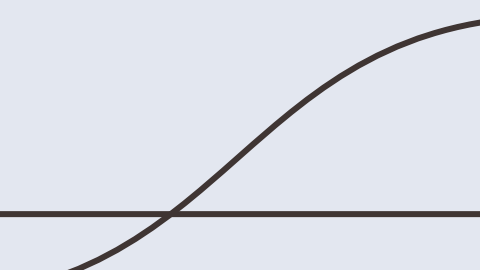




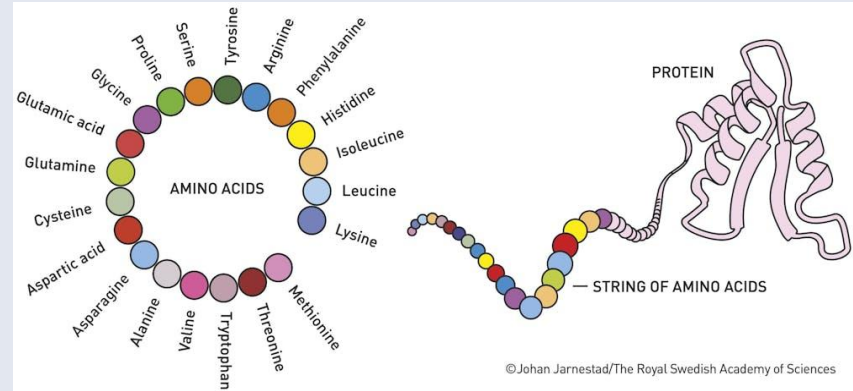
AlphaFold: Predicting Protein Structures

By Kathy Lin
Buckingham Browne and Nichols School
December 8, 2024



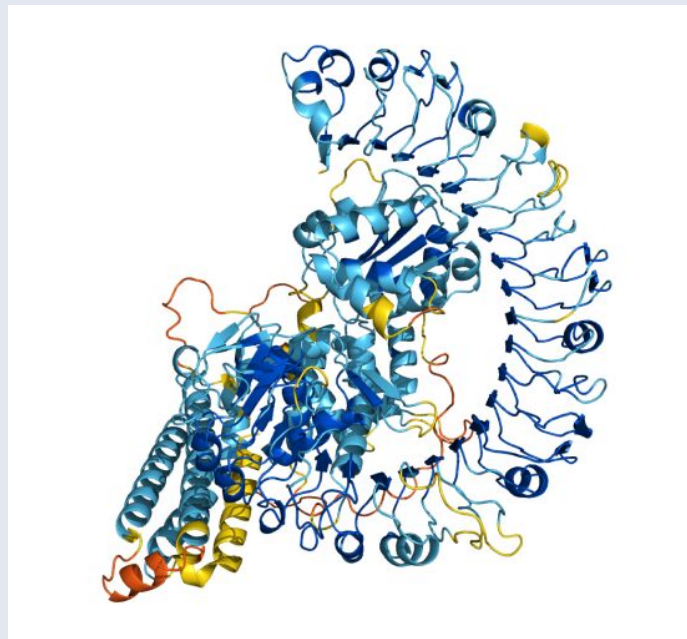
The Protein Folding Problem

- Proteins carry out all of the important functions in all living organisms
- The 3D structure of a protein molecule is essential for its function
- How to determine the structure of proteins based only on its amino acid sequence
- Tons of possible protein folding combinations

