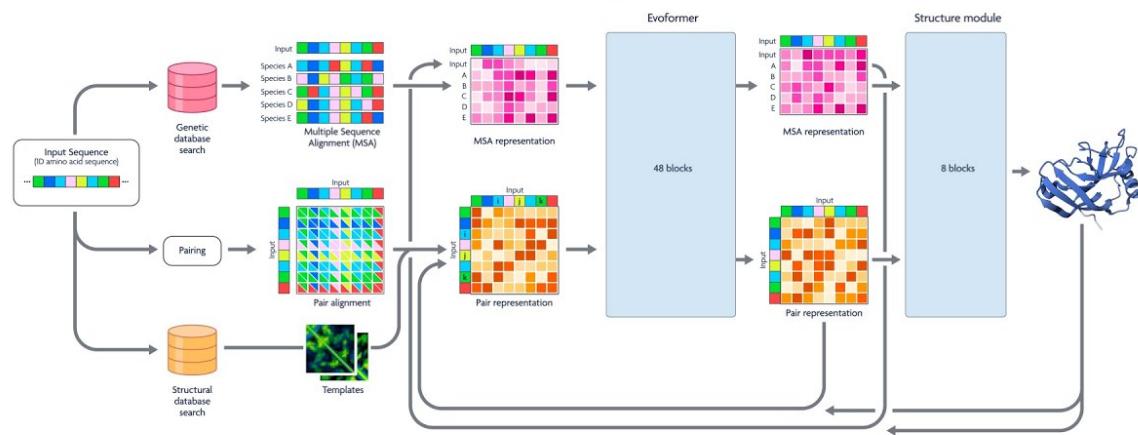


June 30, 2024

By [Alejandro Avella](#)

# What Is AlphaFold?



## Summary:

- AlphaFold is an artificial intelligence system developed by Google DeepMind to predict the 3D structures of proteins from their amino acid sequences.
- Traditional methods like x-ray crystallography and NMR spectroscopy are slow and expensive; AlphaFold offers a rapid and accurate non-experimental alternative.
- AlphaFold 2 uses multiple sequence alignments (MSA) and a pair representation approach to model protein structures efficiently.
- The Evoformer, a unique neural network in AlphaFold 2, refines the MSA and pair representations through iterative calculations over 48 cycles.
- The final structure module integrates refined models to predict 3D atomic coordinates of proteins, useful for drug discovery, genetic variant effects, and protein engineering.

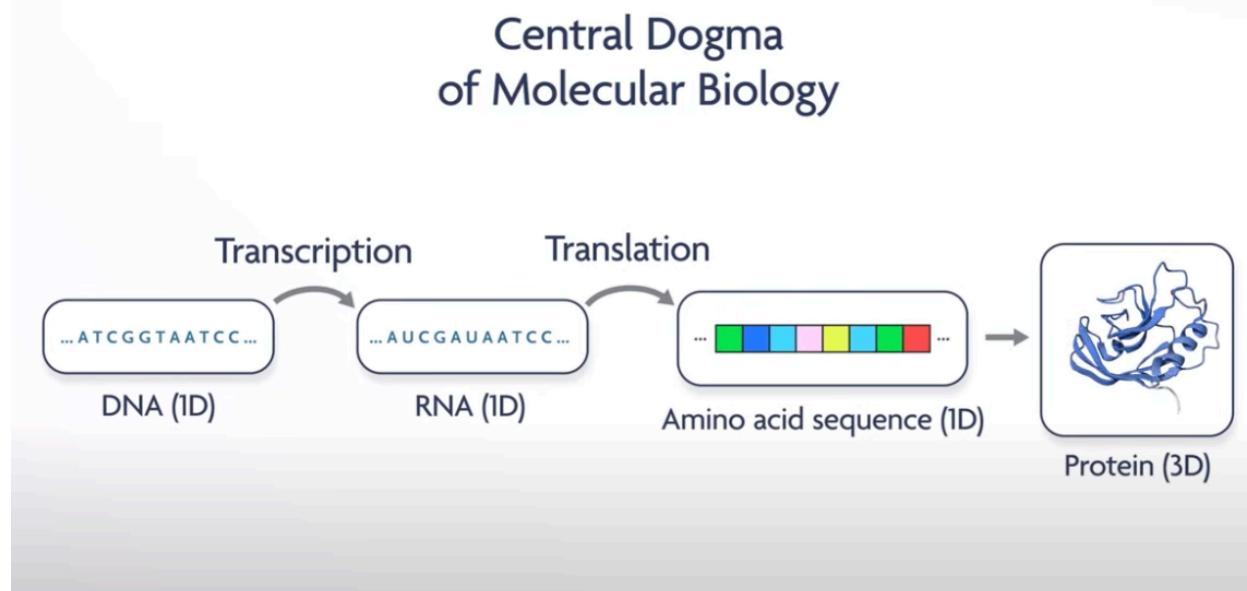
## References:

