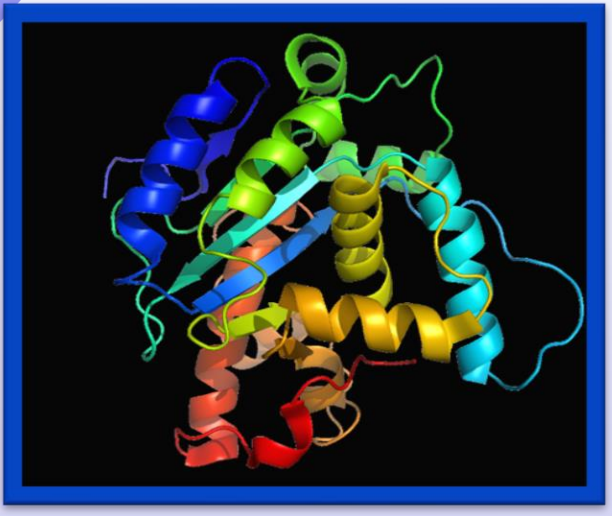


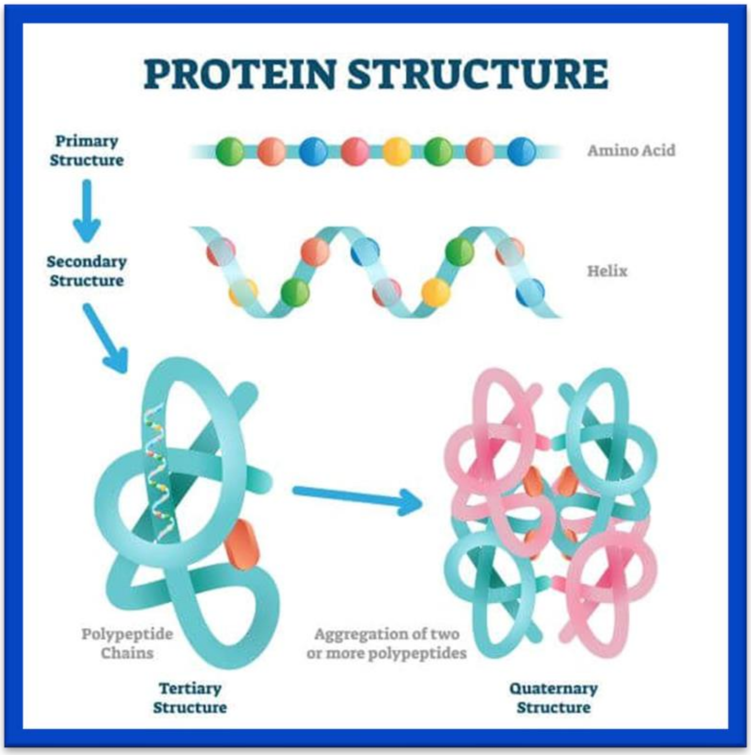
Protein Structure Prediction

* Introduction:

Proteins are essential biochemical components. They serve as the main component of many bodily structures, and a lot of drugs and medical substances are made with the help of natural protein-comprised structures. And a scientific process that helps to understand protein structures and functions and their relationship, is the process of digitally **predicting the protein structure**.

* Explanation:

Predicting the protein structure is a digital way of hypothesizing the three-dimensional structure of proteins from their amino acid sequences, which means predicting the secondary, tertiary, and quaternary structures from the primary structure of the protein.

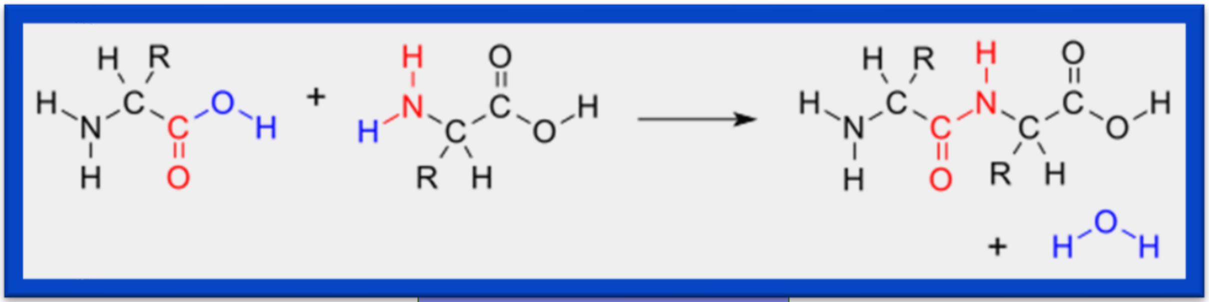


Protein structure prediction is one of the most important goals of computational biology, and as mentioned previously, it also helps in drug and enzyme designing, making it very useful in the branches of medicine and biotechnology.

