Paper II

UNIT 1.1 & 1.3

Q1. Why is structural Bioinformatics studied so widely?

- ➔ Structural bioinformatics, originally known as structural computational biology, predates other forms of bioinformatics. It can be argued that the seminal 1953 article by Watson and Crick is in fact a modelling paper and arguably the first structural bioinformatics paper.
- ➔ Structural Bioinformatics or structural computational biology broadly defined, is a field at the intersection between computer science, physics, chemistry and molecular biology. Historically, the term 'structural bioinformatics' describes data-driven statistical, knowledge-based research of representative non-redundant ensembles of structures to understand the statistical behaviour of the system under investigation.
- ➔ Biologically, structural bioinformatics aims to understand the factors that influence and determine the function of biological macromolecules, the interplay between evolution, kinetics and thermodynamics, the determinants of specificity and selectivity in molecular interactions, the dynamic aspects of macromolecular structures and their effect on function and stability and, finally, the ability to use all these for engineering, design and biotechnology.

Q2. Describe a case study where it states the importance of structural bioinformatics studies.

- ➔ SARS-CoV-2 is a single stranded positive sense RNA virus, belonging to the genus Betacoronavirus and the family *Coronaviridae.* SARS-CoV-2 and two related coronaviruses namely SARS-CoV and MERS-CoV are known to cause severe disease in humans.
- ➔ Since the first sequenced genome of SARS-CoV-2 isolated in Wuhan city, China, over 90,000 genome sequences have been deposited in Global Initiative on Sharing All Influenza Data. Three-dimensional structures were rapidly solved for the key target proteins in SARS-CoV-2 and host proteins, namely the spike protein, RNA-dependent RNA polymerase (RdRp), main protease, Papain-like protease and human angiotensin-converting enzyme 2.
- ➔ Understanding the functional role of 3D structures is essential to understand viral evolution and transmission, as well as to guide therapeutic research. Structural biology has contributed to the successful development of antiviral therapeutics for other human diseases including HIV and influenza.
- ➔ In absence of vaccines against SARS-CoV-2, structure-guided approaches are being used to identify promising candidates for antigens, particularly from the spike protein which is responsible for mediating infection. Several potential drug molecules have been recently proposed for re-purposing based on protein-structure based algorithms.

Q3. List and describe the challenges faced in structural bioinformatics studies.

- ➔ Although considerable progress has been made, advances are still needed in the following areas:
- ➔ **Modelling large or multi-domain proteins and assemblies:** Most proteins are large and multi-domain. Targeting such complexes and assemblies experimentally and looking at them in the context of the complete cell is an emergent challenge.
- ➔ **Biomolecules as dynamic objects:** The accurate modelling of large conformational changes due to ligand binding, allosteric effects, post-translational modifications or as the result of protein-protein interactions. The use of such conformational ensembles as opposed to single static structures is challenging for complex structures.
- ➔ **Modelling 3D RNA structures**: The production of RNA structures is still in its infancy and often requires processing low resolution data.
- ➔ **Small Differences may have drastic effects:** Although, protein structure is resilient to mutation, function is not necessarily as resilient. This combination of structural robustness and functional plasticity is at the core of evolutionary change.

Q4. Discuss a case study where it states integration of structural data with other relevant fields.

- ➔ To integrate heterogeneous and large omics data constitutes not only a conceptual challenge but a practical hurdle in the daily analysis of omics data. With the rise of novel omics technologies and through large-scale consortia projects, biological systems are being further investigated at an unprecedented scale generating heterogenous and often large data sets.
- ➔ Data integration is now a very commonly used notion in life sciences research. However, there is still no unified definition of data integration, nor taxonomy for data-integration methodologies despite some recent efforts on this topic.
- ➔ Research in life sciences has the generic goal to identify the components that make up a living system and to understand the interactions among them that results in the functioning system. Collection of biological data is therefore a method to catalogue the elements of life, but the understanding of a system requires the integration of these data under mathematical and relational models that can describe mechanistically the relationships between their components.
- ➔ We can illustrate the state of affairs on data integration in life science research using a simple example taken from metabolic modelling.
- ➔ In the study of GLY (Glycolysis pathway), G1 is considered to be known as there are a detailed set of genes. Proteins and metabolites already described; however, we are not yet certain that this list contains all involved elements, for example the list does not incorporate the epigenetic marks that may be associated to the regulation of GLY.
- ➔ Even though the system-defined as a set of interactions is fully known, however, the pathway elements and relations may be missing and this does not allow us to determine completeness.
- ➔ The present situation is very fortunate since over the last decades several different types of data were generated and huge efforts were dedicated to create database repositories for different data-types where investigators were encouraged to deposit and share datasets associated with scientific publications.

