2024 Chemistry Nobel

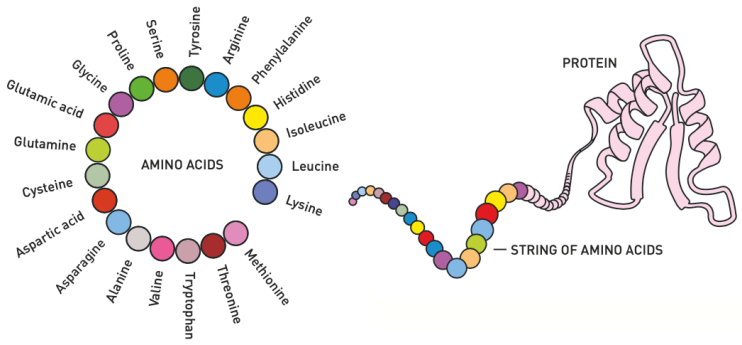
Michael Wollowski

Pertinent Information summarized from:

https://www.nobelprize.org/uploads/2024/10/advanced-chemistryprize2024.pdf and

https://www.nobelprize.org/uploads/2024/10/popular-chemistryprize2024-3.pdf

Proteins



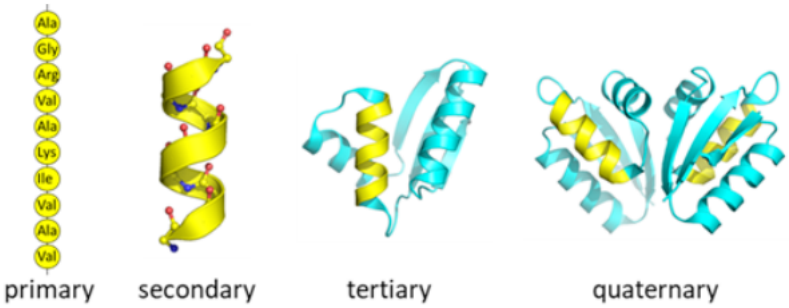
- How is the exuberant chemistry of life possible? Proteins
- They are generally built from 20 amino acids that can be combined in endless ways.
- Using the information stored in DNA as a blueprint, the amino acids are linked together in our cells to form long strings.
- Then the magic of proteins happens: the string of amino acids twists and folds into a distinct sometimes unique – three-dimensional structure:

Protein Folding

- This structure is what gives proteins their function.
- Some become chemical building blocks that can create muscles, horns or feathers.
- Others may become hormones or antibodies.
- Many of them form enzymes, which drive life's chemical reactions with astounding precision.
- The proteins that sit on the surfaces of cells are also important, and function as communication channels between the cell and its surroundings.

Protein Folding

- Proteins have several structures.
- Predicting them is very hard.



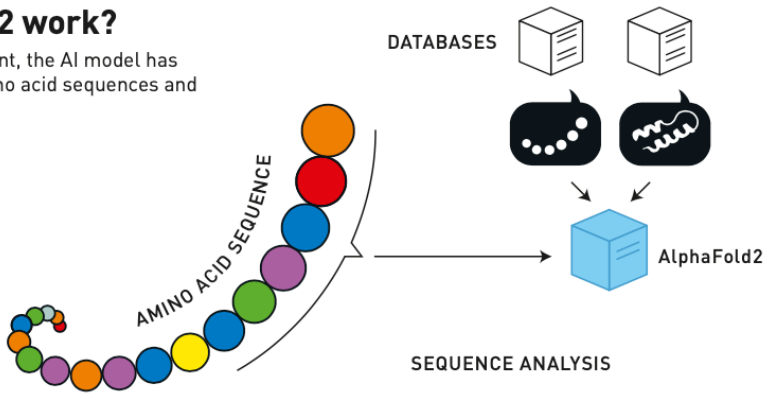
AlphaFold2

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

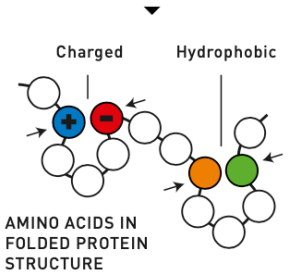
An amino acid sequence with unknown structure is fed into AlphaFold2, which searches databases for similar amino acid sequences and protein structures.

