Chapter 7

STRUCTURAL AND COMPUTATIONAL BIOLOGY

STUDENTS'LEARNING OUTCOMES

After studying this chapter, the students will be able to:

- Define structural biology.
- Explain that structure determination of biomolecules are important.
- Describe how X-ray crystallography works.
- Outline the online databases where biomolecule structures are available.
- Describe computational biology.
- Define sequence homology.
- Define structural homology.

Structural biology deals with the study of three dimensional (3D) structures of macromolecules (including proteins and nucleic acids) at atomic levels. It provides the detailed information about the structure of biomolecule, its functions, dynamics and interaction with ligands and other macromolecules.

7.1- APPLICATIONS OF STRUCTURAL BIOLOGY

Structural biology has a wide range of applications especially in the field of medical research. Some of these are discussed here:

1- Determining the Active sites and Domains

Structural biologists can determine the three-dimensional (3D) structures of macromolecules such as proteins and nucleic acids. The 3D structures reveal the exact location, shape, and environment of the active sites and different domains (distinct structural units with independent functions) of macromolecules. For example, structural studies of the enzyme HIV-1 reverse transcriptase have identified its polymerase domain (which synthesizes DNA) and RNase H domain (which breaks down the RNA strand of RNA-DNA hybrids). Knowing the location and structure of these domains has helped in the design of antiviral drugs that specifically target them. Similarly, the structure of serine proteases reveals its well-defined active site, which is responsible for breaking down peptide bonds.

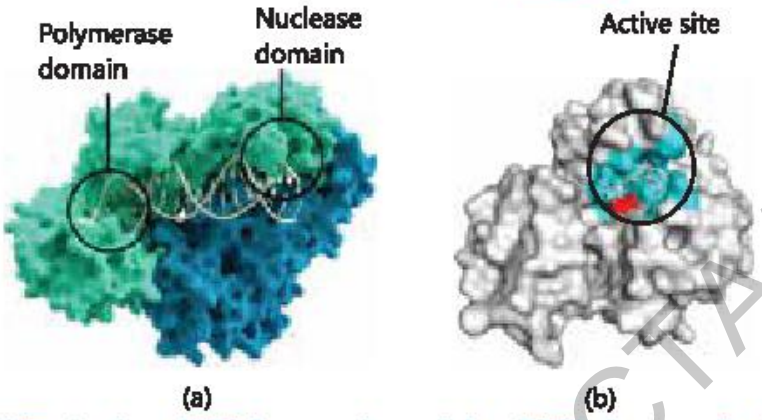

Figure 7.1: (a) 3D structure of HIV-1 reverse transcriptase (b) 3D structure of serine protease

2- Identifying Drug Targets

Structural biology helps scientists find the right place on a disease-causing molecule where a drug can work. These places are usually proteins and are called drug targets. By studying the 3D shape of these proteins, scientists can find specific spots where a drug can attach and stop the protein from working. For example, in COVID-19, scientists used structural biology to study the spike protein of the coronavirus (SARS-CoV-2). This protein helps the virus to enter human cells. By knowing its 3D structure, scientists identified it as a drug target. Thus, they design vaccines and medicines that block the spike protein, preventing the virus from infecting more cells.

3- Identifying Host–Pathogen Interactions

Structural biology also helps in understanding how pathogens (like viruses or bacteria) interact with the host's body cells. This is called host–pathogen interaction. By studying the 3D structures of both the pathogen and the host cell proteins, scientists can see how the pathogen attaches to and enters the host cell, and which molecules are involved in the process. For example, structural biologists studied the spike protein of coronavirus, which sticks out from the surface of the virus. They also look at a protein on human cells that acts as receptor of virus spike protein. So, the scientists discover exactly how the virus enters human cells. This information is vital in developing the drugs that can bind with receptor proteins. Such drug inhibits the interaction of the virus with the receptor and consequently blocks the entry of virus into the host cells.