Q5. To understand the Macromolecular interactions in below given sets:

a) Protein – Protein interaction

- ➔ Protein interactions are fundamentally characterized as stable or transient, and both types of interactions can be either strong or weak. Stable interactions are those associated with proteins that are purified as multi- subunit complexes, and the subunits of these complexes can be identical or different. Haemoglobin and core RNA polymerase are examples of multi-subunit interactions that form stable complexes**.**
- ➔ Proteins bind to each other through a combination of hydrophobic bonding, van der Waals forces, and salt bridges at specific binding domains on each protein. These domains can be small binding clefts or large surfaces and can be just a few peptides long or span hundreds of amino acids. The strength of the binding is influenced by the size of the binding domain.
- ➔ One example of a common surface domain that facilitates stable protein–protein interactions is the leucine zipper, which consists of α-helices on each protein that bind to each other in a parallel fashion through the hydrophobic bonding of regularly-spaced leucine residues on each α- helix that project between the adjacent helical peptide chains. Because of the tight molecular packing, leucine zippers provide stable binding for multi-protein complexes, although all leucine zippers do not bind identically due to non-leucine amino acids in the α-helix that can reduce the molecular packing and therefore the strength of the interaction.
- ➔ Two Src homology (SH) domains, SH2 and SH3, are examples of common transient binding domains that bind short peptide sequences and are commonly found in signalling proteins. The SH2 domain recognizes peptide sequences with phosphorylated tyrosine residues, which are often indicative of protein activation. SH2 domains play a key role in growth factor receptor signalling, during which ligand-mediated receptor phosphorylation at tyrosine residues recruits downstream effectors that recognize these residues via their SH2 domains
- ➔ The SH3 domain usually recognizes proline-rich peptide sequences and is commonly used by kinases, phospholipases and GTPases to identify target proteins. Although both SH2 and SH3

domains generally bind to these motifs, specificity for distinct protein interactions is dictated by neighbouring amino acid residues in the respective motif.

b) Post translational modification: Phosphorylation, ubiquitination, and glycosylation

- ➔ There are many types of protein modification, which are mostly catalysed 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 signalling. The most common PTMs are:
- ➔ **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 signalling 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 tumour suppressor protein. p53 is used in cancer therapeutics and is activated by phosphorylation of its N-terminal by several kinases.
- ➔ **Glycosylation:** Glycosylation involves addition of an oligosaccharide termed 'glycan' to either a nitrogen atom (N-linked glycosylation) or an oxygen atom (O-linked glycosylation). N-linked glycosylation occurs in the amide nitrogen of asparagine, while the O-linked glycosylation occurs on the oxygen atom of serine or threonine. Carbohydrates present in the form of N-linked or O-linked oligosaccharides are present on the surface of cells and secrete proteins. They have critical roles in protein sorting, immune recognition, receptor binding, inflammation, and pathogenicity. For example, N-linked glycans on an immune cell can dictate how it migrates to specific sites. Similarly, it can also determine how a cell recognizes 'self' and 'non-self'.
- ➔ **Ubiquitination:** Ubiquitination involves addition of a protein found ubiquitously, termed 'ubiquitin', to the lysine residue of a substrate. Either a single ubiquitin molecule (monoubiquitinating) or a chain of several ubiquitin molecules may be attached (polyubiquitination). Polyubiquitinated proteins are recognized by the 26S proteasome and are subsequently targeted for proteolysis or degradation. Monoubiquitinated proteins may influence cell tracking and endocytosis.

c) Protein – Nucleic acids interactions

- ➔ Proteins interact with DNA and RNA through similar physical forces, which include electrostatic interactions (salt bridges), dipolar interactions (hydrogen bonding, H-bonds), entropic effects (hydrophobic interactions) and dispersion forces (base stacking). These forces contribute in varying degrees to proteins binding in a sequence-specific (tight) or non–sequence-specific (loose) manner.
- ➔ The DNA- or RNA-binding function of a protein is localized in discrete conserved domains within its tertiary structure. An individual protein can have multiple repeats of the same nucleic acid binding domain or can have several different domains found within its structure. The identity of the individual domains and their relative arrangement are functionally important within the protein. Several common DNA binding domains include zinc fingers, helix-turn-helix, helix-loop-helix, winged helix and leucine zipper.
- ➔ RNA-binding specificity and function are constituted by zinc finger, KH, S1, PAZ, PUF, PIWI and RRM (RNA recognition motif) domains. Multiple nucleic acid binding domains with a single protein can increase specificity and affinity of the protein for certain target nucleic acid sequences, mediate a change in the topology of the target nucleic acid, properly position other

nucleic acid sequences for recognition or regulate the activity of enzymatic domains within the binding protein.

- ➔ **Protein–DNA interactions:** The common DNA-binding domains, helix-turn-helix and zinc finger domains, are incorporated within numerous DNA-binding proteins expressed in the cell. Specificity is derived from higher order interactions involving nucleoprotein complexes. These DNA-binding protein complexes find their target by "sliding" along the genomic DNA until their specific DNA-docking site is discovered. The binding of protein to DNA controls the structure of genomic DNA (chromatin), RNA transcription, and DNA repair mechanisms.
- ➔ **Protein–RNA interactions:** Proteins interact with RNA in order to splice, protect, translate or degrade the message. The first interaction occurs just after transcriptional initiation, when the complement to the promoter sequence is cleaved out of the mRNA and the capping machinery incorporates a "GpppN" cap at the 5' end of the mRNA. This results in recruitment of elongation factors that regulate the reset of mRNA transcription. Elongation is followed by 3'-end processing and splicing, resulting in a mature RNA transcript that is exported to the cytoplasm for translation. All of these processes require significant protein–RNA interactions and are highly regulated and complex. Many of the regulatory elements for this process reside in noncoding 3' and 5' untranslated regions (UTRs) of the mRNA. However, regulatory microRNAs (miRNAs) also occur in coding regions of introns, as well as exons, noncoding genes and repetitive elements.