1. Jumper, J. et al. "Highly accurate protein structure prediction with AlphaFold." *Nature*, 596, pages 583–589 (2021). DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2)
2. Clinical Implications of Basic Research, "A Holy Grail — The Prediction of Protein Structure", by R.B. Altman 10.1056/NEJMcibr2307735 <https://nej.md/3PpWv4F>
3. What is AlphaFold? | NEJM <https://www.youtube.com/watch?v=7q8Uw3rmXyE&t=6s>
4. John Jumper Wikipedia's page: [https://en.wikipedia.org/wiki/John\\_M.\\_Jumper](https://en.wikipedia.org/wiki/John_M._Jumper)

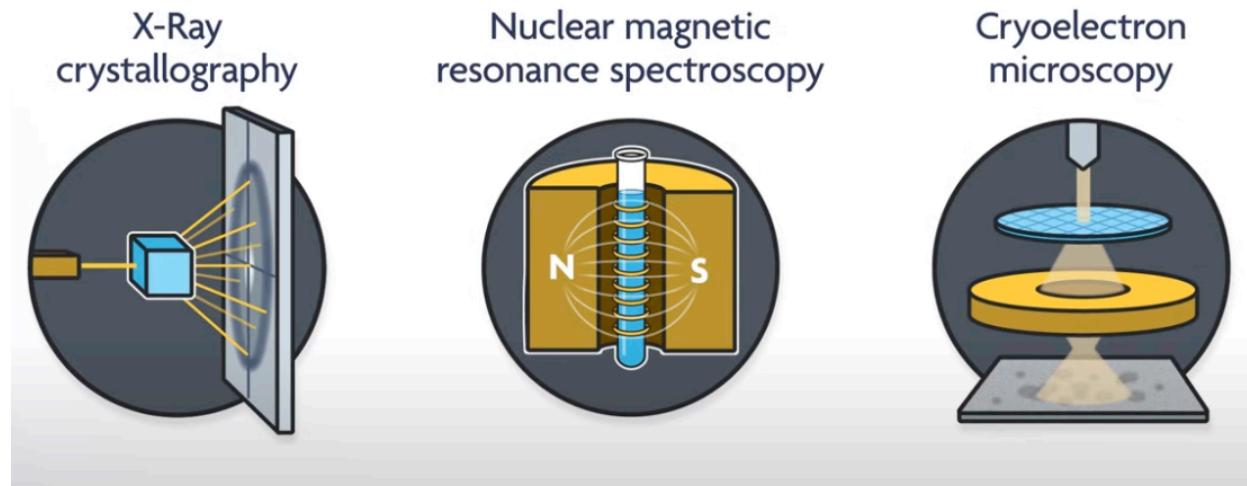
**Title: "[What Is AlphaFold? | NEJM](#)" Transcript:**

"For the last 50 years, researchers have been searching for a way to predict the 3D structure of proteins from the one-dimensional sequence of their amino acids. From the central dogma, we know that DNA is transcribed into RNA and then translated into an amino acid sequence which contains all the necessary information to reliably fold into a three-dimensional protein."