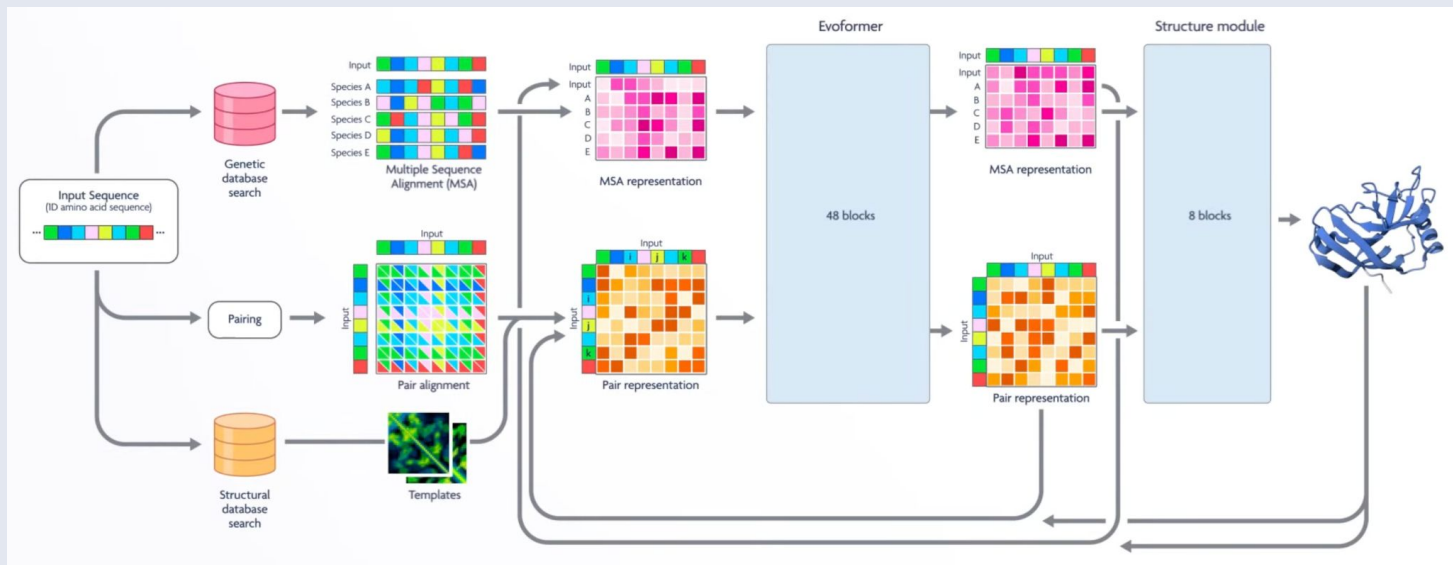


What is AlphaFold?

- Program developed by Google Deepmind
- Uses artificial intelligence to predict the structures of proteins
- Inputs an amino acid sequence and output an image similar to the one on the right

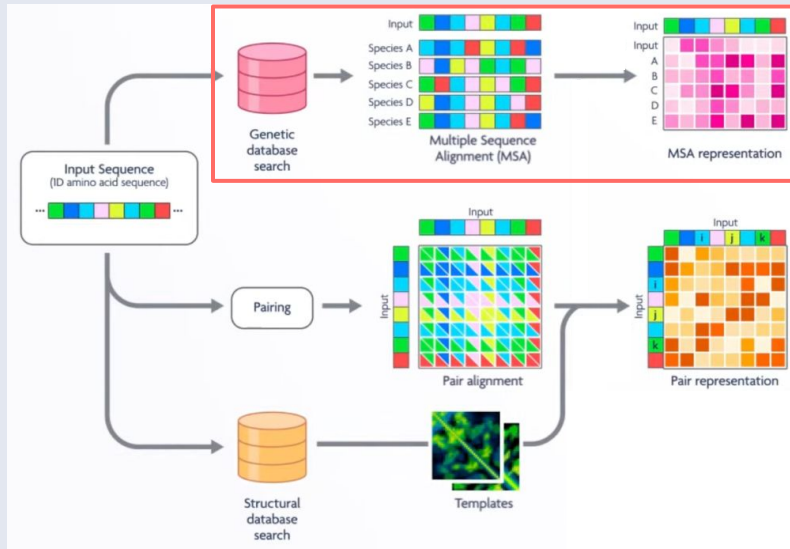


AlphaFold 2 Architecture



AlphaFold 2 takes in an amino acid sequence and outputs a protein structure.