d) Protein -carbohydrates interaction

- ➔ Carbohydrates are one of the four central building blocks of life, along with proteins, nucleic acids and lipids. They are conjugated to proteins (glycoproteins) or lipids (glycolipids) or exist as free ligands. Carbohydrates have higher diversity than proteins and nucleic acids, resulting from the types of residues (e.g., glucose and galactose), anomeric centres (α and β), glycosidic linkage positions and chemical modifications (e.g., sulfation and methylation).
- ➔ Protein–carbohydrate interactions are ubiquitous in nature, and carbohydrates can be noncovalently bound to proteins from different families including lectins, antibodies, sugar transporters and enzymes. These interactions have critical roles in many biological activities such as cell–cell adhesion and communication, tumour growth and metastasis, anticoagulation, inflammation and microbial infection.
- ➔ Hence, carbohydrates and their binding partners have emerged as important targets in fighting against major diseases.

PAPER II

UNIT 1.5

A. What are proteins and explain the role of amino acid in proteins

➔ Proteins are large, complex molecules that play many critical roles in the body. They do most of the work in cells and are required for the structure, function, and regulation of the body's tissues and organs.

➔ Proteins are made up of hundreds or thousands of smaller units called amino acids, which are attached to one another in long chains. There are 20 different types of amino acids that can be combined to make a protein.

➔ The sequence of amino acids determines each protein's unique 3-dimensional structure and its specific function. Amino acids are coded by combinations of three DNA building blocks (nucleotides), determined by the sequence of genes.

B. How Proteins are classified on basis of their R group.

➔ The amino acids present in proteins differ from each other in the structure of their side (R) chains. The simplest amino acid is glycine, in which R is a hydrogen atom. In a number of amino acids, R represents straight of branched carbon chains.

- ➔ Simple amino acids: these have no functional group in their side chain. Example: glycine, valine, alanine, leucine, isoleucine
- ➔ Hydroxy amino acids: these have a hydroxyl group in their side chain Eg: serine, threonine
- ➔ Sulphur containing amino acids: have sulphur in their side chain Eg: cysteine, methionine
- ➔ Aromatic amino acids: have benzene ring in their side chain Eg: phenylalanine, tyrosine
- ➔ Heterocyclic amino acids: having a side chain ring which possess at least on atom other than carbon Eg: Tryptophan, histidine, proline
- ➔ Amine group containing amino acids: derivatives of amino acids in which one of carboxyl group has been transformed into an amide group Eg: Asparagine, glutamine
- ➔ Branched chain amino acids: A branched-chain amino acid (BCAA) is an amino acid having aliphatic side-chains with a branch Eg: leucine, isoleucine, valine
- ➔ Acidic amino acids: have carboxyl group in their side chain Eg: Aspartic and Glutamic acid.
- ➔ Basic amino acids: contain amino group in their side chain Eg: Lysine, Arginine.
- ➔ Imino acid: Amino acids containing a secondary amine group Eg: Proline.

C. Explain different levels of protein along with an example.

➔ **Primary structure**

- ➔ The simplest level of protein structure, primary structure, is simply the sequence of amino acids in a polypeptide chain. For example, the hormone insulin has two polypeptide chains, A and B. Each chain has its own set of amino acids, assembled in a particular order.

➔ **Secondary structure**

- ➔ The next level of protein structure, secondary structure, refers to local folded structures that form within a polypeptide due to interactions between atoms of the backbone. (The backbone just refers to the polypeptide chain apart from the R groups – so all we mean here is that secondary structure does not involve R group atoms.) The most common types of secondary structures are the α helix and the β pleated sheet. Both structures are held in shape by hydrogen bonds, which form between the carbonyl O of one amino acid and the amino H of another.
- ➔ In an α helix, the carbonyl (C=O) of one amino acid is hydrogen bonded to the amino H (N-H) of an amino acid that is four down the chain. (E.g., the carbonyl of amino acid 1 would form a hydrogen bond to the N-H of amino acid 5.) This pattern of bonding pulls the polypeptide chain into a helical structure that resembles a curled ribbon, with each turn of the helix containing 3.6 amino acids. The R groups of the amino acids stick outward from the α helix, where they are free to interact.
- ➔ 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).

➔ **Tertiary structure**

- ➔ The overall three-dimensional structure of a polypeptide is called its tertiary structure. The tertiary structure is primarily due to interactions between the R groups of the amino acids that make up the protein.
- ➔ R group interactions that contribute to tertiary structure include hydrogen bonding, ionic bonding, dipole-dipole interactions, and London dispersion forces. Finally, there's one special type of covalent bond that can contribute to tertiary structure: the disulphide bond. Disulphide bonds, covalent linkages between the sulphur-containing side chains of cysteines, are much stronger than the other types of bonds that contribute to tertiary structure. They act like molecular "safety pins," keeping parts of the polypeptide firmly attached to one another.