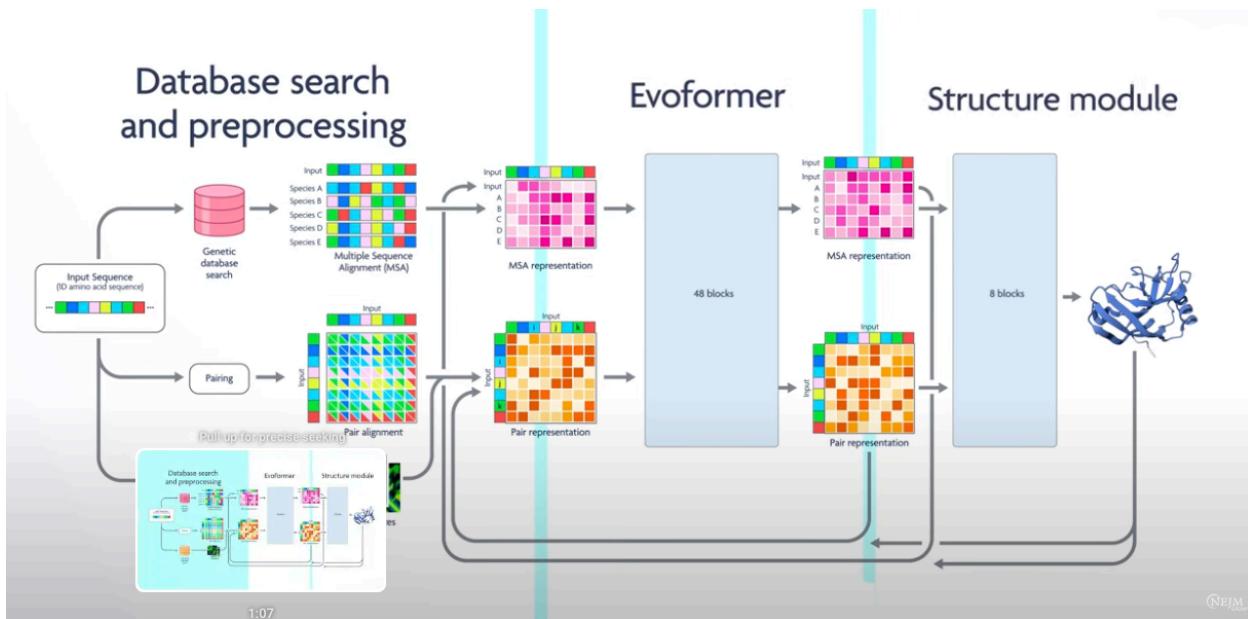
## Central Dogma of Molecular Biology



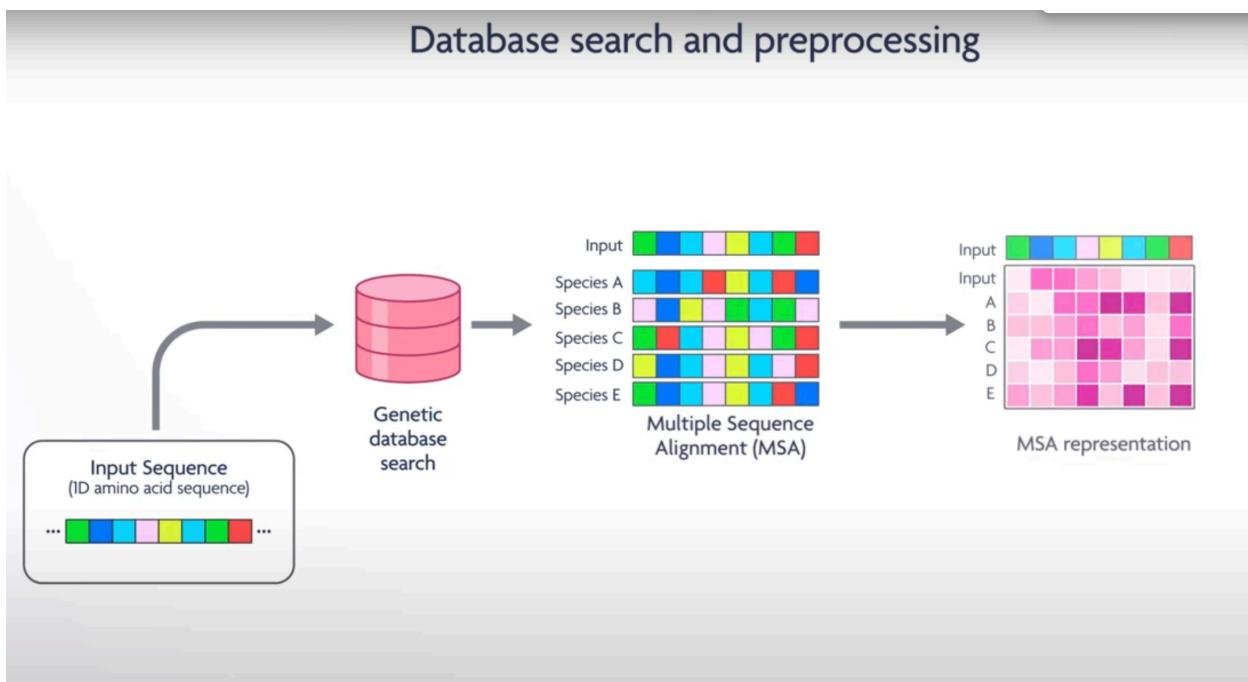
Researchers have successfully determined protein structure experimentally using methods such as x-ray crystallography, nuclear magnetic resonance spectroscopy and cryoelectron microscopy, but these methods are time consuming and expensive.