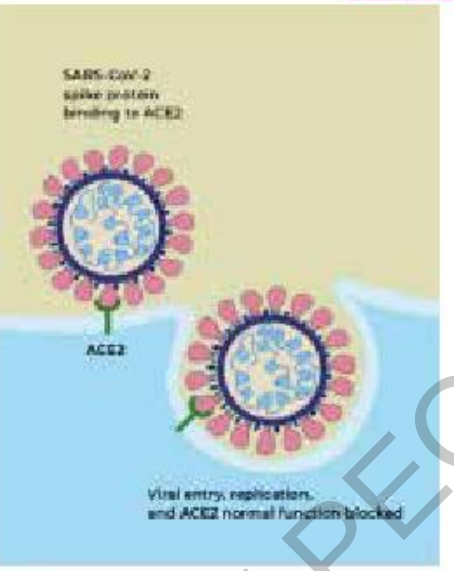

Figure 7.2: Mechanism of corona virus binding with receptorof human cell.

4- Identifying Protein Misfolding

The functionality of proteins depends on the correct folding into three dimensional shapes. Several diseases (including cyctic fibrosis, Parkinsons, Alziemer's) originate due to incorrect folding of proteins. Structural biology provides understanding of intricate folding pathways and how misfolding leads to the diseases.

7.2- X-Ray Crystallography

X-ray crystallography was developed in 1912 by William Henry Bragg and William Lawrence Bragg. They were awarded 1915 Nobel Prize in Physics for their work. Since then it has been used to analyze the diverse substances including minerals, salts, metals, proteins, carbohydrates, nucleic acids and vitamins. In this technique, x-rays beam strikes a crystals and atoms and molecules in the crystals diffract the x-rays beam in specific directions. From the angles and intensities of diffracted beams, a 3D picture of electron density within the crystals are produced. The electron density is exploited to create 3D structure of the molecule.

In order to understand the working of X-ray crystallography, let us take the example of protein structure determination. The method can be divided into following steps:

(I) **Protein crystallization:** Protein crystallization means turning a purified protein into a solid crystal form. Crystals are needed because they arrange protein molecules in a regular, repeating pattern, which is important for getting a clear image during the X-ray process. To make crystals, scientists slowly mix the protein with special solutions that cause the protein molecules to stick together in an orderly way. This process can take hours, days, or even weeks. It often

requires careful control of temperature, pH, and salt concentration. Once a clear and stable protein crystal is formed, it can be used in the next steps.

(II) **Production of a diffraction pattern:** Once a good quality crystal is formed, it is mounted on the x-ray machine. The x-rays beam is bombarded at the crystal at various angles. The atoms in the crystal diffract the x-rays beam and a diffraction pattern (which is a series of spots) is created on the detectors.

(III) **Creating density map:** The angles and intensities of these spots contain information about the arrangement of atoms in the crystal. Diffraction pattern is used to make a density map.

(iv) **Determination of protein structure:** Then the data is analyzed mathematically by using computational programs. These calculations transform data into the 3D structure of protein.

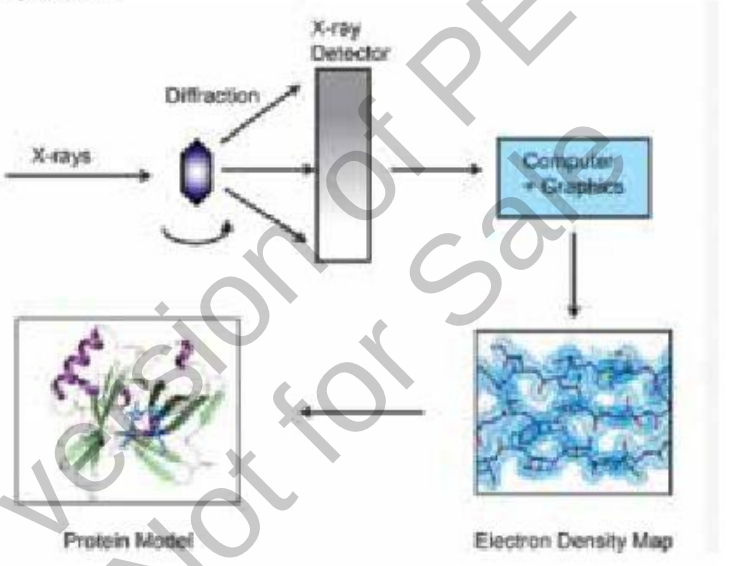

Figure 7.3: Schematic representation of X-ray crystallography

7.3- COMPUTATIONAL BIOLOGY

Computational biology is an interdisciplinary field that uses computational techniques and tools to solve biological problems. It integrates knowledge from biology, computer science, mathematics, and statistics to analyze and interpret biological data. The importance of computational biology lies in its ability to handle large datasets, uncover hidden patterns, and generate predictive models that can lead to new biological insights and applications. Major areas in the computational biology include:

