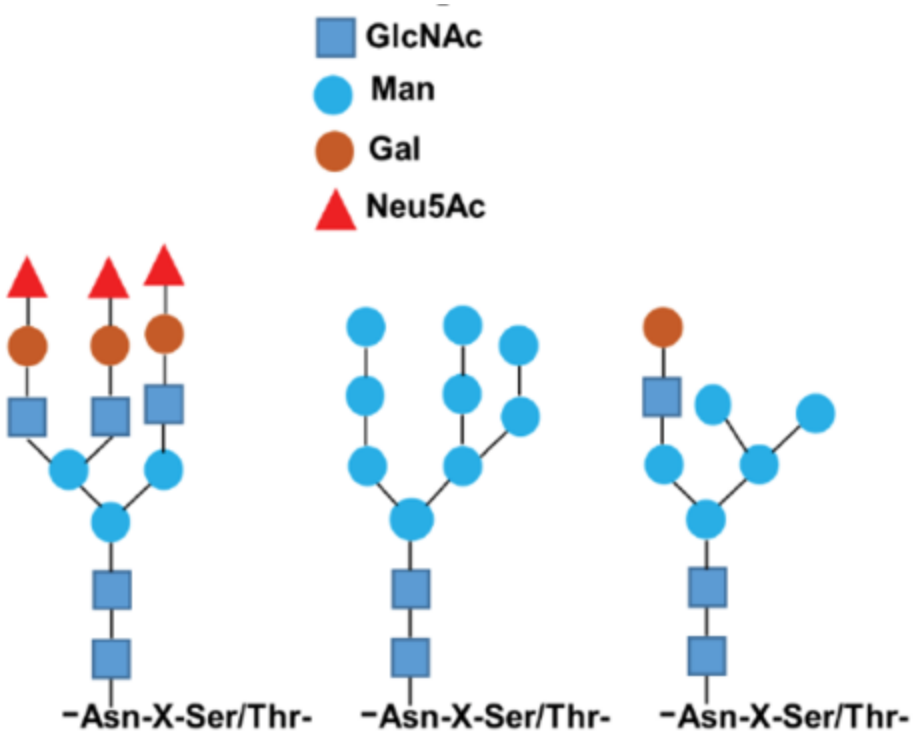
N-linked glycosylation

- N-linked glycosylation is the most frequent type of glycosylation and occurs in eukaryotes and Achaea. Approximately 90% of glycoproteins contain N-linked glycosylation.
- In N-linked **glycosylation, sugars are attached to the amino group of an asparagine (N) side chain**.
- N-linked glycans tend to be large, bulky and diverse, requiring the action of many independent enzymes.



- All N-linked attachments to proteins are initially identical.
- The enzymes that **transfer N-linked glycans** are located in the **lumen of the ER**.
- Upon recognition of the appropriate consensus site, a glycan containing 14 sugar residues is attached en bloc to the protein as it is translated. This glycan is then modified by trimming enzymes.
- When the protein proceeds into the Golgi apparatus, the glycan modifications become diversified by the action of multiple resident glycosyltransferases and glycosidases located in the different stacks.
- Fully glycosylated proteins can contain either complex oligosaccharides containing multiple sugar types, high-mannose oligosaccharides (multiple mannose residues) or a mixture of both.





The structures of N-linked glycosylation proteins. Under the action of glycosidases, N-acetylglucosamine (GlcNAc) and mannose molecules are linked to Asn residues to form the basic N-sugar structures. Galactose (Gal), N-acetylneuraminic acid (Neu5AC), N-acetylgalactosamine (GalNAc) and other molecules participate in the process of glycosylation.

O-linked glycosylation

- O-linked glycosylation is most frequently found on mucins (proteins that form mucous secretions), extracellular matrix components and on antibodies.
- In O-linked glycosylation, a monosaccharide is attached to the hydroxyl group of serine or threonine posttranslationally.
- O-linked glycosylation can also occur on oxidized forms of lysine and proline within certain proteins.
- Sugars are added one at a time by different enzymes and the overall structure of an O-linked glycan is less diverse and complex than an N-linked glycan.
- Although O-linked glycosylation typically occurs in the Golgi apparatus, it has also been reported in the cytosol and nucleus.

Studying glycosylation

Prediction

- Glycosylation enzymes recognize a consensus sequence when initiating glycosylation.
- The consensus sequence for N-linked glycosylation is Asn-X-Ser/Thr (where X is any amino acid except Pro) and more rarely Asn-X-Cys.
- O-linked glycosylation merely requires a serine or threonine without a consensus sequence.
- Protein prediction software can be used to predict potential glycosylation sites on a protein.

Changes in molecular weight

- Glycosylation is usually suspected if a protein contains potential glycosylation sites and migrates slower in a gel than expected based on its predicted molecular weight.
- Often, multiple forms of the protein are observed , with indiscrete higher molecular weight bands due to the presence of multiple glycosylated forms.

Enzymatic and chemical treatment

- Deglycosylating enzymes and chemicals can be used in initial studies to study glycosylation of a protein.
- The protein can then be analyzed by SDS-PAGE and protein staining or Western blotting to detect the protein of interest. A shift in molecular weight upon treatment is indicative of removal of sugar residues. Alternatively, treated samples may be analyzed by mass spectrometry.
- Several enzymes can be used for analysis of glycosylation.
- Endonuclease H removes simple N-linked glycosylations occurring in the ER while PNGase F removes almost all N-linked glycans.
- No single enzyme can cleave O-linked glycans, however the sugars can be trimmed by exoglycosidases, which then makes the glycan susceptible to removal by O-Glycosidase.

- Exposing glycoproteins to an alkaline environment results in sugar release, termed β-elimination. Treatment with sodium hydroxide can result in complete degradation of the glycan. Alternatively, anhydrous hydrazine can be added to a lyophilized glycoprotein to release both N- and O-linked glycans.
- Temperature variation can be used to determine type of glycan is released: O-linked glycans are removed at 60°C while N-linked glycans are released at 95°C. While the glycan is which released intact, the protein may be degraded.

Staining

- Although glycan sugars are usually not reactive to stains, glycoproteins can be detected using the periodic acid-Schiff (PAS) reaction.
- Reaction of sugars with periodic acid oxidizes sugars into aldehydes or ketones, which can be detected using multiple dyes and detected in gels or on membranes.
- The reaction can also be used to couple glycosylated proteins to labels, such as biotin or horseradish peroxidase, for other detection methods.
- Although this method can be used to determine if a protein is glycosylated it does not discriminate between the different types of glycosylation

Affinity detection

Lectins

Lectins are proteins found in animals, plants and microorganisms that specifically bind to sugar molecules. More than 2000 lectins have been described and many are commercially available. Different lectins have various specificities and can be used to detect different types of glycosylation. However, lectins have different affinities for various sugars and recognition can overlap. Lectins can be conjugated to probes for use in blotting or immunohistochemistry or immobilized to solid supports for affinity purification.