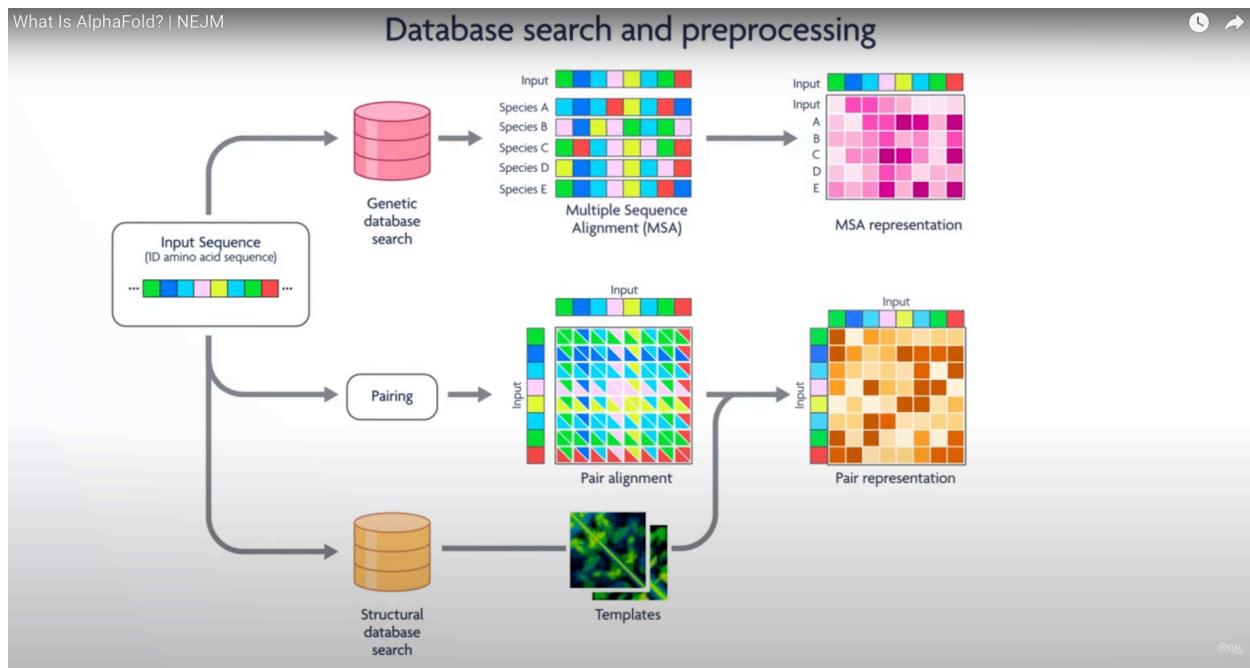
The Alphafold artificial intelligence system developed by Google Deepmind is the first non-experimental method that can rapidly accomplish this with accuracy comparable to experiment. But, how does Alphafold 2 the 2021 iteration of alpha fold speed up the process of obtaining useful 3D protein structure models. Here is a simplified diagram of alpha fold 2's architecture. We'll break it down into three sections



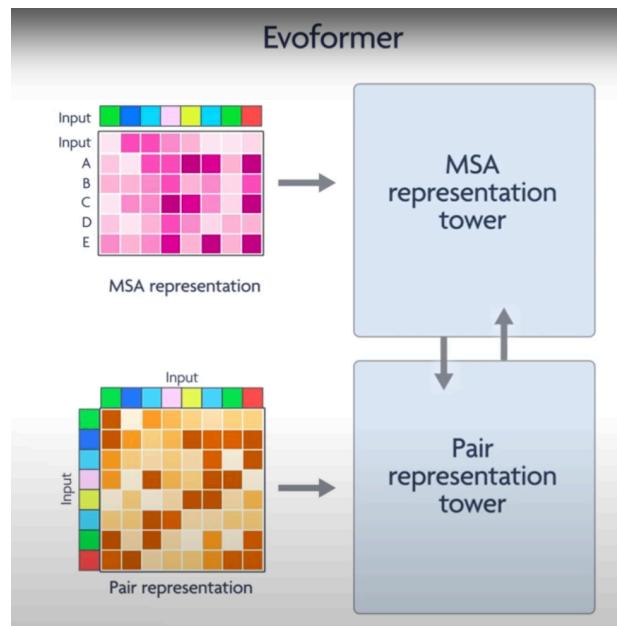
and start here when an input sequence of residues or amino acids is entered Alpha fold Compares it to several databases of protein sequences to extract similar sequences from various organisms and tissues to generate a multiple sequence alignment or MSA and an initial MSA representation



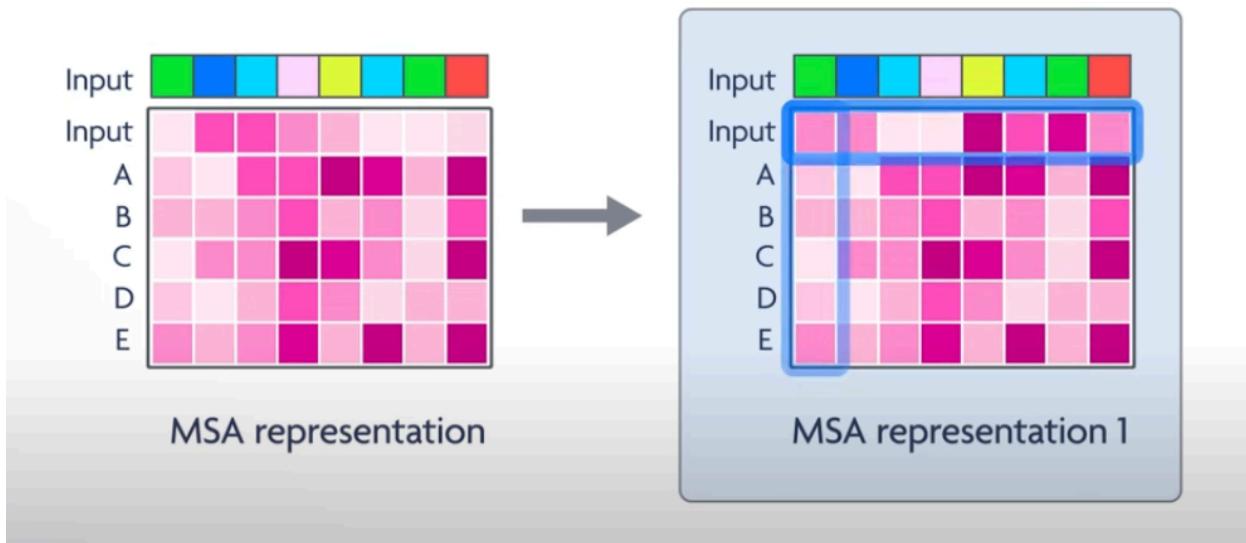
It also pairs the input sequence and searches databases for templates of similar sequence proteins with experimentally determined structures. This is then used to create an initial pair representation of the input sequence representing the relationship between every pair of residues within the target protein.



