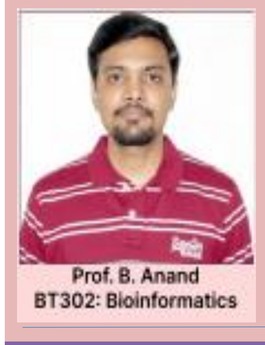


Inferring protein function from sequence and structure

Amgoth Naresh , Priyanshu Prasad ,Aarij

Department of Bioscience and Bioengineering, IIT Guwahati, Assam

Introduction

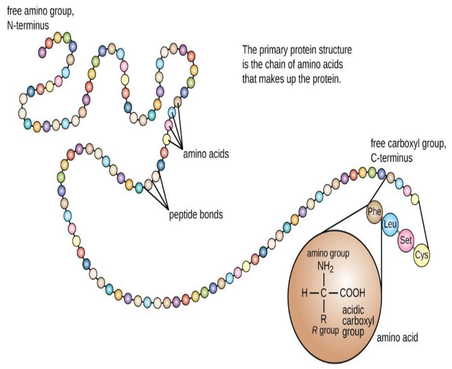
1. Proteins are large, complex molecules that play many critical roles in the body. They do most of the work in cells and are required for the structure, function, and regulation of the body's tissues and organs.

2. The sequence of amino acids in a protein and its three-dimensional structure are intricately linked to its function.

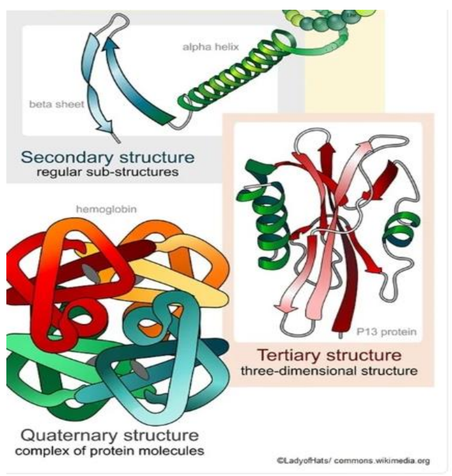
3. We examine the computational tools and algorithms that analyze amino acid sequences and predict structural motifs indicative of specific functions

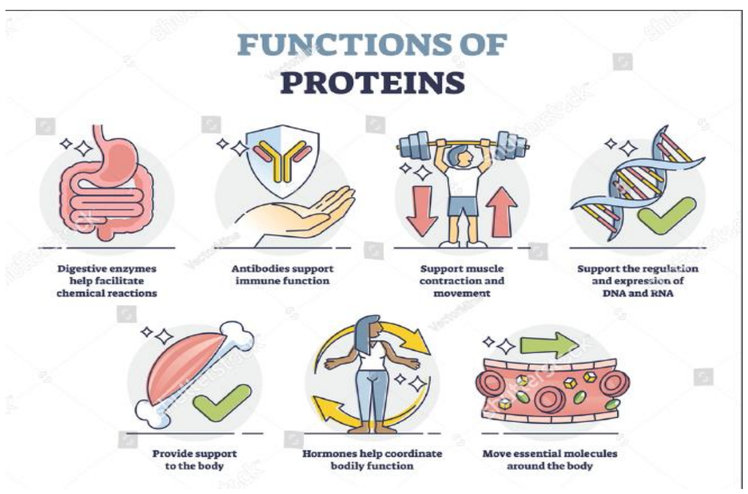
Principle

Protein Sequences:
The linear sequence of amino acids in a protein is akin to a biological code, specifying its composition and order. Variations in sequence confer distinct functionalities, influencing how a protein interacts with other molecules.

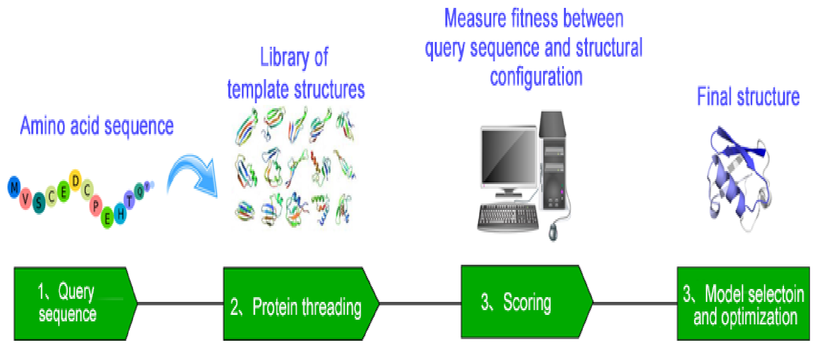


Protein Structures:
The three-dimensional arrangement of atoms in a protein determines its shape and, consequently, its function.
Structural features, such as active sites and binding pockets, provide insights into the potential roles a protein may play.