➔ **Quaternary structure**

- ➔ Many proteins are made up of a single polypeptide chain and have only three levels of structure. However, some proteins are made up of multiple polypeptide chains, also known as subunits. When these subunits come together, they give the protein its quaternary structure. One example of a protein with quaternary structure: haemoglobin. It carries oxygen in the blood and is made up of four subunits, two each of the α and β types.
- ➔ Another example is DNA polymerase, an enzyme that synthesizes new strands of DNA and is composed of ten subunits. In general, the same types of interactions that contribute to tertiary structure (mostly weak interactions, such as hydrogen bonding and London dispersion forces) also hold the subunits together to give quaternary structure.

D. Write a note on Hsp60 chaperon in detail along with along diagram and mechanism.

➔ Hsp60 is a molecular chaperone which assists protein folding in prokaryotes and in eukaryotic cell organelles. In mitochondria, Hsp60 is considered an essential chaperone for the protein import into the mitochondrial matrix. Hsp60 functions together with its co-chaperone Hsp10 also residing in mitochondria.

➔ Most information on Hsp60 and Hsp10 derives from studies on their prokaryotic homologs, GroEL and GroES, respectively. As already mentioned, GroEL and GroES assemble in a barrel-shaped oligomer. In vitro studies by Ryabova and colleagues revealed a self-organization of the GroEL/GroES complex. According to this work, unfolded monomeric GroEL and GroES are folded very rapidly into the oligomerization-competent form. The oligomerization process depends on the concentration of the monomers and, in the case of GroEL, requires the presence of Mg^{2+}, adenine nucleotides and GroES. For GroES oligomerization, no specific factor is needed to induce subunit assembly. GroEL forms an unstable intermediate ring heptamer which is stabilized in the presence of heptameric GroES. Heptameric GroEL and heptameric GroES then constitute a stable two- ring GroEL structure in the absence of Mg^{2+}/ATP(ADP).

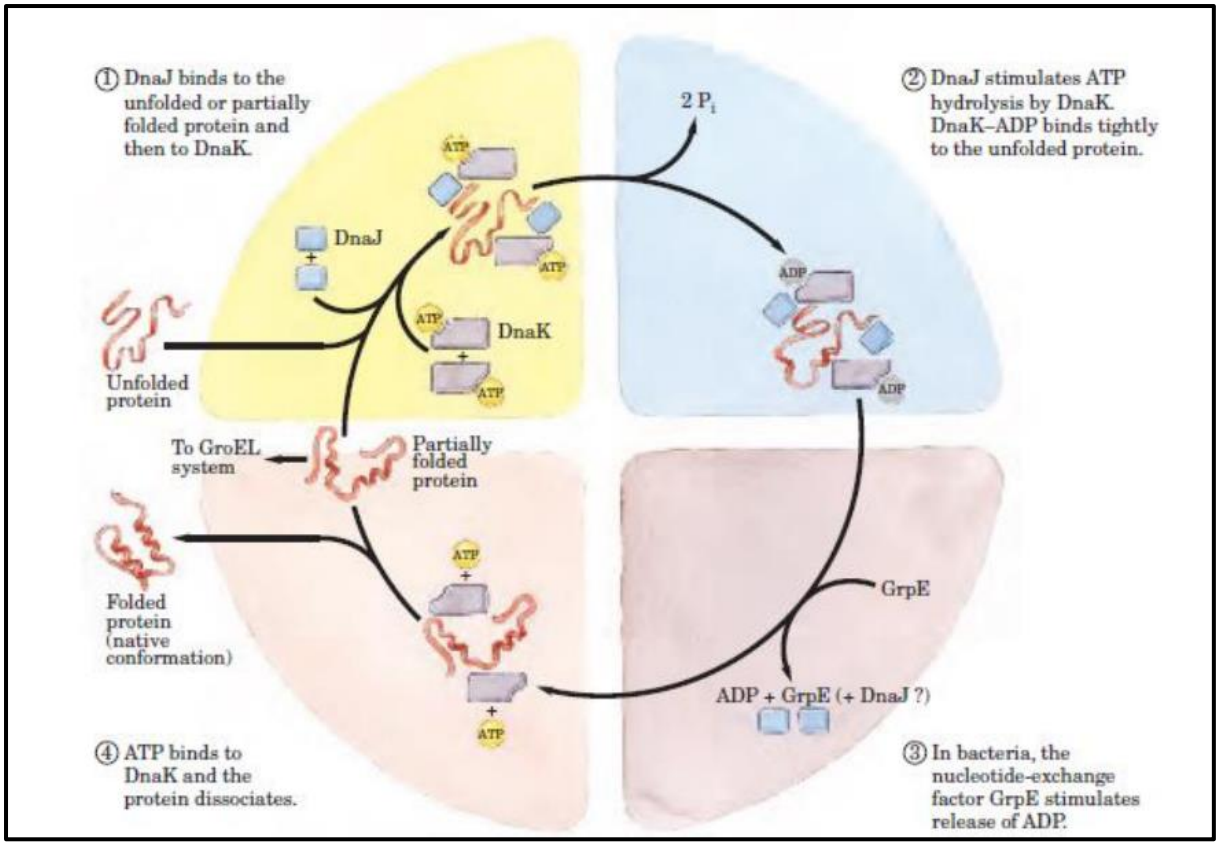
➔ As already mentioned, GroEL and GroES rapidly constitute a GroEL/ADP/GroES complex in the presence of Mg^{2+}/ATP. Since the GroEL chaperonin shows high affinity only to one heptameric GroES, previous studies primarily demonstrated the appearance of one GroEL particle in the complex with one GroES heptamer. GroEL can also form a complex with two GroES heptamers depending on the K^+ concentration and the ATP/ADP ratio. As illustrated in this Figure, a polypeptide destined for folding translocates into the folding chamber of one ring followed by its capping with heptameric GroES. In this GroES-bound ring (cis ring), folding of the unfolded polypeptide occurs in the presence of ATP; the second unliganded ring is referred to as the trans ring. A further folding process is started in the second (trans) ring after release of ADP and the folded protein as well as the dissociation of GroES.