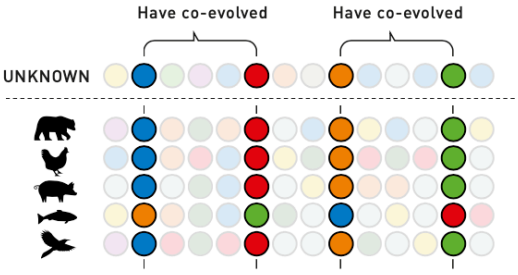


2. SEQUENCE ANALYSIS

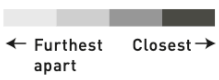
The AI model aligns all the similar amino acid sequences – often from different species – and investigates which parts have been preserved during evolution.

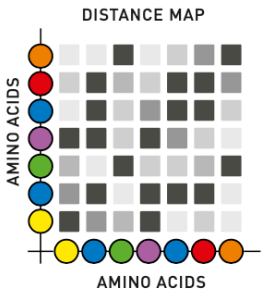
In the next step, AlphaFold2 explores which amino acids could interact with each other in the three-dimensional protein structure. Interacting amino acids co-evolve. If one is charged, the other has the opposite charge, so they are attracted to each other. If one is replaced by a water-repellent (hydrophobic)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





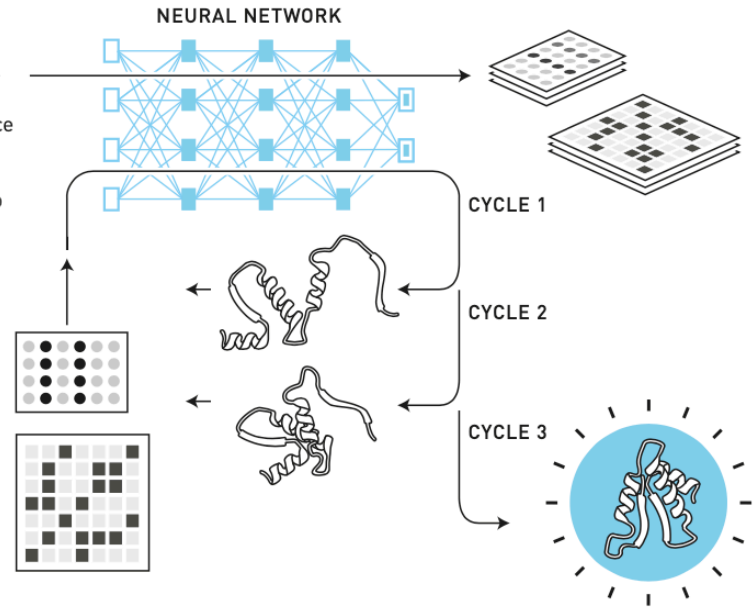
Using this analysis, AlphaFold2 produces a distance map that estimates how close amino acids are to each other in the structure.



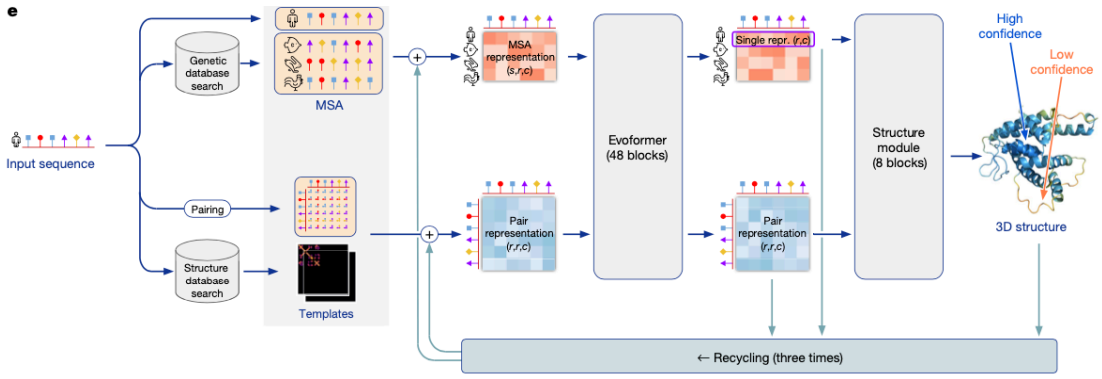


3. AI ANALYSIS
Using an iterative process, AlphaFold2 refines the sequence analysis and distance map. The AI model uses neural networks called transformers, which have a great capacity to identify important elements to focus on. Data about other protein structures – if they were found in step 1 – is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



More Technical View of the Main Processing



Transformers

- AlphaFold1 used a CNN architecture
- AlphaFold2 used a transformer architecture

Pattern Recognition

- Just like AlphaGoZero, MNIST CNN, and Transformers in general,
- These systems are trained to recognize patterns.
- They are not just pattern matching systems.