Methods

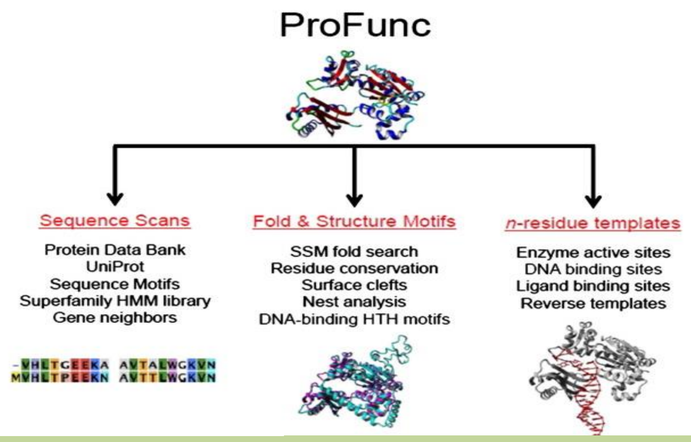


1. Sequence Analysis: Basic Sequence Information: Start by obtaining the protein's primary sequence (amino acid sequence).

2.Structure Analysis: 3D Structure Visualization: The protein structure can be seen experimentally or through structure prediction

3.Sequence Alignment and Structure Data: Sequence-Structure Alignment: Align the protein sequence with its predicted or known 3D structure to correlate specific structural elements.

ProFunc is a metaserver that combines sequence, global structure, and local structure-based methods to obtain a set of function predictions from which one might seek consensus.
First, the protein of unknown function is analyzed by numerous sequence searches.



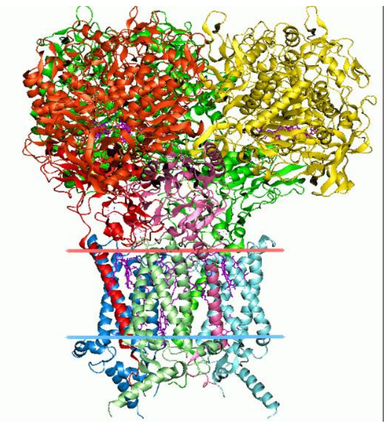
BLAST analysis scans both the PDB and UniProt and uses multiple sequence alignment to determine sequence similarities and detect sequence motifs. Gene neighbors are also examined based on the query protein's predicted location within the genome. The genes located near each other are often functionally related or functionally similar.

Structure-based analyses are performed on the protein of interest.This involves searching a number of databases for global folds or local structures that are similar to the query protein. secondary structure matching (SSM) evaluates the secondary structure elements (SSEs) of the query protein of unknown function and compares them to the SSEs of protein structures within its database. The algorithm retrieves high, strong matches and superimposes the structures with the query protein to give a root mean square deviation (RMSD) so that a common number can be used to compare the results. Finally, ProFunc utilizes other servers to search for 3D templates of proteins with known binding sites.

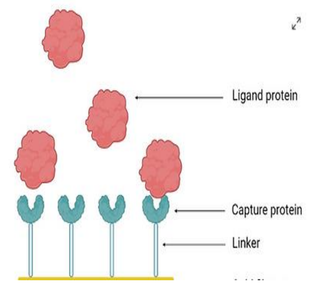
Applications

1. **Drug Discovery:**
Understanding the function of proteins is critical in identifying potential drug targets.
Inference from protein sequences and structures aids in predicting how a protein may respond to specific drugs.

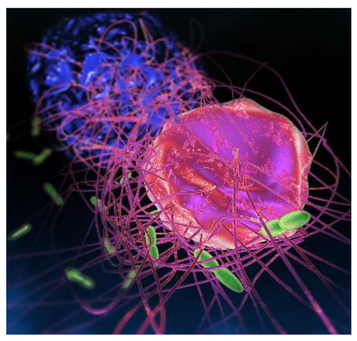
2. **Structural Bioinformatics Tools:**
The development and improvement of bioinformatics tools for sequence and structure analysis contribute to advancements in various applications.
Tools such as BLAST, Pfam, and others play a crucial role in inferring protein function.



3. **Protein-Protein Interactions:**
Inferring protein function is crucial for deciphering complex networks of protein-protein interactions.
Understanding how proteins interact provides insights into cellular processes and signaling pathways.



4. **Disease Mechanisms:**
Investigating the functions of proteins associated with diseases helps elucidate the underlying mechanisms.
Identifying proteins critical to disease pathways can guide the development of targeted therapies.



Observation

1. **Interconnectedness of Sequence and Structure**
2. **Bioinformatics as a Cornerstone**
3. **Potential for Personalized Medicine**
4. **Need for Experimental Validation**
5. **Technological Advancements Driving Progress**