E. Elaborate on the enzymes that catalyses the protein folding.

- ➔ Cells contain at least two types of enzymes that catalyse protein folding by breaking and re-forming covalent bonds. The formation of disulphide bonds between cysteine residues is important in stabilizing the folded structures of many proteins. Protein disulphide isomerase, which was discovered by Christian Anfinsen in 1963, catalyses the breakage and re-formation of these bonds. For proteins that contain multiple cysteine residues, protein disulphide isomerase (PDI) plays an important role by promoting rapid exchanges between paired disulphides, thereby allowing the protein to attain the pattern of disulphide bonds that is compatible with its stably folded conformation. Disulphide bonds are generally restricted to secreted proteins and some membrane proteins because the cytosol contains reducing agents that maintain cysteine residues in their reduced (—SH form), thereby preventing the formation of disulphide (S—S) linkages. In eukaryotic cells, disulphide bonds form in the endoplasmic reticulum, in which an oxidizing environment is maintained. Consistent with the role of disulphide bonds in stabilizing secreted proteins, the activity of PDI in the endoplasmic reticulum is correlated with the level of protein secretion in different types of cells.
- ➔ The second enzyme that plays a role in protein folding catalyses the isomerization of peptide bonds that involve proline residues. Proline is an unusual amino acid in that the equilibrium between the cis and trans conformations of peptide bonds that precede proline residues is only slightly in favour of the trans form. In contrast, peptide bonds between other amino acids are almost always in the trans form. Isomerization between the cis and trans configurations of prolyl peptide bonds, which could otherwise represent a rate-limiting step in protein folding, is catalysed by the enzyme peptidyl prolyl isomerase. This enzyme is widely distributed in both prokaryotic and eukaryotic cells and can catalyse the refolding of at least some proteins. However, its physiologically important substrates and role within cells have not yet been determined.

F. Write a note on Hsp70 chaperon in detail along with along diagram and mechanism

- ➔ Folding for many proteins is facilitated by the action of specialized proteins. Molecular chaperones are proteins that interact with partially folded or improperly folded polypeptides, facilitating correct folding pathways or providing microenvironments in which folding can occur. Two classes of molecular chaperones have been well studied. Both are found in organisms ranging from bacteria to humans. The first class, a family of proteins called Hsp70, generally have a molecular weight near 70,000 and are more abundant in cells stressed by elevated temperatures (hence, heat shock proteins of Mr 70,000, or Hsp70). Hsp70 proteins bind to regions of unfolded polypeptides that are rich in hydrophobic residues, preventing inappropriate aggregation. These chaperones thus “protect” proteins that have been denatured by heat and peptides that are being synthesized (and are not yet folded). Hsp70 proteins also block the folding of certain proteins that must remain unfolded until they have been translocated across membranes. Some chaperones also facilitate the quaternary assembly of oligomeric proteins. The Hsp70 proteins bind to and release polypeptides in a cycle that also involves several other proteins (including a class called Hsp40) and ATP hydrolysis. Figure 4–30 illustrates chaperone assisted folding as elucidated for the chaperones DnaK and DnaJ in E. coli, homologs of the eukaryotic Hsp70 and Hsp40. DnaK and DnaJ were first identified as proteins required for in vitro replication of certain viral DNA molecules (hence the “DNA” designation).
- ➔ The cyclic pathway by which chaperones bind and release polypeptides is illustrated for the E. coli chaperone proteins DnaK and DnaJ, homologs of the eukaryotic chaperones Hsp70 and Hsp40. The chaperones do not actively promote the folding of the substrate protein, but instead prevent aggregation of unfolded peptides. For a population of polypeptides, some fraction of the polypeptides released at the end of the cycle are in the native conformation. The remainder are rebound by DnaK or are diverted to the chaperonin system. In bacteria, a protein called GrpE interacts transiently with DnaK late in the cycle (step 3), promoting dissociation of ADP and possibly DnaJ. No eukaryotic analogue of GrpE is known.