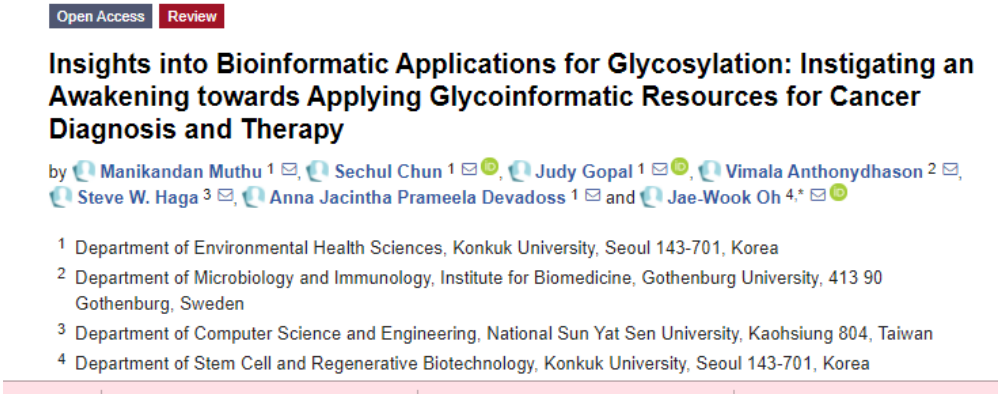
Antibodies

Anti-glycan antibodies are used to study glycosylation of proteins. Antibodies can have greater specificity than lectins and target particular linkages. Commercial antibodies are available that

recognize all 5 types of glycosylation linkages. The antibodies can be used in any immunodetection assay (e.g. Western blot, FACS, etc)

Mass spectrometry

Mass spectrometry can be used to characterize and quantitate the glycans associated with proteins. However, glycans are extremely variable and often have similar mass, making analysis by mass spec challenging. Often, multiple types of analyses are required. To study the linkage site, the protein and associated glycans is often left intact. In other studies, the glycans are removed sequentially by enzymatic and chemical treatments and the glycans are analyzed separately from the protein.



Open Access Review

Insights into Bioinformatic Applications for Glycosylation: Instigating an Awakening towards Applying Glycoinformatic Resources for Cancer Diagnosis and Therapy

by Manikandan Muthu [1], Sechul Chun [1], Judy Gopal [1], Vimala Anthonydhason [2], Steve W. Haga [3], Anna Jacintha Prameela Devadoss [1] and Jae-Wook Oh [4,*]

[1] Department of Environmental Health Sciences, Konkuk University, Seoul 143-701, Korea
[2] Department of Microbiology and Immunology, Institute for Biomedicine, Gothenburg University, 413 90 Gothenburg, Sweden
[3] Department of Computer Science and Engineering, National Sun Yat Sen University, Kaohsiung 804, Taiwan
[4] Department of Stem Cell and Regenerative Biotechnology, Konkuk University, Seoul 143-701, Korea

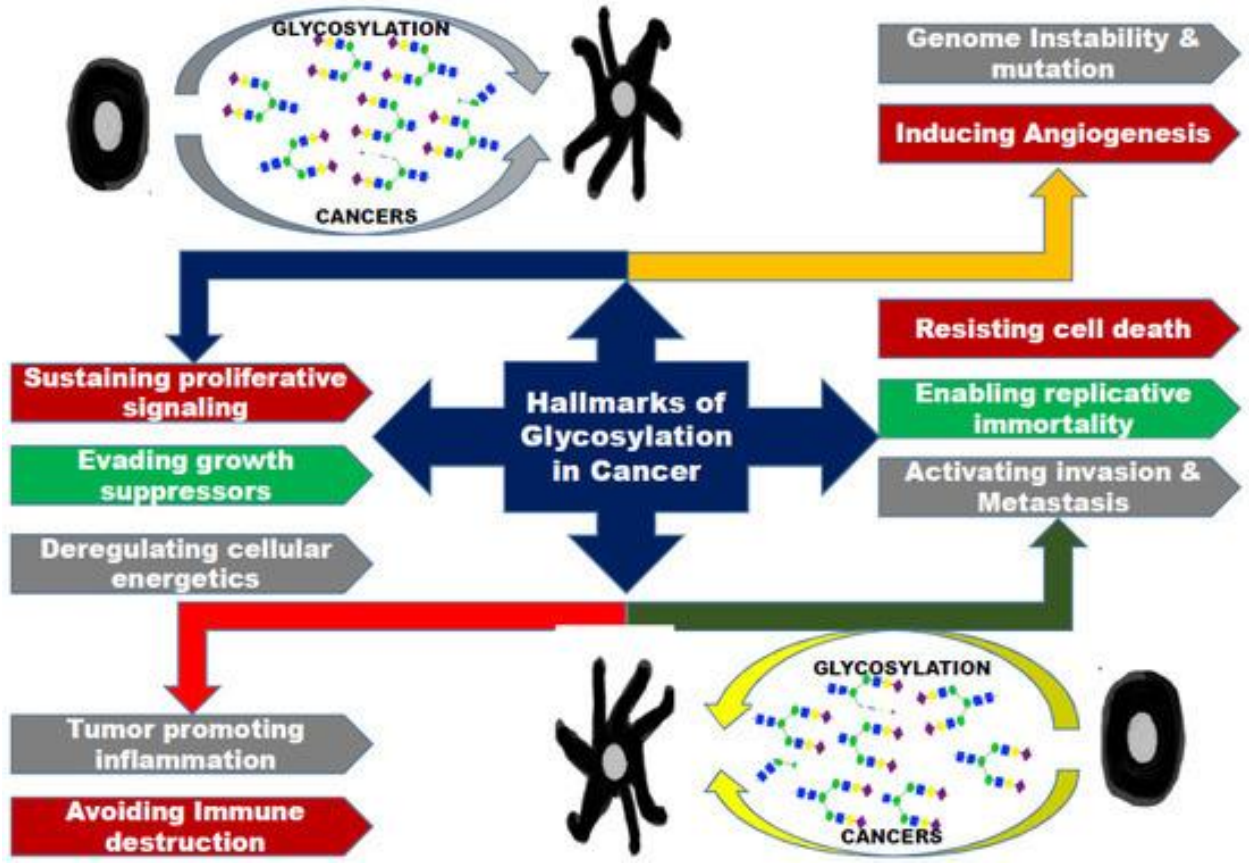
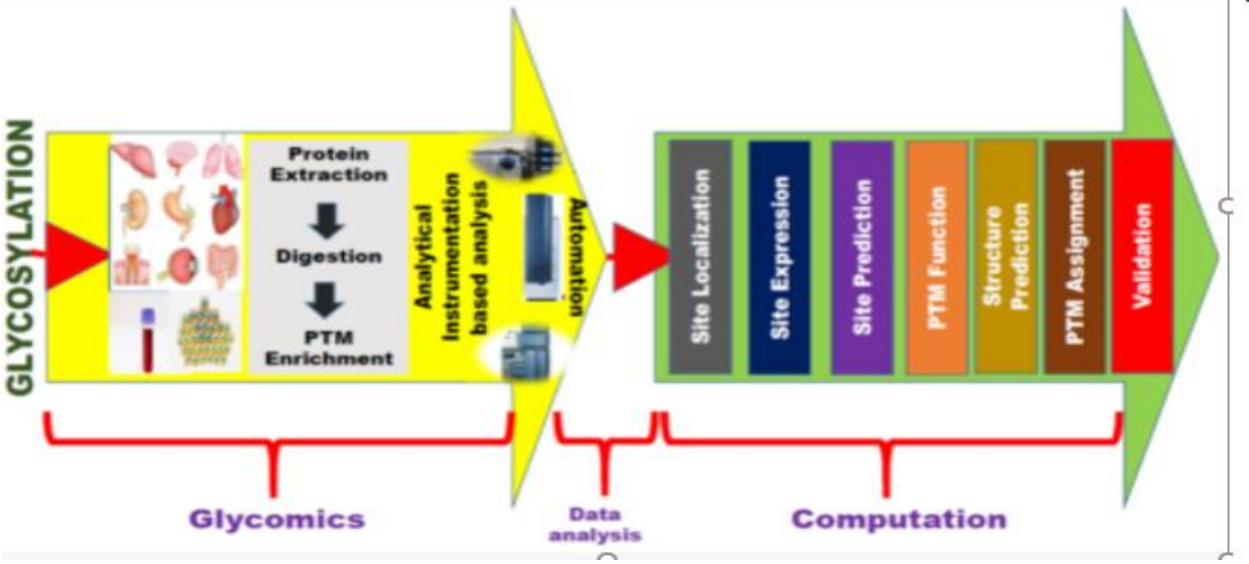