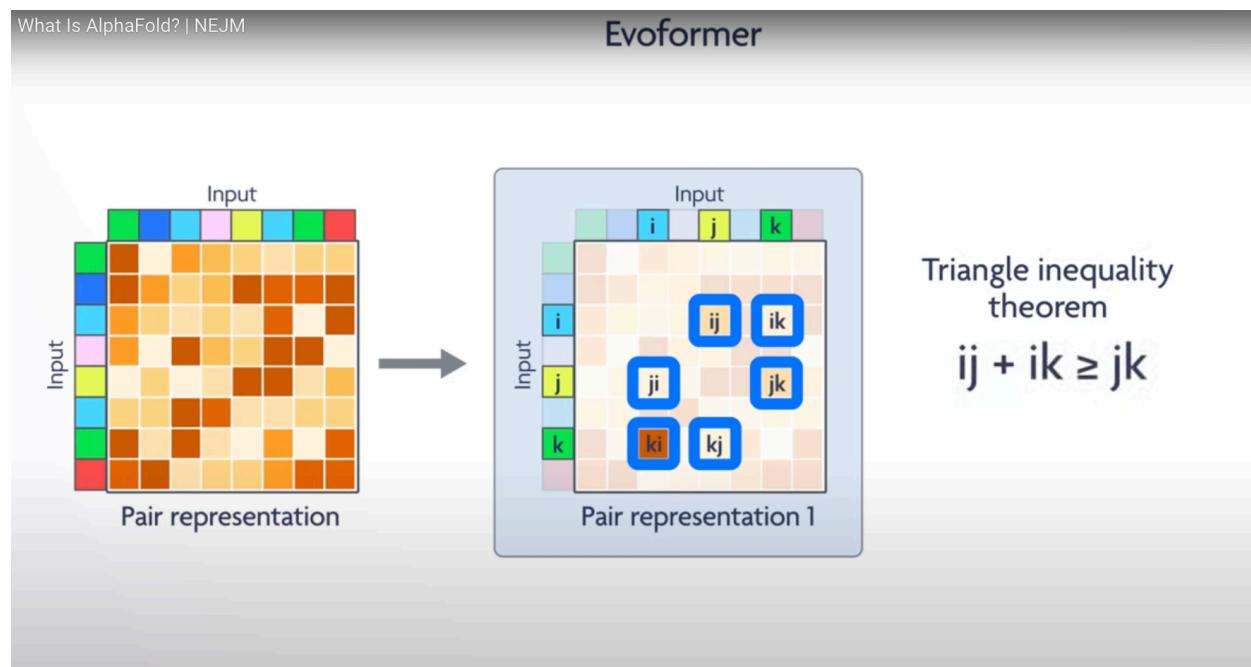
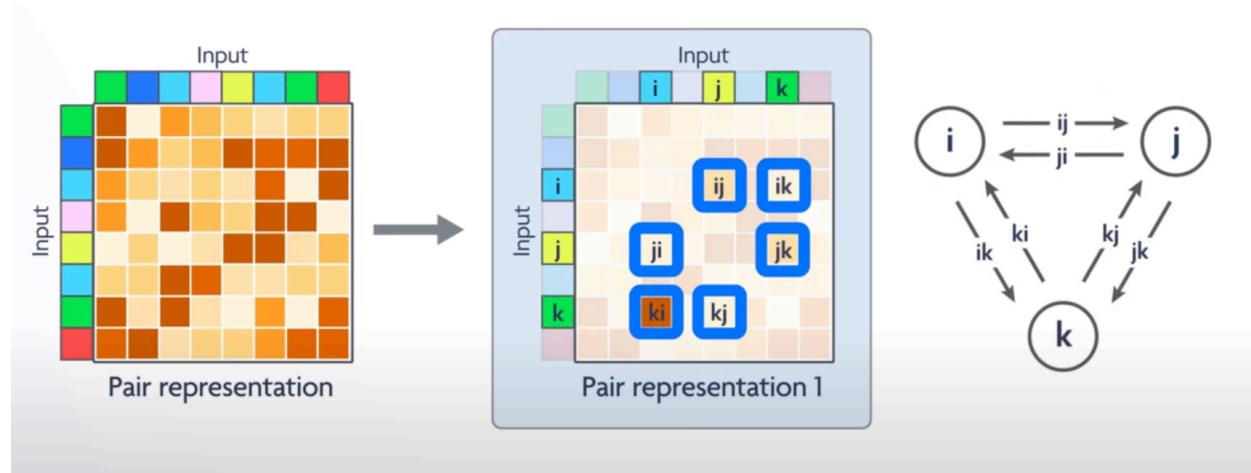
From here we can move on to section 2. The **Evoformer** is a neural network unique to Alphafold. It consists of two towers that can communicate information to each other.