(I) Genomics is the study of genomes, which are the complete set of DNA within a single cell of an organism. Genomics involves sequencing, assembling, and

analyzing the function and structure of genomes. It helps in understanding genetic variations, gene function, and evolutionary relationships.

(ii) **Proteomics** is the large-scale study of proteins, including their structures and functions. Proteins are essential molecules that perform many functions within organisms. Proteomics aims to map the entire set of proteins (the proteome) produced by an organism and understand their interactions and roles in cellular processes.

(iii) **Bioinformatics** is the application of computer technology to manage and analyze biological data. Bioinformatics tools and techniques are used to store, retrieve, and analyze DNA, RNA, and protein sequences.

Applications of Computational Biology

Though computational biology has vast application, some of these are discussed here.

(i) **Drug Discovery:** Computational biology helps in identifying potential drug targets and simulating the effects of drugs on biological systems. It accelerates the drug discovery process by predicting how drugs interact with proteins and other molecules.

(ii) **Genetic Research:** By analyzing DNA sequences, computational biology helps identify genetic variations associated with diseases. It aids in understanding the genetic basis of diseases and can lead to the development of personalized medicine.

(iii) **Evolutionary Biology:** Computational tools are used to compare genetic information across different species, helping to reconstruct evolutionary relationships and understand the process of evolution.

Key Databases: Here are some of the key databases used to analyze nucleic acid and proteins. **GenBank:** https://www.ncbi.nlm.nih.gov/nuccore/

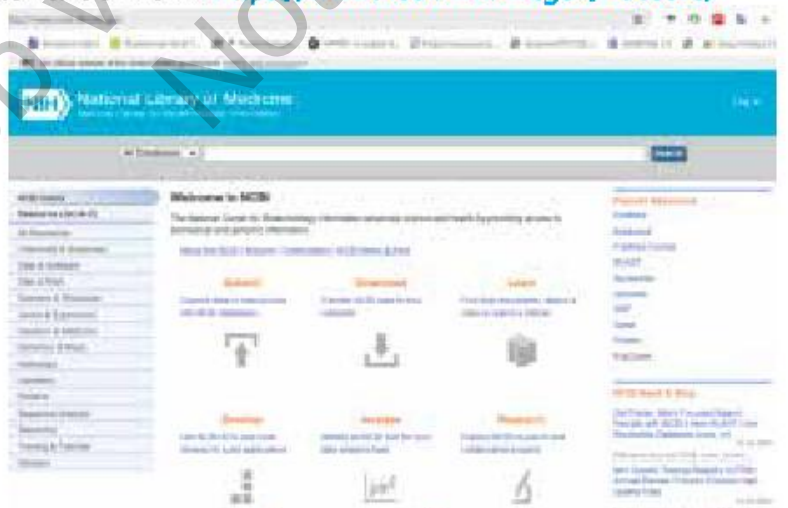

Figure 7.4: Screenshot of GenBank database

It is a comprehensive public database of nucleotide sequences and supporting bibliographic and biological annotations. It provides access to a vast repository of DNA sequences from various organisms, facilitating genetic research and comparative genomics.

Protein Data Bank (PDB)

This database provides 3D structural data of large biological molecules, such as proteins and nucleic acids. It is important for studying the structures of macromolecules, understanding their functions, and designing drugs that target specific protein structures.

Figure 7.5: Screenshot of protein databank

Ensembl

It is a genome browser providing information on genome sequences, gene models, and comparative genomics for various species. Ensembl helps to access and visualize genomic data, supporting studies in genomics and evolutionary biology.

Key Algorithms

In addition to above mentioned databases, some algorithms being used in data analysis are discussed below.