Figure 6 Work-flow of glycosylation research, beginning from wet lab analysis—instrumentational analysis culmination in bioinformatics based computation (data analysis).



Bioinformatic tool applications in various Glycosylation associated studies

Bioinformatics Resources	URL	Application
	Glycosylation Structure Related Resources	
UniCarb KB	http://unicarbkb.org/	Repository for glycan structures of glycoproteins
GlycoMod	http://web.expasy.org/glycomod/	Software tool for N-linked and O-linked glycan structures prediction
GlycosuiteDB	http://www.glycosuite.com	Relational database of glycoprotein glycan structures and their biological sources
GlycoDeNovo	https://www.cs.brandeis.edu/~hong/Research/GlycoDeNovo/GlycoDeNovo.htm	Algorithm for Accurate de novo Glycan Topology Reconstruction from Tandem Mass Spectras
Glycoforest	https://glycoforest.expasy.org/	Partial de-novo algorithm for sequencing glycan structures based on MS/MS spectra
Sweet-II	http://www.glycosciences.de/modeling/sweet2/doc/index.php	Tool to construct 3D models of saccharides from their sequences using standard nomenclature
GlyProt	http://www.glycosciences.de/modeling/glyprot/php/main.php	Tool to connect N-glycans in silico to a given 3D protein structure.

Bioinformatic tools for glycosylation in cancer applications.

Bioinformatics Resources	Cancer Type	URL	Application	F
RNA sequencing analysis of genes	prostate cancer cell lines and patients	A composite sequencing server	RNA sequencing analysis of identified a set of 700 androgen-regulated genes.	
Gene ontology (GO)	prostate cancer cell lines and patients	Gene ontology (GO) is general functional annotation server	identified 72 terms with significant gene enrichment ($p < 0.05$) and defined glycosylation as an androgen-regulated process in prostate cancer cells.	
GlycoBase (inactive now)	Breast cancer cells	(https://glycobase.nibrt.ie):	database of experimentally determined glycan structures originally developed from the EurocarbDB project	
GlycoDigest	Breast cancer cells	(http://www.glycodigest.org):	a tool that simulates exoglycosidase digestion based on controlled rules acquired from expert knowledge and experimental evidence available in GlycoBase	
GlycoMarker	Breast cancer cells	https://glycobase.nibrt.ie/glycomarker	Web application/server for Biomarker discovery, identifies markers in LC profiles	
using KEGG, DAVID and Ingenuity databases, uniprot database	Breast cancer cells	KEGG, DAVID and Ingenuity databases, uniprot database these are very general functional analysis and sequence databases	Significant change in the expression profiling of glycosylation patterns of various proteins associated with Triple negative breast cancer was identified. Differential aberrant glycosylated proteins in breast cancer cells with respect to non-neoplastic cells	

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOINFORMATICS

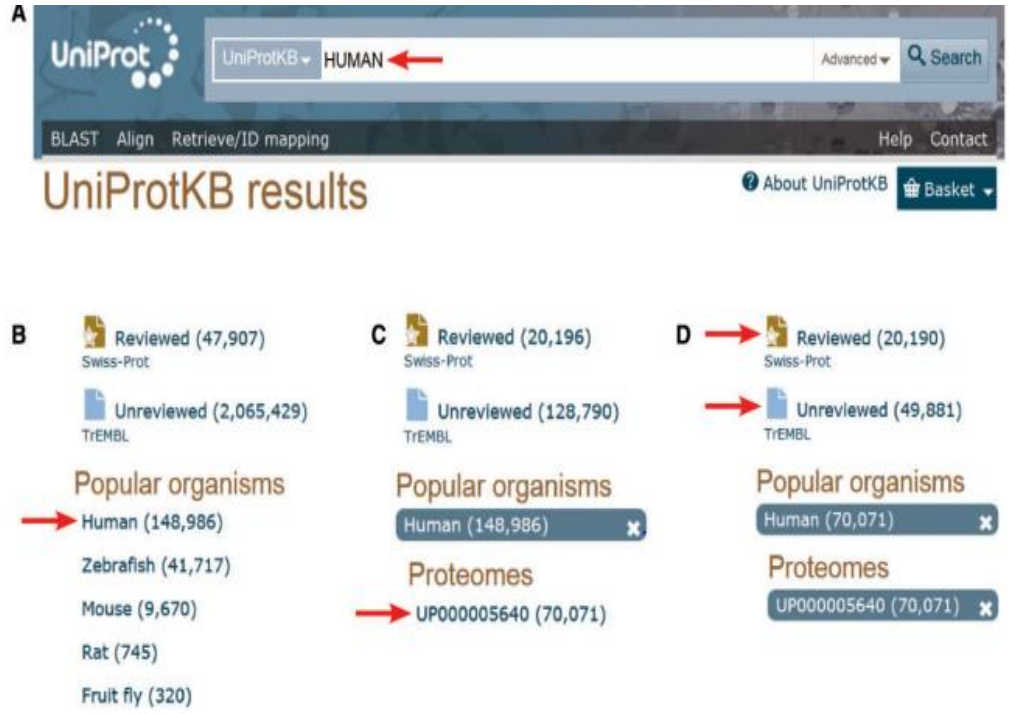
UNIT – 5- SBIA1303 – Proteomics and Interactomics