MSA Representation



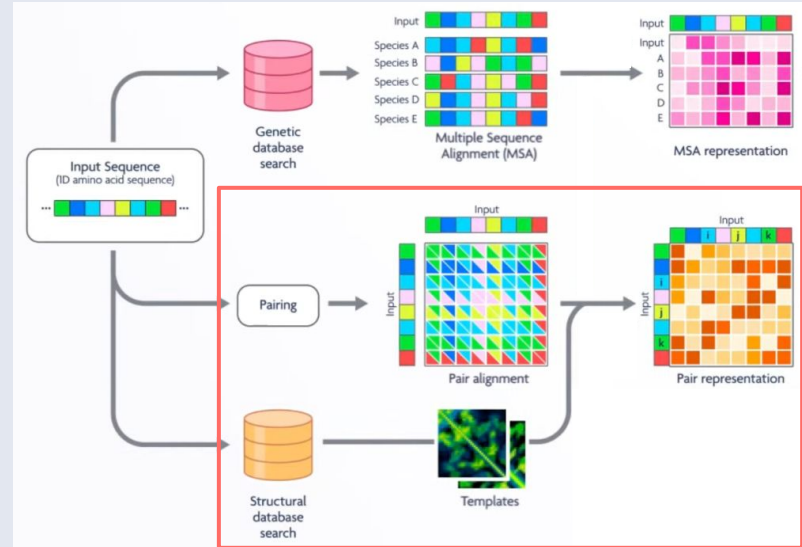
- Short for Multiple Sequence Alignment
- List of amino acid sequences of other proteins that are similar to the input sequence
 - Shows the evolutionary history of the protein
 - Probably evolved from the same source
 - Repeated/similar sections have similar structure and are integral to the function of the protein

- Proteins can mutate by insertion, substitution, or deletion
 - Can't compare positions of amino acids
- Machine learning algorithm tries to find best alignment by inserting a gap token represented by a dash
 - Uses the Needleman–Wunsch algorithm

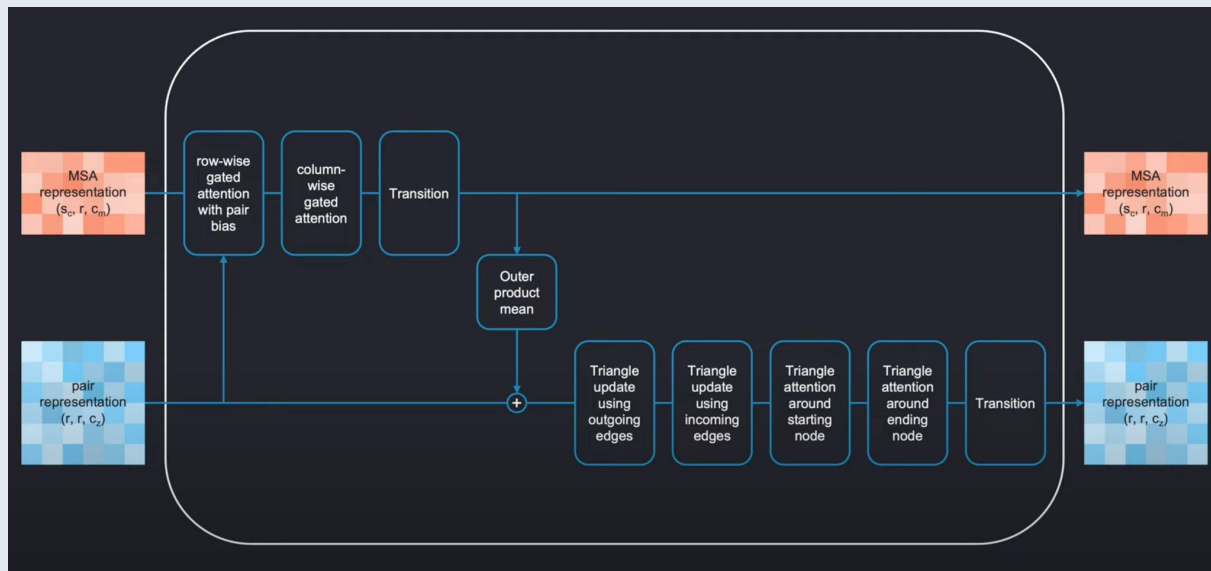
```
PIAQIHILEGRSDEQKETL - IREVSEAIRSLDAPLTSVRVILTEWMAKGHFGIGTELASK  
WIAQIDILETRSDEQLLHLPPEVSEAIRSGDAILVSKRVIQT - EMWFGHFGAGTYLASW
```