G. Write a note on:

a. Anfinsen experiment

➔ The Nobel-winning scientist Christian Anfinsen proposed several decades ago that the three-dimensional structure of a protein is a direct consequence of its primary structure. That is, proteins find the folded conformation that is lowest in energy, meaning the conformation in which favourable interactions are maximized and unfavourable interactions are minimized. To test this hypothesis, Anfinsen denatured (unfolded) a protein and then asked if it would fold back into its native conformation in a test tube. He used the enzyme ribonuclease A, which cleaves RNA. This protein is relatively small, and because it is an enzyme, its folded state could be assessed by measuring its enzymatic activity.

➔ Anfinsen took two samples of ribonuclease A and added both a reducing agent and 8 M urea (CO(NH2)2) to each of them. The reducing agent was added in order to break the four disulphide bonds that stabilize the structure of ribonuclease A. Urea causes proteins to unfold. At high concentrations, urea alters the extensive hydrogen bonding network between water molecules (urea contains both hydrogen bond donors and acceptors) and thus interferes with the forces that promote protein folding (in effect, it makes the hydrophobic effect less important in protein folding). Anfinsen then took one of the samples and used dialysis to remove the urea. Next, Anfinsen treated this sample with an oxidizing agent to promote disulphide bond formation. The protein was then tested for enzymatic activity, and it was found to have 90% of the activity of the original sample, implying that most of the protein molecules had refolded into their correct, or native, shape. In a parallel experiment, Anfinsen took the other sample and, after it had been treated with the reducing agent and urea, oxidized it first to promote disulphide bond formation and then removed the urea. In other words, he reversed the order of the final two steps. This sample was only 1-2% as active as the original, implying that the procedure of oxidizing while the protein was still denatured was ineffective in restoring enzymatic activity. By oxidizing the protein before removing the urea, Anfinsen "trapped" conformations containing incorrect disulphide bonds. These bonds then prevented the protein from re-adopting its native structure.

➔ Anfinsen's experiment showed that the native structure of ribonuclease A will form following denaturation provided that premature oxidation is prevented. Therefore, the protein is

intrinsically capable of finding its lowest-energy conformation. Anfinsen concluded that the information required to fold a protein into its native, lowest-energy conformation is entirely contained within its sequence of amino acids.

b. Allosteric proteins

➔ A protein whose shape is changed when it binds a particular molecule. In the new shape the protein's ability to react to a second molecule is altered. Allosteric proteins, such as haemoglobin, are assemblies of functional units, which undergo quaternary structural transitions in response to concentration changes of a specific ligand. Functional properties of haemoglobin ligation intermediates indicate that the tertiary structural changes induced by the ligand do not promote an equilibrium of quaternary structures. The term allostery means "other sites." Allosteric proteins, such as haemoglobin, are "intelligent" molecules that vary their activity in response to environmental stimuli in the form of concentration changes of ligands, such as ions, metabolites, and macromolecules.

c. Misfolding of protein results in various diseases (any 2 in detail)

➔ In general, proteins can only execute their various biological functions when they are appropriately folded. Their amino acid sequence encodes the relevant information required for correct three-dimensional folding, with or without the assistance of chaperones. The challenge associated with understanding protein folding is currently one of the most important aspects of the biological sciences.

➔ Misfolded protein intermediates form large polymers of unwanted aggregates and are involved in the pathogenesis of many human diseases, including Alzheimer's disease (AD) and Type 2 diabetes mellitus (T2DM). AD is one of the most prevalent neurological disorders and has worldwide impact; whereas T2DM is considered a metabolic disease that detrimentally influences numerous organs, afflicts some 8% of the adult population, and shares many risk factors with AD.

➔ Research data indicates that there is a widespread conformational change in the proteins involved in AD and T2DM that form β-sheet like motifs. Although conformation of these β-sheets is common to many functional proteins, the transition from α-helix to β-sheet is a typical characteristic of amyloid deposits. Any abnormality in this transition results in protein aggregation and generation of insoluble fibrils.

➔ The abnormal and toxic proteins can interact with other native proteins and consequently catalyse their transition into the toxic state. Both AD and T2DM are prevalent in the aged population. AD is characterized by the accumulation of amyloid-β (Aβ) in brain, while T2DM is characterized by the deposition of islet amyloid polypeptide (IAPP, also known as amylin) within beta-cells of the pancreas. T2DM increases pathological angiogenesis and immature vascularisation.

➔ This also leads to chronic cerebral hypoperfusion, which results in dysfunction and degeneration of neuroglial cells. With an abundance of common mechanisms underpinning both disorders, a significant question that can be posed is whether T2DM leads to AD in aged individuals and the associations between other protein misfolding diseases.

➔ Parkinson's disease is thought to be triggered by the misfolding of a protein known as alpha-synuclein. This is supported by the observation that mutations in the SNCA gene, which encodes alpha-synuclein, cause familial disease. Misfolded alpha-synuclein forms aggregates, or clumps, and accumulates in nerve cells (neurons) in the brain that generate dopamine, a signalling molecule that plays a role in motor function. Eventually, aggregated alpha-synuclein spreads to other brain cells in regions involved in cognition, sleep, and mood.