Unit V Developments in Proteomics

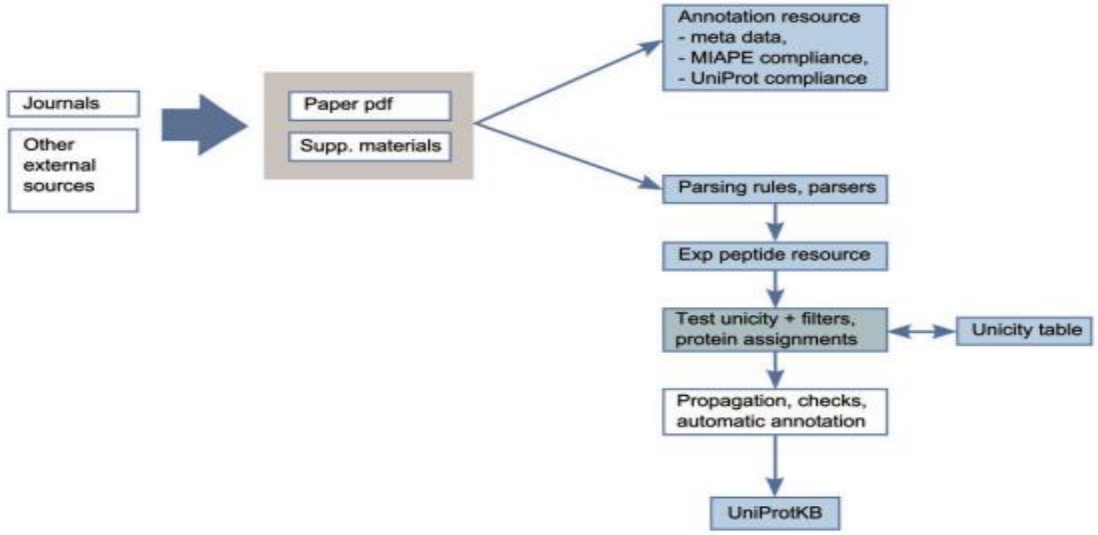
- Proteome databases.
- Proteome project.
- Comparative proteomics, Endogenous peptides and Peptidomics.
- Metaproteomics, Proteogenomics

The UniProtKB guide to the human proteome

Advances in high-throughput and advanced technologies allow researchers to routinely perform whole genome and proteome analysis. For this purpose, they need high-quality resources providing comprehensive gene and protein sets for their organisms of interest. Using the example of the human proteome, we will describe the content of a complete proteome in the UniProt Knowledgebase (UniProtKB). We will show how manual expert curation of UniProtKB/Swiss-Prot is complemented by expert-driven automatic annotation to build a comprehensive, high-quality and traceable resource. We will also illustrate how the complexity of the human proteome is captured and structured in UniProtKB.



Accessing the human proteome from the UniProt web site (http://www.uniprot.org). A. One can directly type ‘HUMAN’ in the search box. B. Then select ‘Human’ in the ‘Popular organisms’ section on the left.C. There is a single proteome in the ‘Proteomes’ section for this organism, UP000005640, a direct link allowing access to the entries composing the human complete proteome. D. There, one still has the possibility to select the 20 197 expertly ‘Reviewed’ entries of the Swiss-Prot section of UniProtKB from the 49 496 additional ‘unreviewed’ entries of UniProtKB/TrEMBL.



Annotation scheme for integration of proteomics data in UniProtKB.

WHAT IS OPENPROT?

OpenProt is the first database that enforces a polycistronic model of eukaryotic genome annotations. Thus, OpenProt annotates known proteins (called RefProts) but also novel isoforms and novel proteins (called altProts). It also provides supporting evidence for each protein, such as mass-spectrometry and ribosome profiling detection, protein homology and predictions of funtional domains.

WHY USE OPENPROT?

Annotations substantially shape today's Research by drawing the scope of possibilities. When using OpenProt, you will gain a better view of the proteomic complexity incumbent to each gene and each transcript. By gathering experimental evidence, OpenProt is a data-driven protein database. All data are freely available and can be downloaded for in-house analyses.

HOW TO USE OPENPROT?

OpenProt offers multiple downloads, in particular for mass-spectrometry based proteomics analyses, as well as a search page and a genome browser that allows users to interrogate the database.

THE CONCEPT BEHIND OPENPROT

The current annotation model

Current genome annotations hold arbitrary criteria for Open Reading Frame (ORF) annotation, such as:

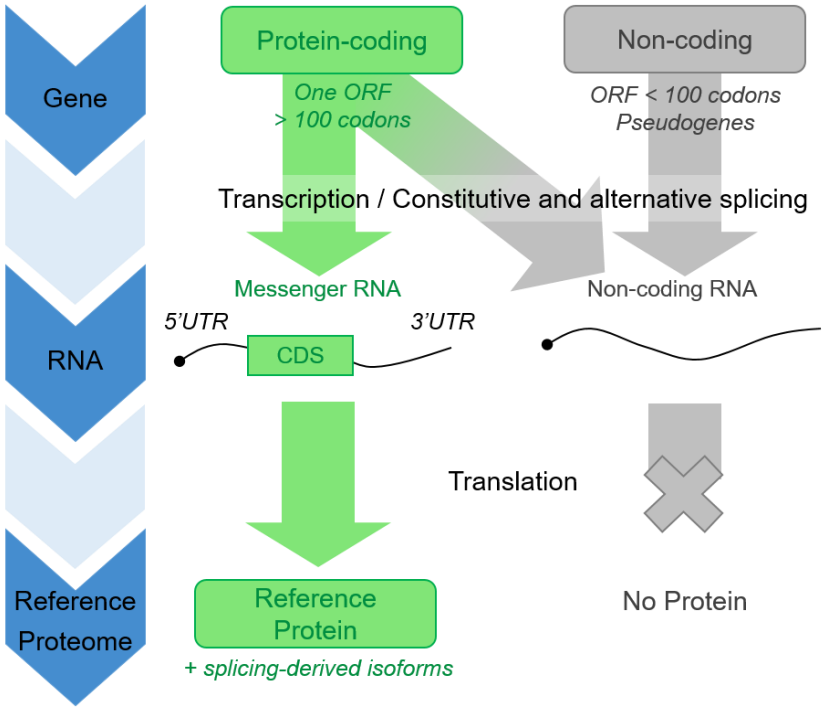
a minimal ORF length of 100 codons;

a single ORF per transcript (monocistronic);

transcripts that do not meet the above criteria are non-coding (ncRNAs);

transcripts from unprocessed pseudogenes are non-coding (ncRNAs).

With the rare exception of previously characterized examples, these rules are applied and considerably shape the annotated protein landscape (Reference Proteome). Yet, a wealth of experimental data highlights the pitfalls of such annotation model.

