SHANKAR
IAS PARLIAMENT
Information is Empowering

Protein Structure Prediction and Computational Design

Why in News?

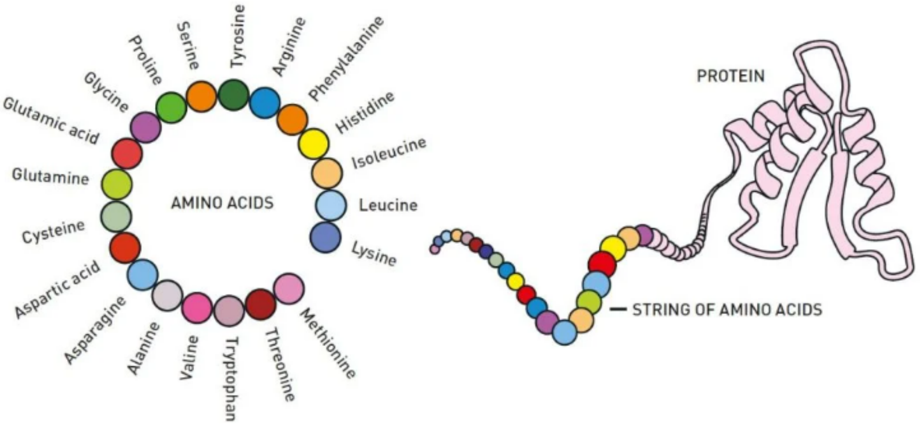
The Nobel Prizes for 2024 in Chemistry was awarded to David Baker “for computational protein design” along with Demis Hassabis and John M. Jumper “for protein structure prediction”.

How does a protein find its unique structure?

- **Proteins** - They are fundamental to almost all biological processes and are built from 20 amino acids joined into long strings.

  Haemoglobin is a protein that transports oxygen and insulin helps absorption of glucose from blood.

- **Protein Structure** - The string of amino acids twists and folds into a distinct three-dimensional structure which gives proteins their function.
- All the information about how the protein folds is present in the amino acid sequence.



- Anything that impacts protein production can have consequences for human health.
- **Enzymes** - Certain kinds of proteins, called enzymes, can speed up biochemical reactions within the body, while others can provide structural support to cells and tissues.
- Then there are some proteins that help in immune response, while others can store nutrients or energy.
- **Decoding Proteins** - Traditional methods of decoding the structure of proteins, through x-ray crystallography, is a laborious and time-consuming process taking months/years.
- **Database** - Over the years, the structures of thousands of proteins have been

catalogued and a database containing all known sequences of amino acids has also been created.

What are AlphaFold and Rosetta?

- **AlphaFold** – It is an artificial intelligence model that could predict the structures of millions of proteins.
- Hassabis and Jumper received the other half of the prize for developing this model.
- **Use of AI** – AlphaFold was upgraded using AI neural networks called transformers.

AI neural networks can find patterns in enormous amounts of data in a more flexible manner.

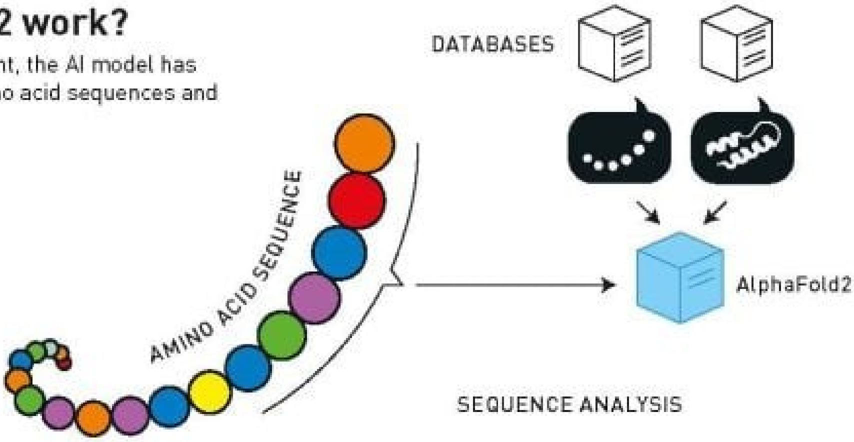
- AlphaFold2 was trained on the vast information in the databases of all known protein structures and amino acid sequences.
- The AI tool predicts the structures of proteins using known sequences of amino acids from the database.
- These predictions were then matched to catalogued protein structures in the other database.
- **Accuracy** - The AI model is not perfect, but it estimates the correctness of the structure it has produced, so researchers know the reliability of their prediction.