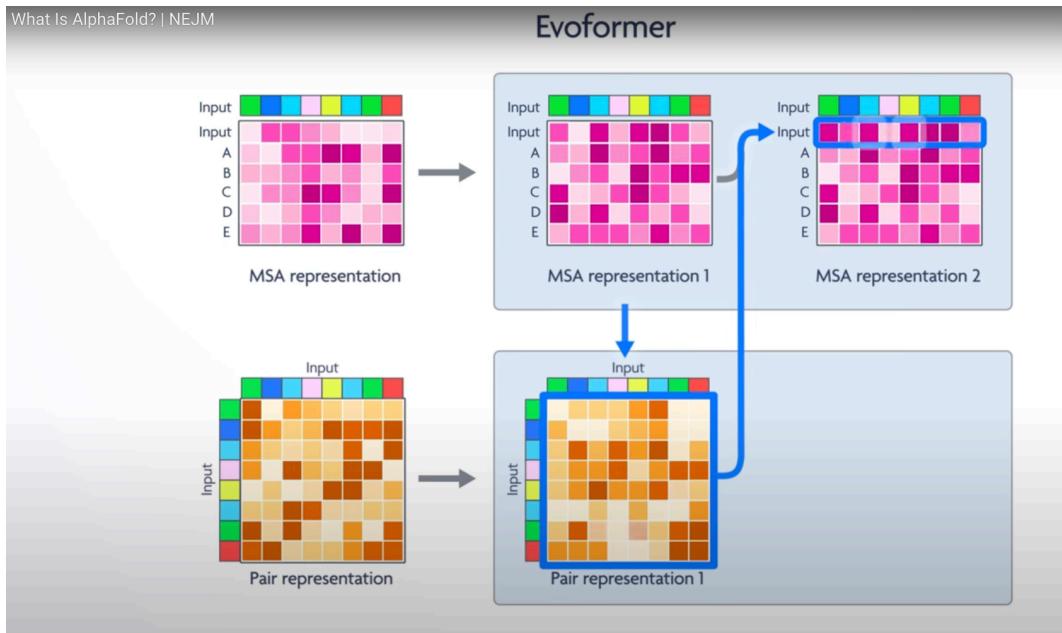
For the MSA representation the neural network prioritizes looking for row ***wise*** relationships between residue pairs in the input sequence, before considering column ***wise*** information that evaluates the importance of each residue, in context of the other sequences.