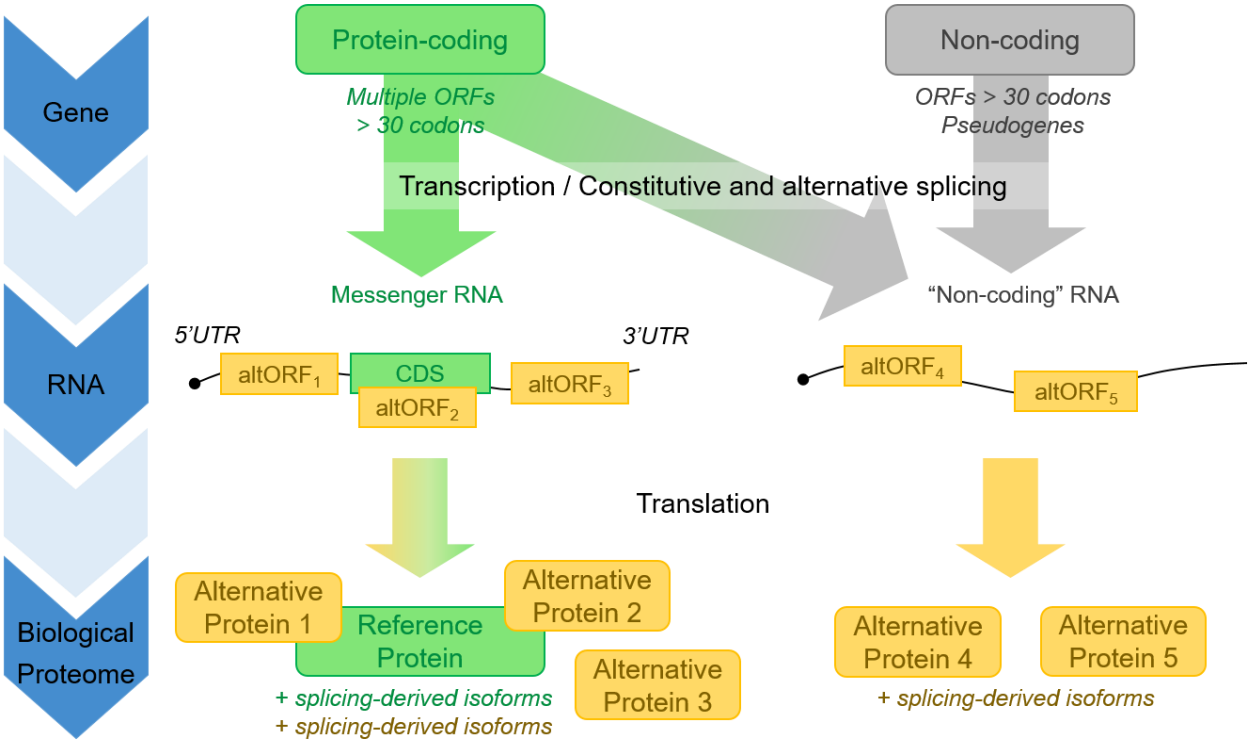


The OpenProt annotation model

OpenProt challenges the aforementioned arbitrary criteria, and thus:

- annotates any ORF longer than 30 codons;
- annotates multiple ORF per transcript (polycistronic);
- annotates ORFs within ncRNAs;
- annotates ORFs within pseudogenes transcripts.

Thus OpenProt annotates known proteins (Reference Proteins or RefProts), novel isoforms and novel proteins (Alternative Proteins or AltProts). It offers a deeper, and more realistic, view of the proteome (Biological Proteome).



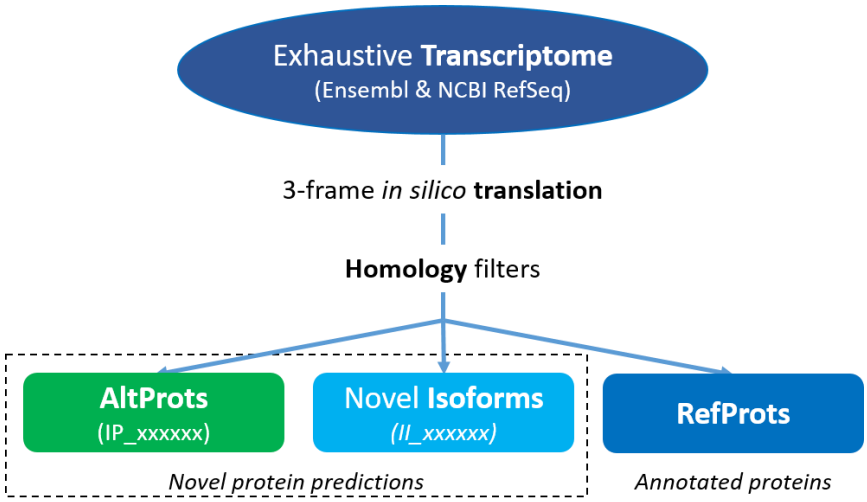
OPENPROT: HOW DOES IT WORK?

OpenProt prediction pipeline

The OpenProt prediction pipeline first retrieves transcripts (RNAs) from two well-used annotations (Ensembl and NCBI RefSeq). This constitutes an exhaustive transcriptome.

A 3-frame in silico translation then yields the ORFeome: any ORF longer than 30 codons in any frame of any transcript. This ORFeome is then filtered to categorize predicted ORFs. The first filter retrieves all known protein (all ORF already annotated in Ensembl, NCBI RefSeq, and/or UniProtKB), these are the RefProts. The second filter looks at the homology of the currently not annotated ORFs with the RefProt of the same gene (if applicable), and retrieves novel predicted isoforms.

The remaining ORFs encode novel proteins, called AltProts.



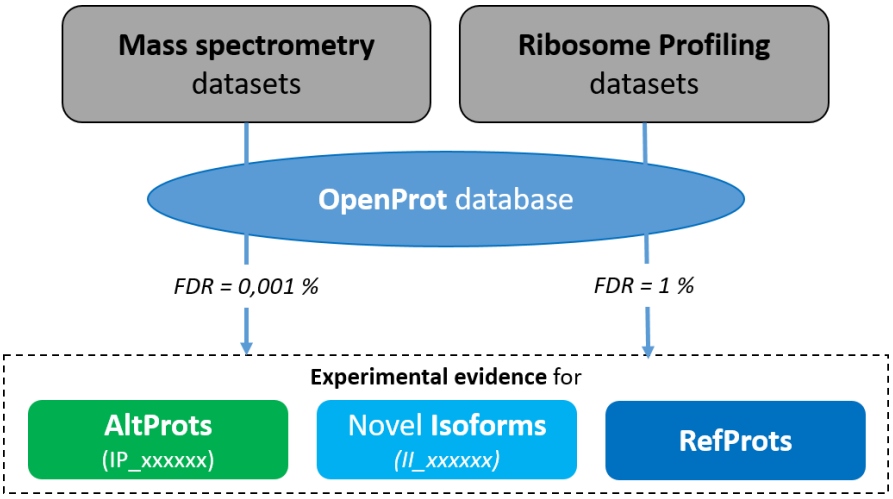
OpenProt evidence pipeline

To increase confidence in ORF expression, since random ORFs are a possibility, OpenProt also cumulates several lines of evidence, such as:

Conservation evidence: for every ORF annotated, OpenProt identifies orthologs and paralogs (across the 10 species currently supported by OpenProt).