How does AlphaFold2 work?

As part of AlphaFold2's development, the AI model has been trained on all the known amino acid sequences and determined protein structures.

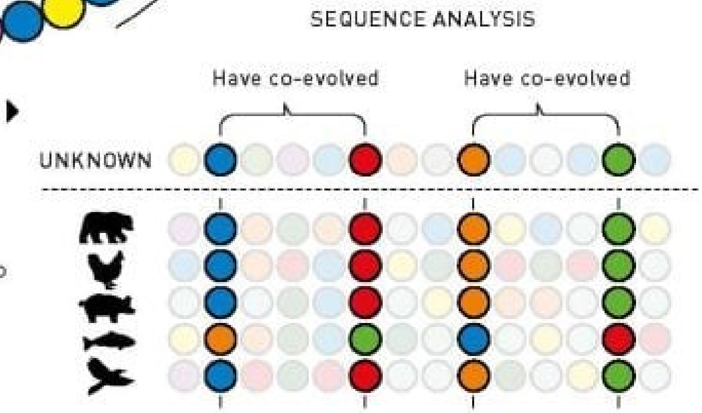
1. DATA ENTRY AND DATABASE SEARCHES

An amino acid sequence with unknown structure is fed into AlphaFold2, which searches databases for similar amino acid sequences and protein structures.

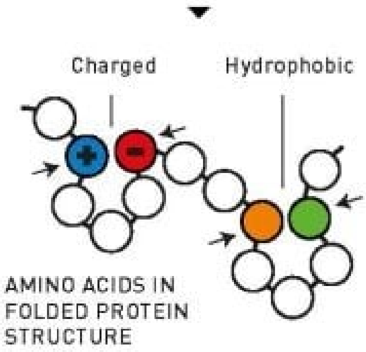


2. SEQUENCE ANALYSIS

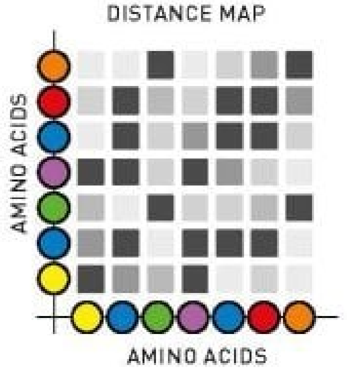
The AI model aligns all the similar amino acid sequences – often from different species – and investigates which parts have been preserved during evolution.



In the next step, AlphaFold2 explores which amino acids could interact with each other in the three-dimensional protein structure. Interacting amino acids co-evolve. If one is charged, the other has the opposite charge, so they are attracted to each other. If one is replaced by a water-repellent (hydrophobic) amino acid, the other also becomes hydrophobic.



Using this analysis, AlphaFold2 produces a distance map that estimates how close amino acids are to each other in the structure.