*General Protein Structure:

Proteins are folded into three-dimensional shapes categorized by their amino acid sequences. Proteins have four levels of complexity: **Primary**, **Secondary**, **Tertiary**, and **Quaternary**.

The primary structure of proteins is basically a linear chain of amino acids by which it's defined. Amino acids are linked together to form a Polypeptide chain. Each amino acid is linked to the next by **peptide bonds**.



Peptide Bond

Proteins consist of 20 different amino acids, and these can be classified according to the chemical structure of the side chains.

Protein structures can be classified into fold families and superfamilies without the focus on its function.

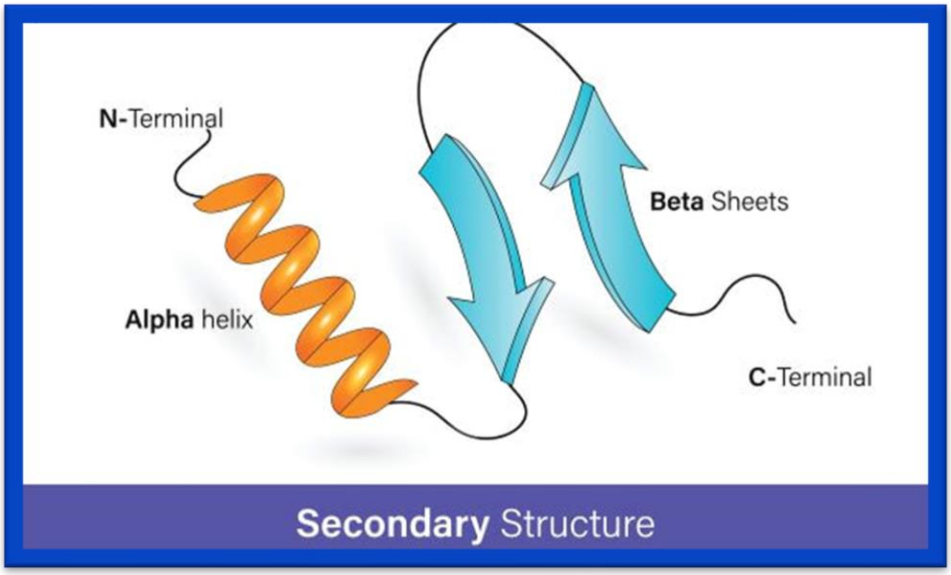
* Secondary Structure of Proteins and its Prediction:

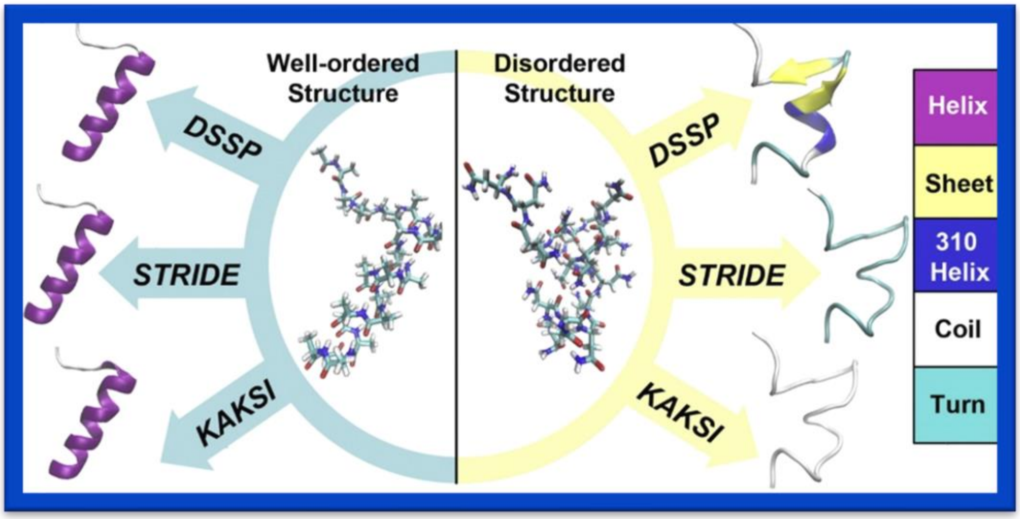
The secondary structure of proteins has specialized elements, which are **α-helices**, and **β-sheets**. These elements can vary widely in length and are connected by **loops and turns**.