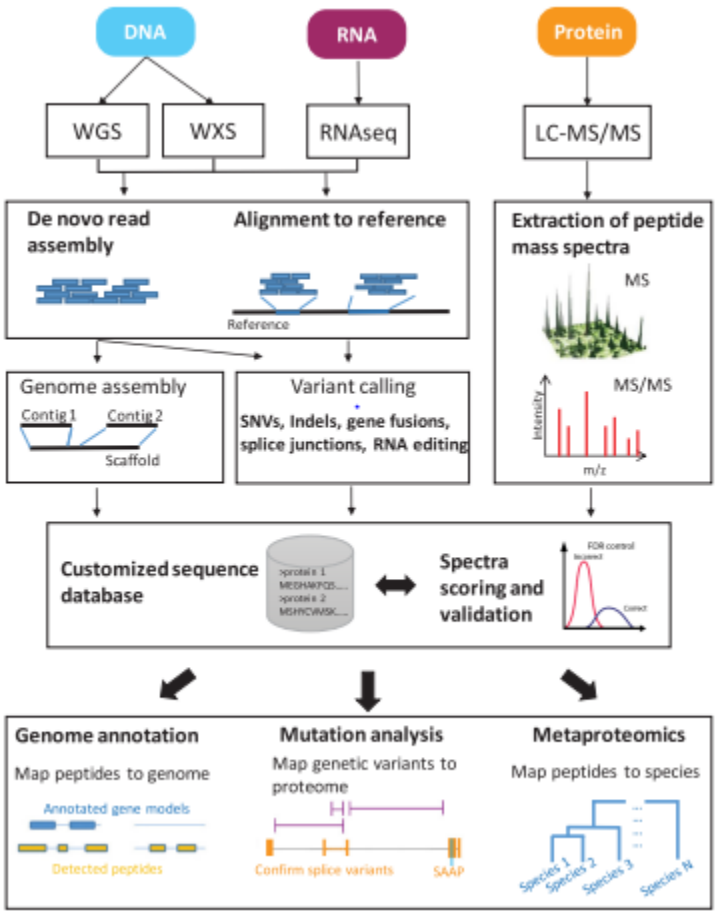
Translation evidence: OpenProt retrieves publicly available ribosome profiling datasets and re-analyses them using the Price algorithm with a stringent 1 % FDR. This gathers translation evidence for any ORF annotated in OpenProt.

Expression evidence: OpenProt retrieves publicly available mass spectrometry datasets and re-analyses them using multiple search engines, and a stringent 0,001 % FDR. This gathers expression evidence for any ORF annotated in OpenProt.

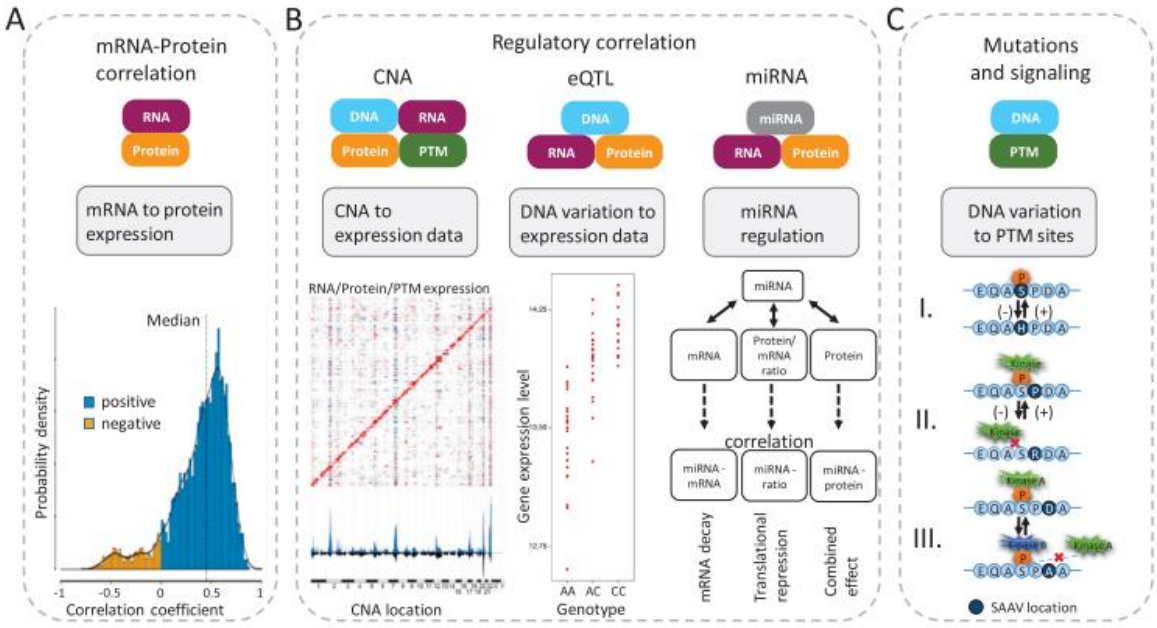


Methods, Tools and Current Perspectives in Proteogenomics

Sequence-centric proteogenomics



Sequencing-based technologies to sequence DNA (whole genome sequencing, WGS; whole exome sequencing, WXS) and RNA (RNA-seq) generate millions of short sequencing reads that are assembled into genomes, exomes or transcriptomes by either de novo or template-based approaches by alignment to a reference sequence. Sample-specific sequence aberrations are determined and nucleotide sequences are transformed into personalized, amino acid-centric sequence databases. Peptide mass spectra derived by LC-MS/MS analysis from a matching sample are then scored and validated against the personalized database enabling the detection of sample-specific peptide sequences. Depending on the scope of the proteogenomic project, these peptides can then be used to (1) aid genome annotation by detection of peptides in unannotated genome regions; (2) identify tumor-specific mutations translated into the proteome as well as novel protein splice variants; and (3) detect species-specific peptides in microbial communities.