Pair Representation

- AlphaFold also pairs each amino acid with every other amino acid in the sequence
- It searches in many databases for the templates of proteins with similar sequences that have already determined structures
 - Other scientists have figured them out through experimental methods
- This is used to create an initial pair representation
 - Represents the relationship between every pair of amino acids



Evoformer



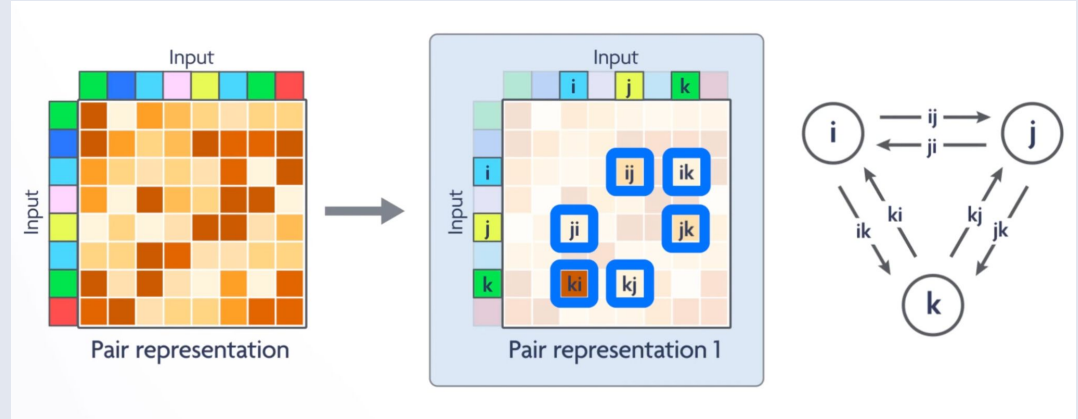
- A neural network that contributes the most parameters to AlphaFold
 - Full model has around 93 million and the Evoformer contributes 88 million
- Updates the MSA and Pair representations
- Has two main stacks that communicate with each other through two information channels

MSA Representation Stack

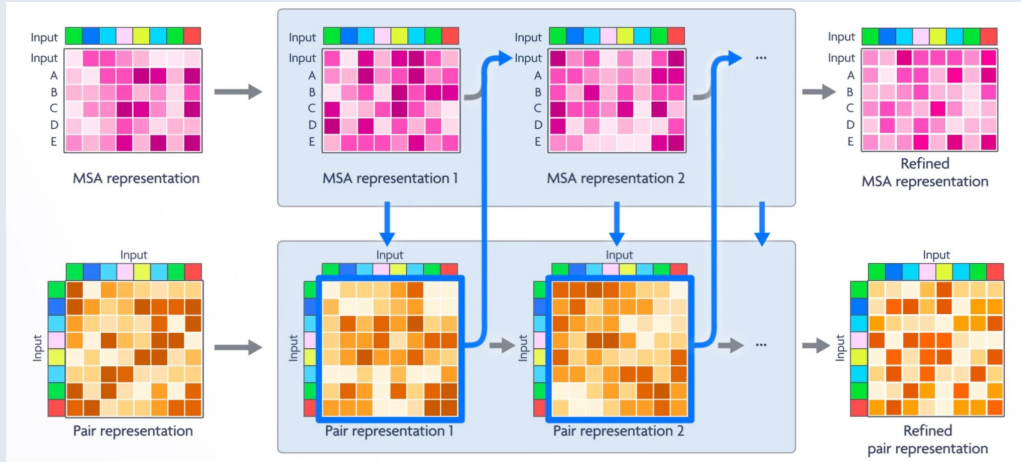
- Model wants to do attention on the MSA representation, but it would be too big
 - Instead splits into two parts
- Row-wise attention between the different amino acids in the input sequence
 - Is able to find the parts of the sequence that are important to the structure/function of the protein
- Column-wise attention
 - Evaluates the importance of each amino acid in the context of the other sequences from the other proteins