PAPER III

UNIT 1.6

Q. Write the principle and working of the following:

1. **X-Ray crystallography**
   - ➔ In a single-crystal X-ray diffraction measurement, a crystal is mounted on a goniometer, which is used to position the crystal at selected orientations. The crystal is illuminated with a finely focused monochromatic beam of X-rays, leading to a diffraction pattern of regularly spaced spots known as reflections. X-ray crystallography works in a manner of elastic scattering with the outgoing X-rays having the same energy and wavelength as the incoming X-rays, which get an altered direction after diffraction. A crystallographer can then produce a three-dimensional picture of the density of electrons within the crystal by measuring the angles and intensities of these diffracted beams under the assistance of the mathematical method Fourier transforms. From this electron density, the mean positions of the atoms, chemical bonds, crystallographic disorder, and some other information in the crystal can be determined. Poor resolution or even errors may occur if the crystals are too small, or not uniform enough in their internal makeup.
   - ➔ The technique of single crystal X-ray crystallography has three basic steps. The first and usually most difficult step is to produce an adequate crystal of the studied material. The crystal should be sufficiently large with all dimensions larger than 0.1 mm, pure in composition and regular in structure, and have no significant internal imperfections such as cracks or twinning. The crystal is subsequently placed in an intense beam of X-rays, usually of a single wavelength, to produce regular reflection pattern. The angles and intensities of diffracted X-rays are measured with each compound having a unique diffraction pattern. Previous reflections disappear and new ones appear along with the gradual rotation of the crystal, and the intensity of every spot is recorded at every orientation of the crystal. Multiple data sets may have to be collected since each set covers slightly more than half a full rotation of the crystal and typically contains tens of thousands of reflections. Ultimately, these collected data are combined computationally with complementary chemical information to obtain and refine a model from the arrangement of atoms within the crystal. The final refined model of the atomic arrangement is called a crystal structure and usually stored in a public database.
2. **NMR**
   - ➔ Nuclear magnetic resonance (NMR) spectroscopy is the study of molecules by recording the interaction of radiofrequency (Rf) electromagnetic radiations with the nuclei of molecules placed in a strong magnetic field.
   - ➔ NMR Spectroscopy Principle:
     - Many nuclei have spin, and all nuclei are electrically charged, according to the NMR principle. An energy transfer from the base energy to a higher energy level is achievable when an external magnetic field is supplied.
     - All nuclei are electrically charged and many have spin.
     - Transfer of energy is possible from base energy to higher energy levels when an external magnetic field is applied.
     - The transfer of energy occurs at a wavelength that coincides with the radio frequency.
     - Also, energy is emitted at the same frequency when the spin comes back to its base level.
     - Therefore, by measuring the signal which matches this transfer the processing of the NMR spectrum for the concerned nucleus is yield.
   - ➔ NMR Spectroscopy Working:
     - Place the sample in a magnetic field.
     - Excite the nuclei sample into nuclear magnetic resonance with the help of radio waves to produce NMR signals.

- These NMR signals are detected with sensitive radio receivers.
- The resonance frequency of an atom in a molecule is changed by the intramolecular magnetic field surrounding it.
- This gives details of a molecule's individual functional groups and its electronic structure.
- Nuclear magnetic resonance spectroscopy is a conclusive method of identifying monomolecular organic compounds.
- This method provides details of the reaction state, structure, chemical environment and dynamics of a molecule.

3. Cryo-electron microscopy

➔ Cryo-electron microscopy (Cryo-EM) is a type of transmission electron microscopy that allows for the specimen of interest to be viewed at cryogenic temperatures. Following years of improvement, the cryo-electron microscope has become a valuable tool for viewing and studying the structures of various biological molecules.

➔ Transmission Electron Microscopy (TEM) refers to a technique where the image (of specimen) is formed by directing a high energy electron beam at a thin sample. While cryo-electron microscopy encompasses a number of experimental methods (imaging intact tissue sections, imaging plunge frozen cells and virus etc), these methods are based on the principle of imaging radiation-sensitive specimens in a transmission electron microscope.

➔ Transmission electron microscopes use the same working principle as the ordinary light microscope. However, rather than using the limited wavelength of light, electrons with much lower wavelengths are used as the source of light. For the TEM, there are two types of electron sources that are commonly used. These include the thermionic electron guns and the field emission guns. Whereas electrons are emitted from such heated filaments as bent tungsten or sharp lanthanum hexaboride crystal in thermionic electron guns, they are emitted from a sharp, pointed cathode by a string electric field in field emission guns. For TEM, the electron gun uses electronic coils as well as high voltages to accelerate the electrons to extremely high speeds (thus shorter waves). The electrons then travel through the anode, an aperture and into the vacuum tube. Unlike the light microscope, the TEM has in place electromagnetic lenses that bend the electron beam though the Lorentz force. The lenses also direct the beam through the tube and onto the specimen.