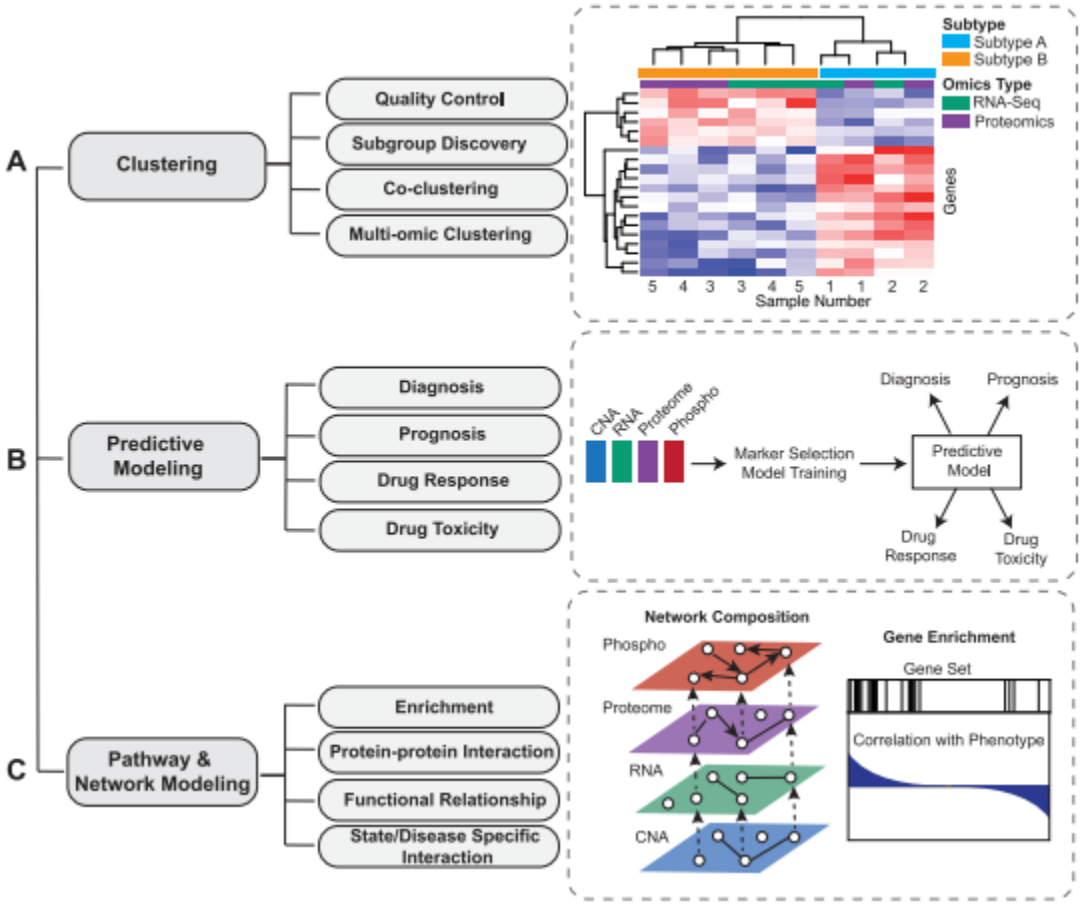
Proteogenomic relationships

A, Correlation analysis of mRNA and protein pairs across samples enables the assessment of global correlation structure which typically centers between correlation coefficients of 0.3 and 0.5.

B, Regulatory effects on RNA and protein expression levels caused by copy number aberrations (CNA), genetic variants (eQTL) and microRNAs (miRNAs) can be studied by different correlation-based approaches. CNA cis and trans effects on RNA, protein and PTM expression can be determined by correlating each gene copy number at a given locus to all quantified features in RNA, protein or PTM space across all samples. Expression quantitative trait loci (eQTL) analysis can be used to identify DNA sequence

variants affecting RNA/protein expression levels in the sample population being studied. Global miRNA analysis accompanied with mRNA or protein profiling enables the assessment of miRNA mediated regulation of mRNA and protein expression.

C, Integrative analysis of genetic variants and PTM sites like phosphorylation can identify functional consequences of genetic variants at the molecular level. Mutations that directly affect serine, threonine and tyrosine residues can result in destruction or genesis of phosphosites (I); mutations adjacent to phosphosites can result in removal or addition of phosphosites (II) or change the kinase that recognizes the phosphorylation site (III).



Integrative modeling.

Overview of sub-topics in integrative modeling of proteogenomic data. A, Clustering techniques illustrating a schematic of multi-omic hierarchical clustering analysis resulting in the identification of two subtypes, B, Predictive modeling for disease diagnosis, prognosis, drug response and drug toxicity using multiple data

modalities and, C, proteogenomic pathway and network modeling, including informing network composition and pathway and GO term enrichment.

Computational frameworks and resources for intersecting PTMs and mutations

Name	URL	Reference	Remarks
PhosphoPOINT	http://kinase.bioinformatics.tw*	(116)	Human kinase interactome and phospho-protein database
PhosSNP	http://phossnp.biocuckoo.org/	(15)	Database of mutations predicted to impact phosphorylation status of proteins
PhosphoVariant	NA	(14)	Database for definite and possible variants changing phosphosites
PTMvar	phosphosite.org	(122)	Database intersecting non-synonymous SNPs and PTM sites
ActiveDriver	http://individual.utoronto.ca/reimand/ActiveDriver/	(121)	Prediction of 'active' phosphosites in proteins that are specifically and significantly mutated in cancer genomes
ReKINect	http://rekinect.science/home	(16)	Prediction of network attacking mutations (NAMs) from NGS data
MIMP	http://mimp.baderlab.org/	(203)	Characterization of genetic variants that specifically alter kinase-binding sites in proteins
g2pDB	www.g2pdb.org	(115)	Database of auto-curated PTM sites mapped to their genomic locations

Metaproteomics

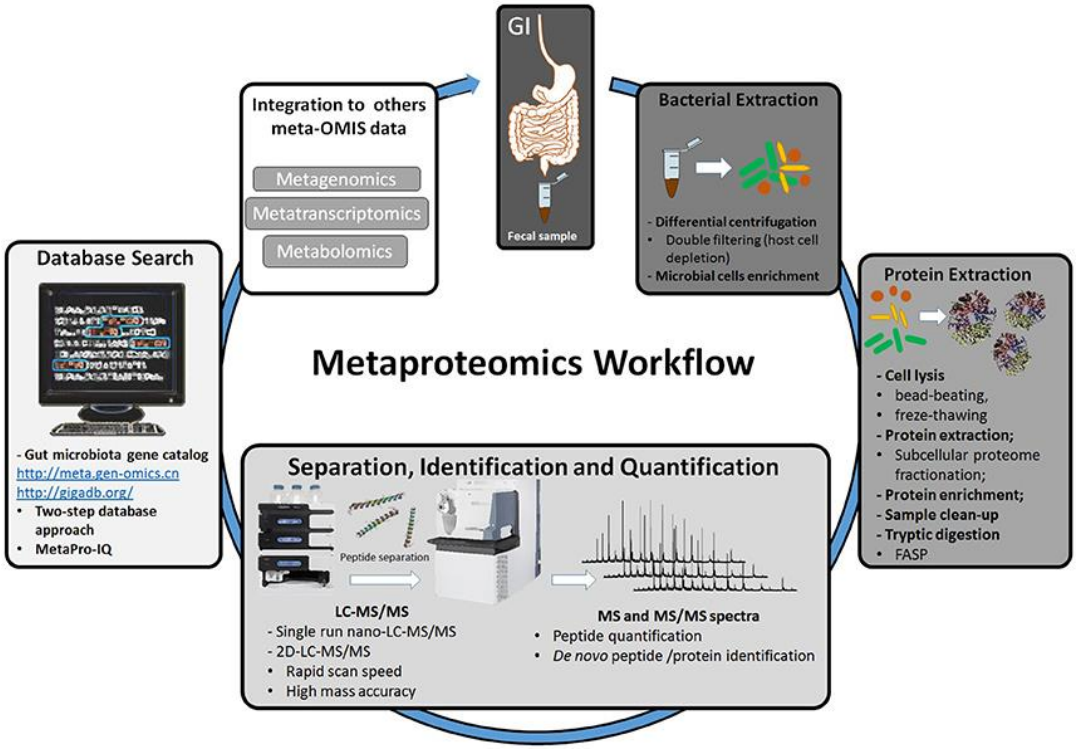
Metaproteomics (also Community Proteomics, Environmental Proteomics, or Community Proteogenomics) is an umbrella term for experimental approaches to study all proteins in microbial communities and microbiomes from environmental sources.