A-helices are the more common secondary structure element of a protein than β-sheets, and in it, a hydrogen bond is formed between the N-H group and the C=O group of the amino acid.

The general method of predicting the secondary structure of a protein relies on the amino acid sequence and secondary structure elements. This data is then compared to results of already developed algorithms, **like the DSSP algorithm**.

The modern methods of predicting the secondary structure were claimed to reach 80% accuracy after the help of sequence alignments and machine learning.





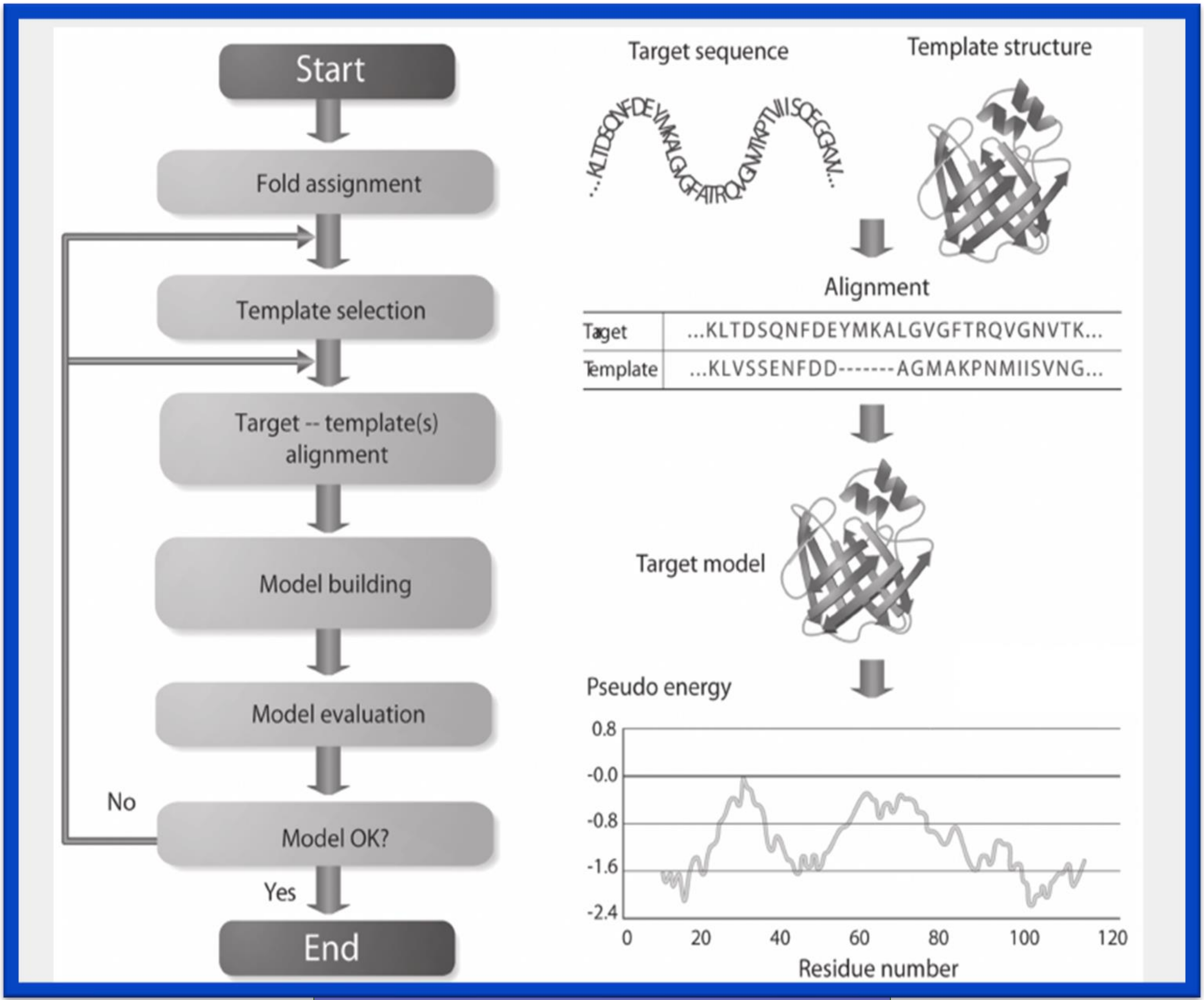
* Tertiary Structure of Proteins and its Prediction:

This refers to the three-dimensional arrangement of the atoms that constitute a protein molecule. It relates the precise spatial coordination of secondary structure elements and the location of all functional groups of a single polypeptide chain.