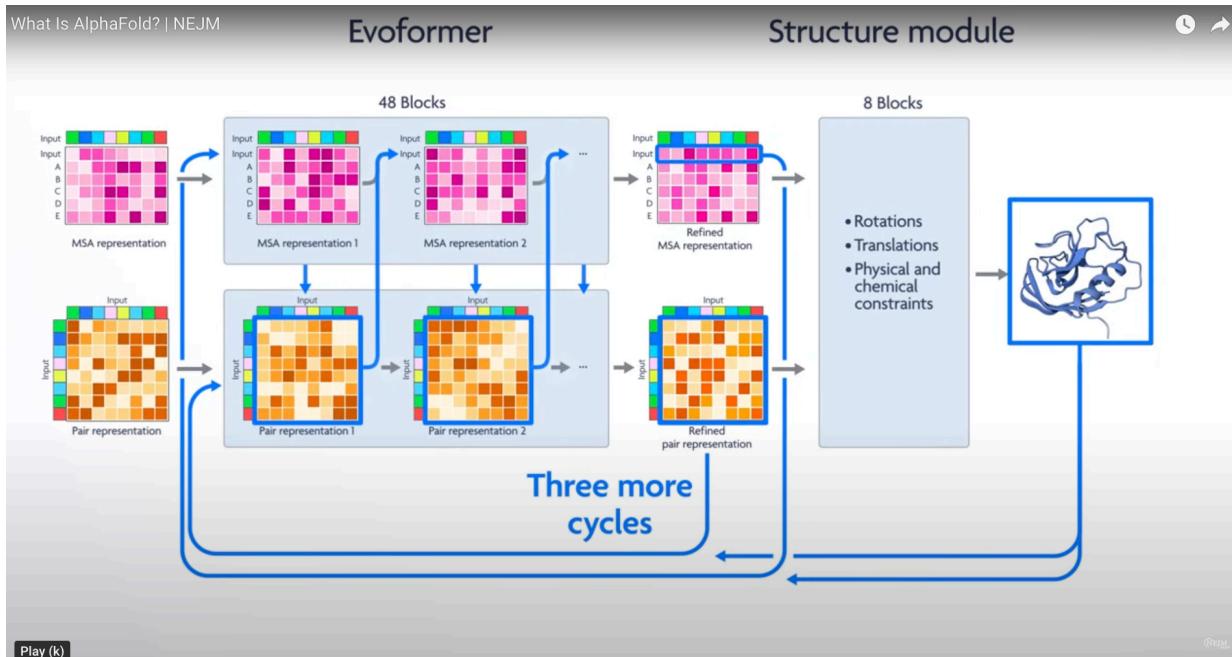


The pair representation tower evaluates the relationships between every two residues, which can be thought of as nodes to refine the proximities or edges between the two. It achieves this by triangulating the relationship of each node in a pair relative to a third node. The goal of this process is to help the network satisfy the triangle inequality theorem, where the sum of two edges on the triangle must be equal to or greater than the third.

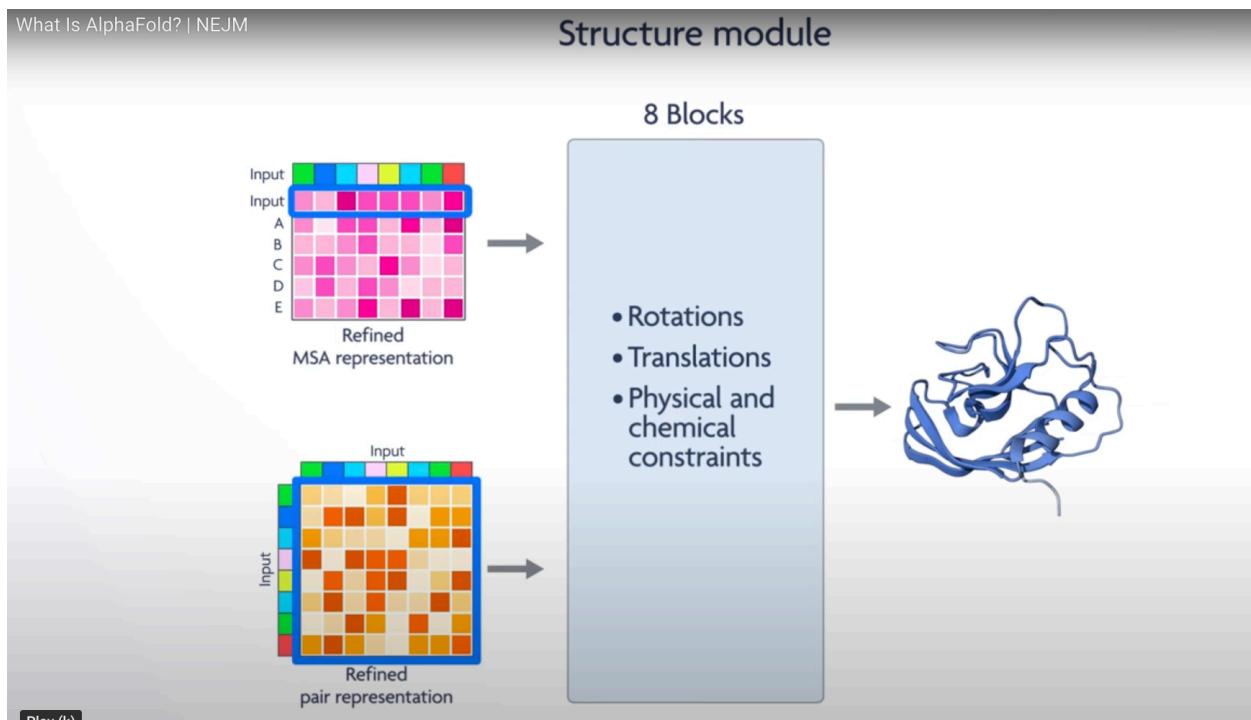


So, how and where do the two towers communicate. Before the pair representation calculates the triangulations on each edge. It considers the updated MSA residue relationships and updates each edge accordingly. The results of the pair representation data is then used in the MSA's row wise weighting of pair relationships in the input sequence prior to its next round of row and column wise evaluations. These individual and cross-communicated calculations happen 48 times in the Evoformer before creating the refined models of the initial MSA impair representations.