What is the purpose of metaproteomics?

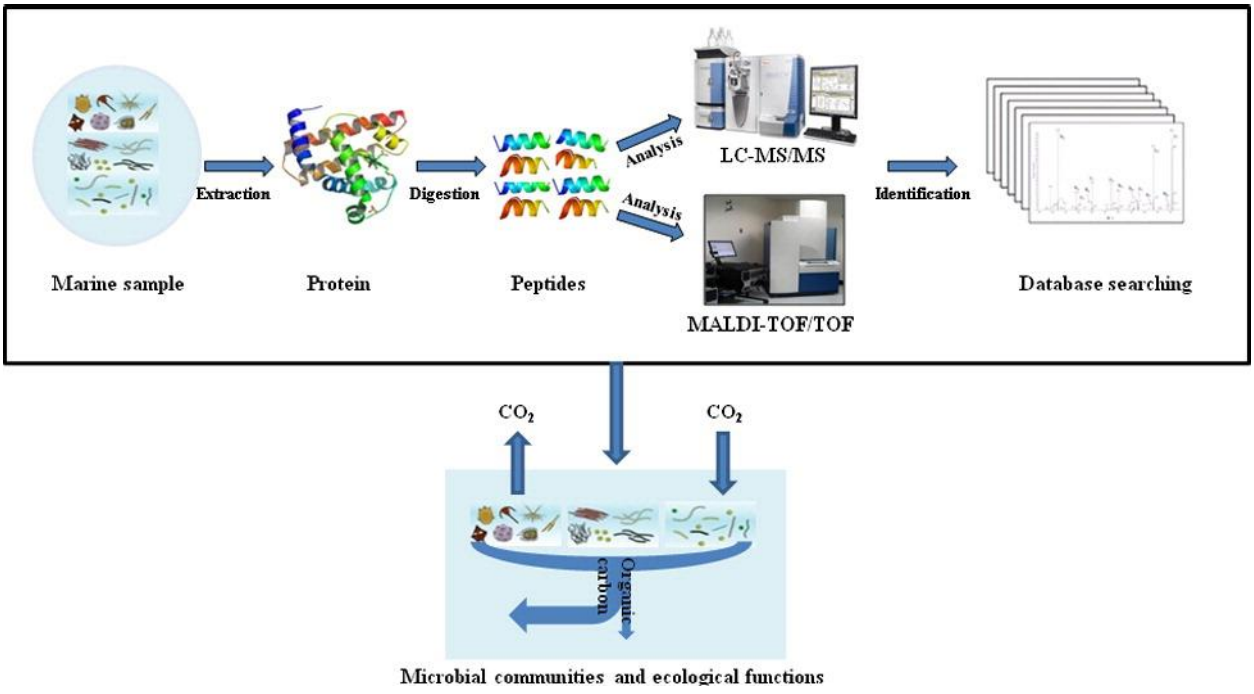
Metaproteomics is the large-scale identification and quantification of proteins from microbial communities and thus provides direct insight into the phenotypes of microorganisms on the molecular level.

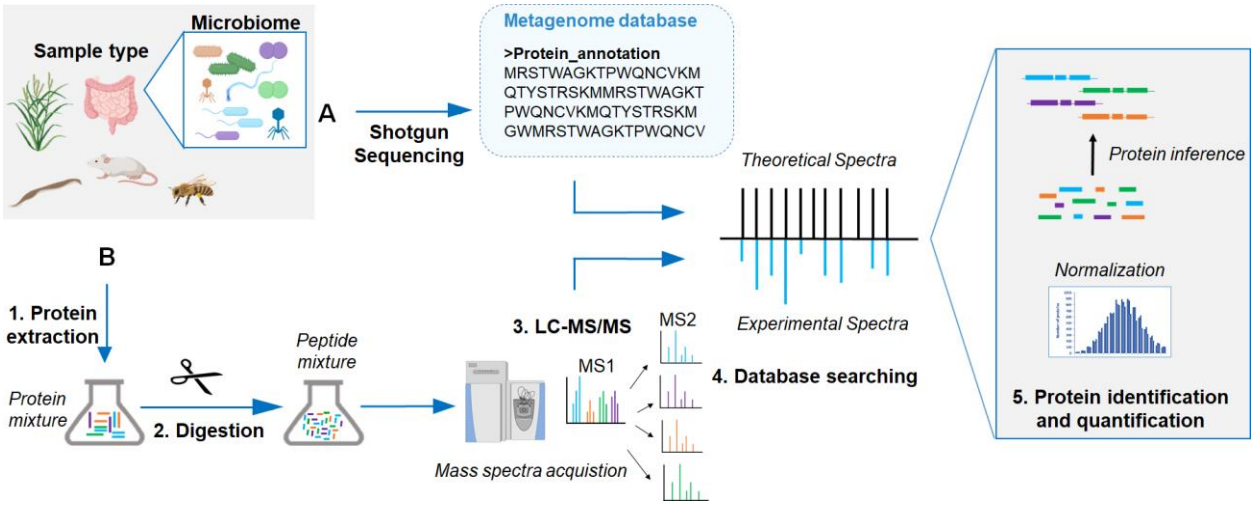
- Metaproteomics provides direct evidence of physiological and metabolic activities of a complex system
- Metaproteomics has shown its powerful potential in the study of marine ecosystem.
- The application of metaproteomic approaches to complex marine samples still faces considerable challenges.
- Metaproteomics is a new field within the 'omics' science which investigates protein expression from a complex biological system and provides direct evidence of physiological and metabolic activities. Characterization of the metaproteome will enhance our

understanding of the microbial world and link microbial communities to ecological functions. Recently, the availability of extensive metagenomic sequences from various marine microbial communities has extended the postgenomic era to the field of oceanography. Although still in its infancy, metaproteomics has shown its powerful potential with regard to functional gene expression within microbial habitats and their interactions with the ambient environment as well as their biogeochemical functions.



Representation of a workflow in a metaproteomic analysis of fecal sample extracted from the gut, starting from the bacterial extraction process, followed by microbial protein extraction and enrichment, protein and peptide separation, identification, and quantification and MS/MS search against genome databases. FASP; filtered-aided sample preparation.





General workflow employed in metaproteomics experiments.

(A) Metagenome shotgun sequencing can be used to generate the reference database for metaproteomics. (B) Metaproteomics workflow including generation of peptides and high-resolution MS analysis. LC, liquid chromatography; MS, mass spectrometry.