Most tertiary structure modelling methods are optimized for modelling the tertiary structure of **single protein domains** (domains are regions of the protein chain that are independently folding). A step called **Domain Parsing** (domain boundary prediction), is usually done first to split a protein into potential structural domains. And then, with the help of machine learning, the rest of the tertiary structure prediction can be done **comparatively from known structures**. The structures of individual domains are then **docked together in a process called Domain Assembly to create the final tertiary structure**.

But how does that comparison process (Comparative Protein Modelling) work? Comparative protein processes use already solved or predicted structures as a foundation or a starting point. This process is effective, because although the number of actual proteins is massive, there is a limited set of tertiary structural motifs or elements to which most proteins belong.

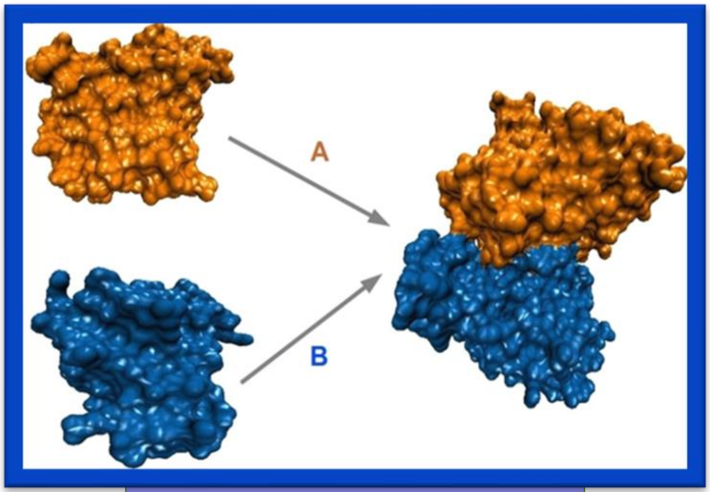
It has been suggested that there are only around 2000 distinct protein folds in nature, but there are many millions of different proteins. Comparative protein modeling can combine with evolutionary covariation in the structure prediction.



Comparative Protein Modelling process

* Quaternary Structure of proteins and its Prediction:

It’s the complex structure of many proteins, and it’s comprised of two or more protein subunits. The quaternary structures of proteins are known or can be predicted with high accuracy. **Protein–protein docking methods** can be used to predict the structure of those complexes. Protein-protein docking is basically the computational modelling of complexes made up of two or more interacting biological macromolecules.