➔ Basically, the transmission electron microscope can be said to have three important systems. These include:

- **The Electron Gun -** The electron gun is the part of the microscope that is responsible for producing the electron beam. Here, the condenser system is responsible for focusing the electron beam on to the specimen sample.
- **Image Producing System -** This is composed of the objective lens, intermediate and projector lenses as well as a movable stage. The lenses are also involved in the focusing of electrons, which helps form the magnified image.
- **Image Recording System** - This part of the microscope is mostly composed of a fluorescent screen that helps in producing an image that can be seen with the eye. Here, a digital camera also helps capture the images for documentation.
- **In Cryo-Electron Microscopy** the sample under observation is usually frozen (frozen-hydrated) for preservation purposes. Here, a very thin slide of the specimen may be rapidly plunged into a liquid ethane bath and viewed in their natural state. Solvents like water or a salt solution is used to ensure that the sample remains stable.

References:

- ➔ Gutmanas, A., Oldfield, T. J., Patwardhan, A., Sen, S., Velankar, S., & Kleywegt, G. J. (2013). The role of structural bioinformatics resources in the era of integrative structural biology. Acta Crystallographica Section D Biological Crystallography, 69(5), 710–721. https://doi.org/10.1107/s0907444913001157
- ➔ Orengo, C., Velankar, S., Wodak, S., Zoete, V., Bonvin, A. M., Elofsson, A., Feenstra, K. A., Gerloff, D. L., Hamelryck, T., Hancock, J. M., Helmer-Citterich, M., Hospital, A., Orozco, M., Perrakis, A., Rarey, M., Soares, C., Sussman, J. L., Thornton, J. M., Tuffery, P., . . . Schneider, B. (2020). A community proposal to integrate structural bioinformatics activities in ELIXIR (3D-Bioinfo Community). F1000Research, 9, 278. https://doi.org/10.12688/f1000research.20559.1
- ➔ Samish, I., Bourne, P. E., & Najmanovich, R. J. (2014). Achievements and challenges in structural bioinformatics and computational biophysics. Bioinformatics, 31(1), 146–150. https://doi.org/10.1093/bioinformatics/btu769
- ➔ Waman, V. P., Sen, N., Varadi, M., Daina, A., Wodak, S. J., Zoete, V., Velankar, S., & Orengo, C. (2020). The impact of structural bioinformatics tools and resources on SARS-CoV-2 research and therapeutic strategies. Briefings in Bioinformatics, 22(2), 742–768. https://doi.org/10.1093/bib/bbaa362
- ➔ Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., Ballestar, E., Bongcam-Rudloff, E., Conesa, A., & Tegnér, J. (2014). Data integration in the era of omics: current and future challenges. BMC Systems Biology, 8(Suppl 2), I1. https://doi.org/10.1186/1752-0509-8-s2-i1
- ➔ Sharma, A., Kumar, G., Sharma, S., Walia, K., Chouhan, P., Mandal, B., & Tuli, A. (2021). Methods for binding analysis of small GTP-binding proteins with their effectors. Biomolecular Interactions Part A, 235–250. https://doi.org/10.1016/bs.mcb.2021.06.003
- ➔ Haurowitz, F. and Koshland, . Daniel E. (2020, December 1). protein. Encyclopedia Britannica. https://www.britannica.com/science/protein
- ➔ Glassco, J. (2019, November 27). Protein Structure: Primary, Secondary, Tertiary, Quatemary Structures. LLS Health CDMO. https://lubrizolcdmo.com/technical-briefs/protein-structure/#:%7E:text=Proteins%20are%20macromolecules%20and%20have,%2C%20secondary%2C%20tertiary%20and%20quaternary.
- ➔ Bukau, B., & Horwich, A. L. (1998). The Hsp70 and Hsp60 Chaperone Machines. Cell, 92(3), 351–366. https://doi.org/10.1016/s0092-8674(00)80928-9
- ➔ Nagradova, N. (2007). Enzymes Catalyzing Protein Folding and Their Cellular Functions. Current Protein & Peptide Science, 8(3), 273–282. https://doi.org/10.2174/138920307780831866
- ➔ LabXchange. (2020, January 20). Labxchange. https://www.labxchange.org/library/pathway/lx-pathway:c02e813d-1e2a-4091-a70e-42d183abdec6/items/lx-pb:c02e813d-1e2a-4091-a70e-42d183abdec6:html:f00de44e
- ➔ Perrella, M., & Russo, R. (2003). Allosteric Proteins: Lessons to be Learned From the Hemoglobin Intermediates. Physiology, 18(6), 232–236. https://doi.org/10.1152/nips.01451.2003
- ➔ Chaudhuri, T. K., & Paul, S. (2006). Protein-misfolding diseases and chaperone-based therapeutic approaches. FEBS Journal, 273(7), 1331–1349. https://doi.org/10.1111/j.1742-4658.2006.05181.x
- ➔ Smyth, M. S. (2000). x Ray crystallography. Molecular Pathology, 53(1), 8–14. https://doi.org/10.1136/mp.53.1.8
- ➔ Aryal, S. (2021, February 4). Nuclear Magnetic Resonance (NMR) Spectroscopy. Microbe Notes. https://microbenotes.com/nuclear-magnetic-resonance-nmr-spectroscopy/
- ➔ Milne, J. L. S., Borgnia, M. J., Bartesaghi, A., Tran, E. E. H., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A., & Subramaniam, S. (2012). Cryo-electron microscopy - a primer for the non-microscopist. FEBS Journal, 280(1), 28–45. https://doi.org/10.1111/febs.12078