BLAST (Basic Local Alignment Search Tool): It is used for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA sequences. It helps identify homologous sequences, predict functions of unknown genes, and study evolutionary relationships.

FASTA: It is a sequence alignment tool that compares a query sequence to a database of sequences to find regions of similarity. It is used for searching protein

and nucleotide databases, identifying sequence homology, and analyzing sequence alignments.

7.4- SEQUENCE HOMOLOGY

Sequence homology refers to the similarity between DNA, RNA, or protein sequences due to shared ancestry. Homologous sequences have evolved from a common ancestral sequence and can be categorized into two main types: (i) **Orthologs:** Sequences in different species that originated from a common ancestral gene during speciation. Orthologs often retain the same function across species. (ii) **Paralogs:** Sequences within the same species that originated from gene duplication. Paralogs can evolve new functions even if they originally arise from the same ancestral gene.

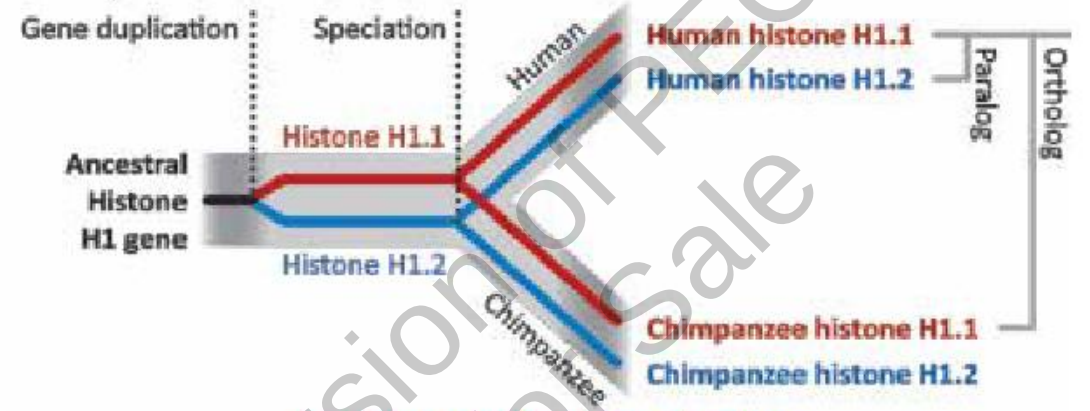

Figure 7.6. Types of Homologous Sequences

Sequence homology provides an insight into the evolutionary relationships between organisms. By comparing homologous sequences, scientists can infer the evolutionary history and divergence of species. Furthermore sequence homology provides a clue about the function of an unknown gene or protein. If an unknown gene/protein is homologous to a gene/protein with a known function, it is likely to have a similar function. Additionally, Identifying homologous genes involved in diseases across different species helps in understanding disease mechanisms and developing treatments. Homologous genes in model organisms can be studied to gain insights into human diseases.

Structural Homology

Structural homology refers to the similarity in the three-dimensional structures of proteins or other macromolecules due to shared ancestry. Proteins with similar structures often perform similar functions, even if their sequences are not highly similar. The three-dimensional structure of a protein provides critical information about its function. Understanding structural homology helps in predicting the function of newly discovered proteins. Furthermore, structural homology is crucial in

drug design, as drugs are often designed to interact with specific protein structures. Understanding the structural relationships between proteins can help in designing more effective drugs. Also studying the structural homology of proteins helps in understanding the evolutionary processes that shape protein functions and interactions.