Protein-protein docking

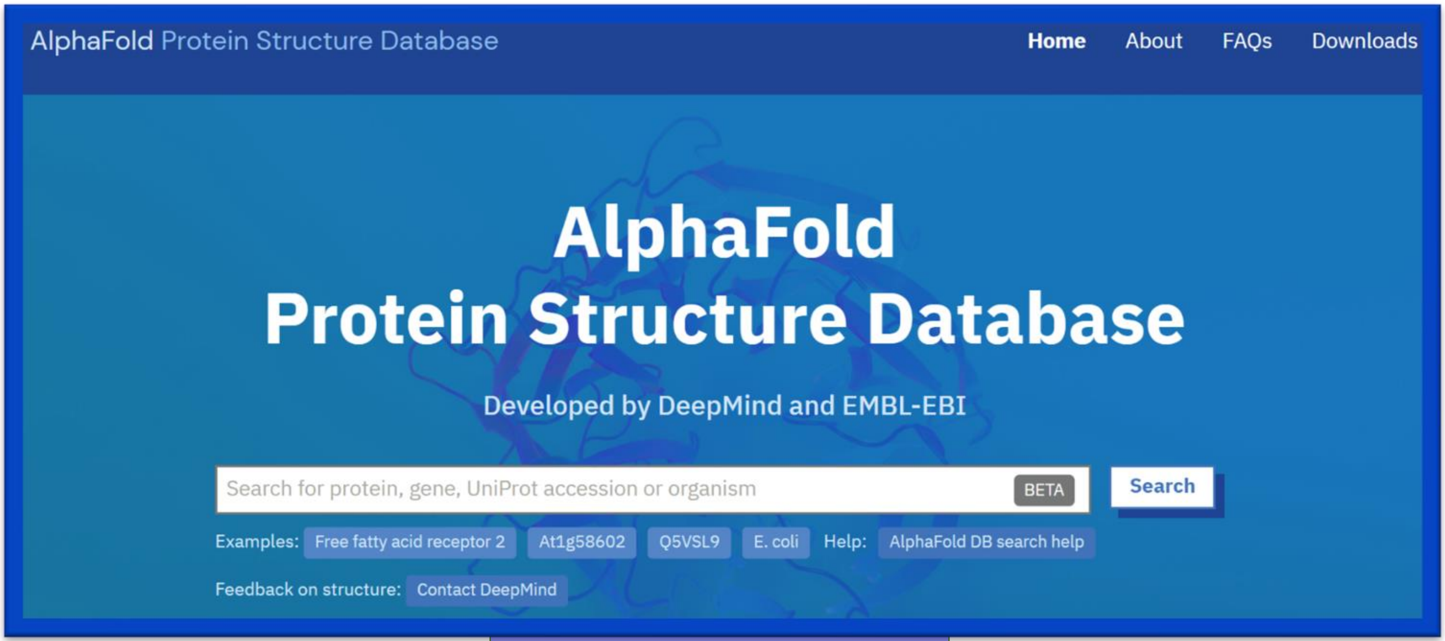
Information on the effect of mutations at specific sites on the complex helps to understand the complex structure, functionality, and to guide docking methods.

* Protein Structure Prediction Software:

There's a lot of computer software tools for protein structure prediction. The concept of these software includes the aforementioned Comparative Protein Modelling, fold recognition, and transmembrane helix and signal peptide prediction. There are protein structure prediction programs that use AI technology to predict the structure, and one program that was reported to have the best performance is a program called **AlphaFold**.

AlphaFold relies on a **Neural Network**, which directly predicts the 3D coordination of all non-hydrogen atoms for a given protein using the amino acid sequence and aligned homologous sequences (Sequences sharing ancestry).

The AlphaFold network consists of a trunk which processes the inputs through repeated layers and a structure module which introduces an **explicit 3D structure (the output)**.



AlphaFold Protein Structure Database

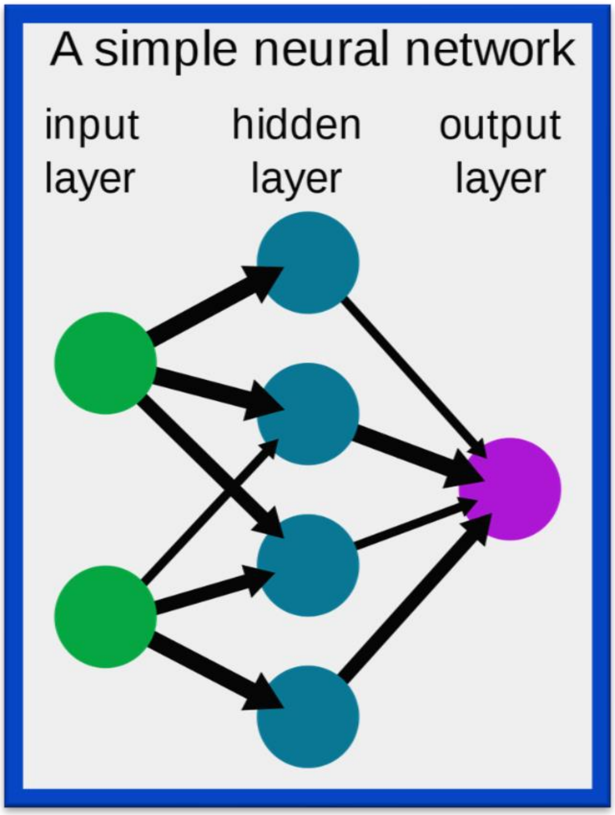
Home About FAQs Downloads

AlphaFold Protein Structure Database

Developed by DeepMind and EMBL-EBI

Search for protein, gene, UniProt accession or organism
BETA
Search
Examples: Free fatty acid receptor 2
At1g58602
Q5VSL9
E. coli
Help: AlphaFold DB search help
Feedback on structure: Contact DeepMind

AlphaFold page



A simple neural network