Pair Representation Stack

- The pair representation stack evaluates the relationship between each amino acid
- This can be thought of as nodes and the edges between them, like a graph
- It triangulates the relationship of each pair of nodes in relation to another node
- The goal of this is to satisfy the triangle inequality, which is where the sum of two edges on a triangle must be greater than the third edge



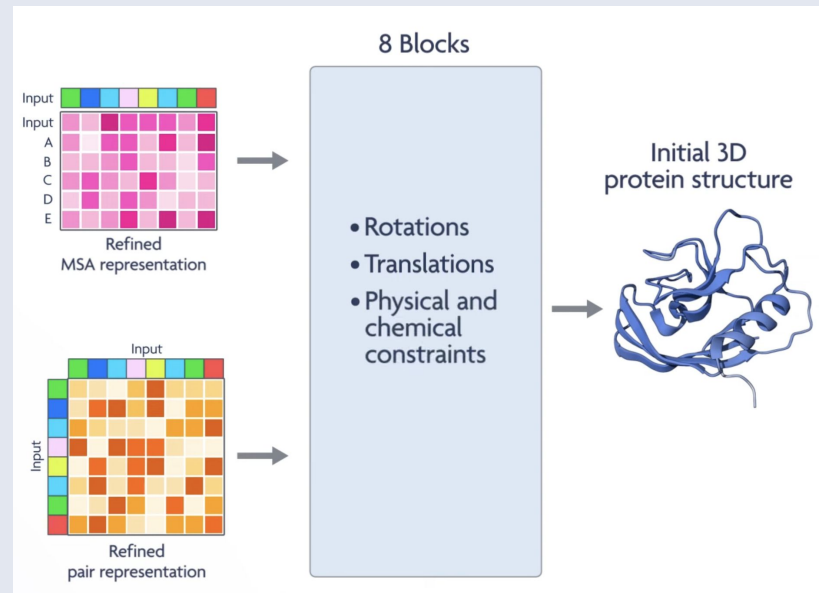
Communication Between Stacks



- Pair representation stack takes the information from the updated MSA representation and changes the pair representation accordingly
- Updated pair representation is used in the new MSA representation and so on for a total of 48 times
- The refined MSA and pair representations are created

Structure Module

- Refined MSA and pair representations are put into the Structure Module
- Performs rotations and translations on the amino acids
- Reveals an initial guess of the 3D protein structure
- It also applies physical and chemical constraints based on the atomic bond angles
- Iterated back through the Evoformer and Structure Module three more times to produce the final structure

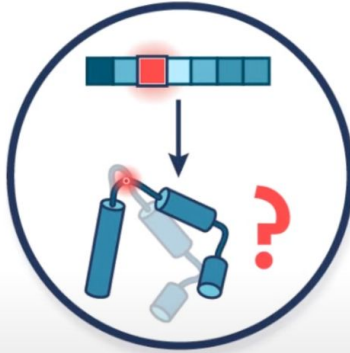


Applications of AlphaFold

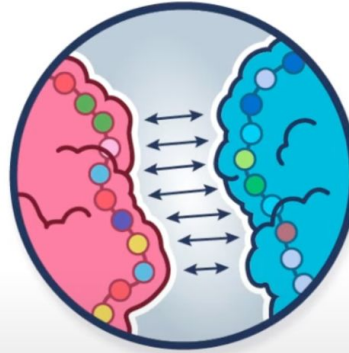
Discovering
drugs



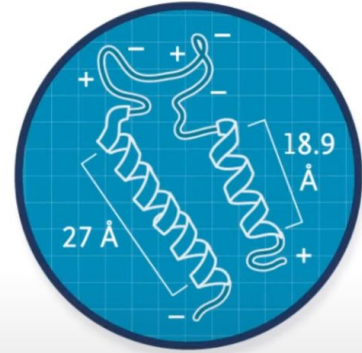
Effect of
genetic variants



Modeling
protein–protein
interactions



Engineering
artificial
proteins



Thanks

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

Please keep this slide for attribution