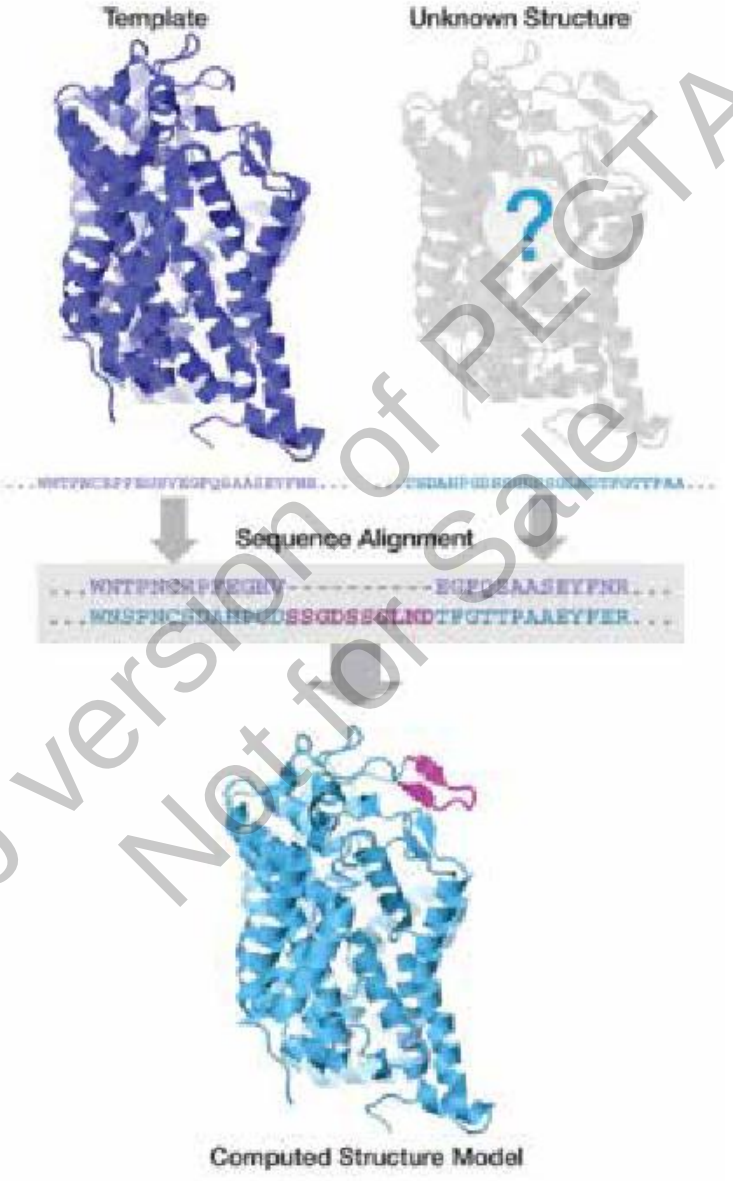

Figure 7.7: Structural homology of protein

EXERCISE

SECTION 1: MULTIPLE CHOICE QUESTIONS

1. Generally, the function of a protein depends on its:
   (a) One-dimensional structure (b) Two-dimensional structure
   (c) Three-dimensional structure (d) Four-dimensional structure
2. The protein domains are:
   (a) Functional and structural units within protein
   (b) Secondary structural elements
   (c) Linear sequences of amino acids
   (d) Specific regions for post-translational modification
3. The first step in x-ray crystallography experiment is:
   (a) Compute an electron density (b) Build a model of your molecule
   (c) Measure a diffraction pattern (d) Grow a crystal
4. What is primary role of computational biology?
   (a) Using computer algorithms to analyze data
   (b) Identifying genetic mutations
   (c) Studying protein functions
   (d) Analyzing the expression patterns
5. Which computational approach is used to predict protein structure based on amino acid sequence?
   (a) Multiple sequence alignment (b) Homology modelling
   (c) Clustering analysis (d) BLAST searches

SECTION 2:SHORT QUESTIONS

1. Define domains of the protein.
2. How corona virus enters the host cells?
3. Define genomics.
4. Differentiate between genomics and proteomics.
5. What is GenBank. Describe it briefly.
6. Write a short note on protein data bank.

SECTION 3: LONG QUESTIONS

1. Describe the applications of structural biology.
2. Write a note on principle and working of x-ray crystallography.
3. Briefly describe key databases of computational biology.

INQUISITIVE QUESTIONS

1. Consider there is a pandemic of a new unknown disease, and the causative agent is a virus. You also know that virus belongs to X family. How structural biology can be helpful in preventing the disease?
2. Suppose you find an unknown protein and determine amino acid sequence by Edman degradation/mass spectrometry. How you can exploit the computational biology to predict the structure and function of the protein.
3. Homology models of macromolecules differ from experimentally determined structures of the macromolecules. Please comment.
4. Draw a flow chart to describe the steps involved in drug development till its prescription.