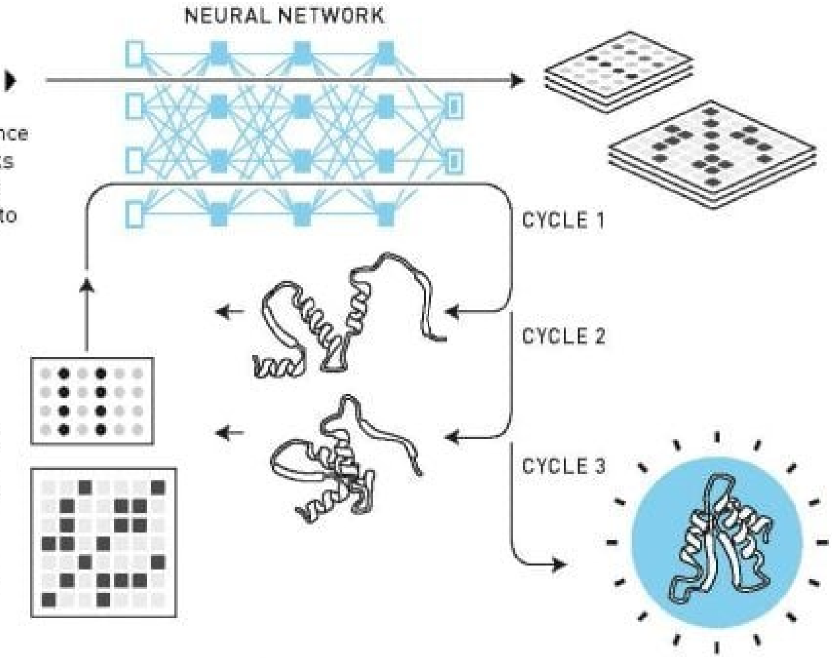


3. AI ANALYSIS

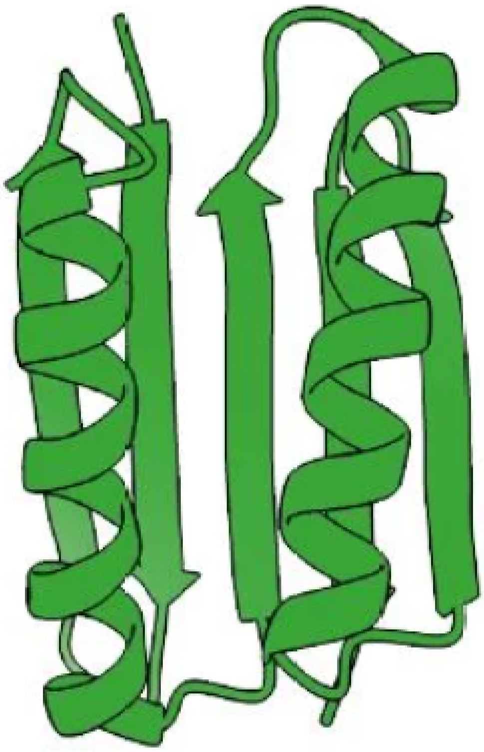
Using an iterative process, AlphaFold2 refines the sequence analysis and distance map. The AI model uses neural networks called transformers, which have a great capacity to identify important elements to focus on. Data about other protein structures – if they were found in step 1 – is also utilised.

4. HYPOTHETICAL STRUCTURE

AlphaFold2 puts together a puzzle of all the amino acids and tests pathways to produce a hypothetical protein structure. This is re-run through step 3. After three cycles, AlphaFold2 arrives at a particular structure. The AI model calculates the probability that different parts of this structure correspond to reality.



- **Rosetta** – It is the software developed by David Baker to predict protein structures by entering the amino acid sequences as input.
- **Creating New Protein Structure** – The Rosetta process is reversed to get the amino acid sequence for the desired protein structure, which enabled to create entirely new proteins.
- The new proteins can perform functions that naturally-synthesised proteins are not designed to.
- **Top 7** - It was the first protein constructed by Rosetta that was entirely different to all known existing proteins.



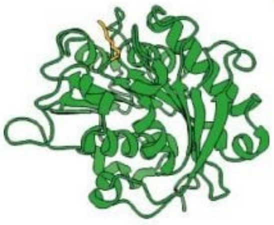
Top 7 Protein

What are the significances of this invention?

- **Better Understanding of Protein** – Their work enabled the better visualisation of protein structure and creating new proteins based on our requirements.
- AlphaFold2 has calculated the structure of all human proteins and predicted the structure of virtually all the 200 million proteins that researchers have so far discovered when mapping Earth’s organisms.
- **Time Saving** – Calculating the protein structure took years once and now takes just a few minutes.
- **Better Understanding of Life** - It allows us to better understand how life functions.
- **Clarity on Diseases** – Knowing the relation between protein structure and amino acids , helped in understating some diseases.
- **Understanding Antibiotic Resistance** – Understanding the bacterial enzymes that cause antibiotic resistance will help in discovering ways to prevent it.
- **Plastic Decompostion** – Synthetic enzymes and proteins that can decompose plastic can be designed using this tool.

- **Development of Bioproducts** - The ability to create proteins can lead to new nanomaterials, targeted pharmaceuticals, more rapid development of vaccines, minimal sensors and a greener chemical industry.

2022: Part of a huge molecular structure in the human body. More than a thousand proteins form a pore through the membrane surrounding the cell nucleus.



2022: Natural enzymes that can decompose plastic. The aim is to design proteins that can be used to recycle plastic.

2023: A bacterial enzyme that causes antibiotic resistance. The structure is important for discovering ways of preventing antibiotic resistance.

References

1. The Indian Express | Nobel Prize in Chemistry 2024
2. The Hindu | Nobel Prizes for 2024 in Chemistry
3. Nobel Prize | 2024 Chemistry

SHANKAR
IAS PARLIAMENT
Information is Empowering