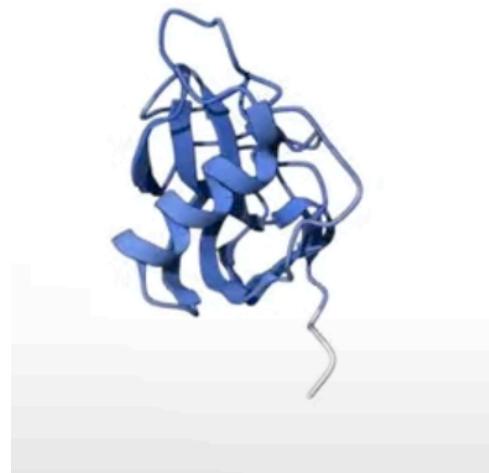




The final section involves another neural network called the **structure module**. It takes the refined models and performs rotations and translations on each amino acid revealing an initial guess of the 3D protein structure. It also applies physical and chemical constraints dictated by atomic bonds, angles and torsional angles. The refined models as well as the output of the structure module is iterated back through the Evoformer and structure module process three more times for a total of four Cycles before it arrives at the final result predicted 3D Atomic coordinates for the proteins 3D structure.



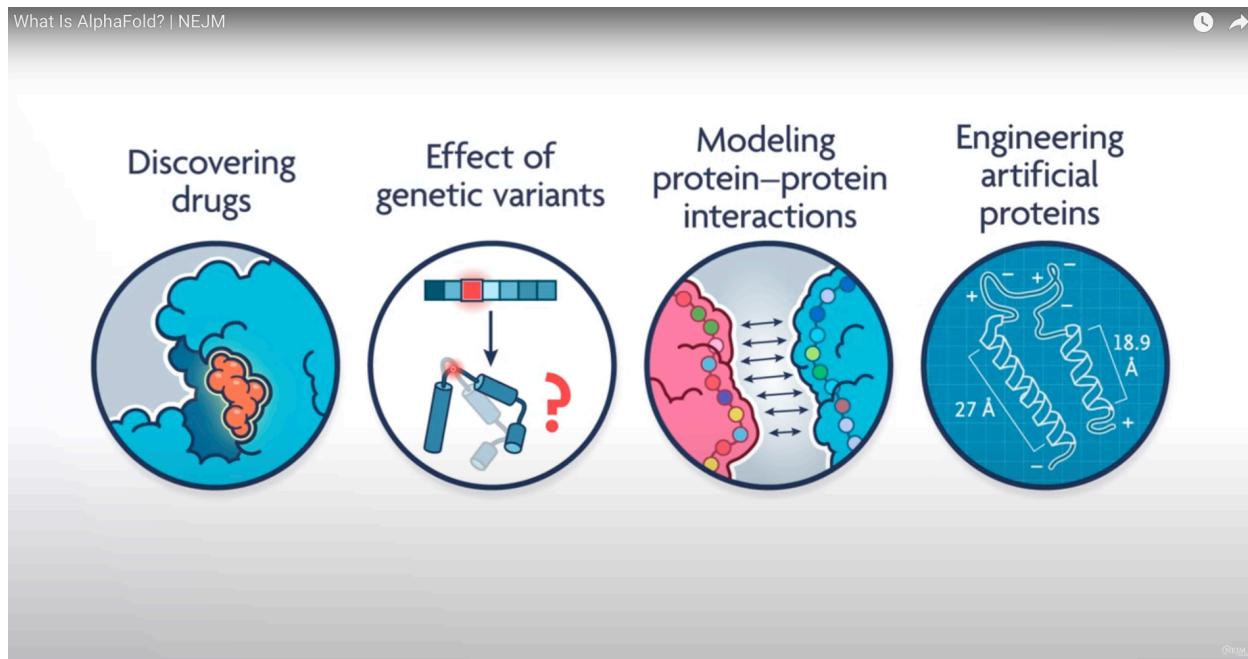
## Predicted 3D coordinates



It's important to note that experimentally determine 3D structures of proteins are almost always more accurate than predicted 3D structures and should be used preferentially when available potential applications of predicted.

3D Atomic coordinates from alpha fold include discovering drugs that bind tightly to protein pockets, estimating the effect of genetic variants that change the amino acids of a protein on protein structure and function, modeling interfaces of proteins that engage in protein-protein interactions and Engineering proteins with new functions for medicine biotechnology Agriculture and the broader environment

What Is AlphaFold? | NEJM



The screenshot shows a mobile browser displaying an article from the New England Journal of Medicine (NEJM) titled "What Is AlphaFold?". The page features a navigation bar at the top with icons for back, forward, and search. Below the title, there are four circular icons, each representing a different application of protein structure prediction:

- Discovering drugs:** Shows a blue protein structure with an orange drug molecule bound to its active site.
- Effect of genetic variants:** Shows a DNA sequence with a red mutation, followed by a blue test tube and a question mark, indicating how predictions can help understand the impact of genetic changes.
- Modeling protein–protein interactions:** Shows two protein structures, one pink and one blue, interacting with each other via several double-headed arrows representing interactions.
- Engineering artificial proteins:** Shows a blue protein structure with a zigzagging backbone, labeled with dimensions of 27 Å and 18.9 Å, representing the ability to design new protein architectures.

This video only touches on the concepts involved in Alpha fold 2's architecture to learn more read the original manuscript published in nature also be sure to check out the related article a "*Holy Grail the prediction of protein structure*" by Russ Altman published in the New England Journal of Medicine"