Predicting Protein Folding

“Improved protein structure prediction using potentials from deep learning”

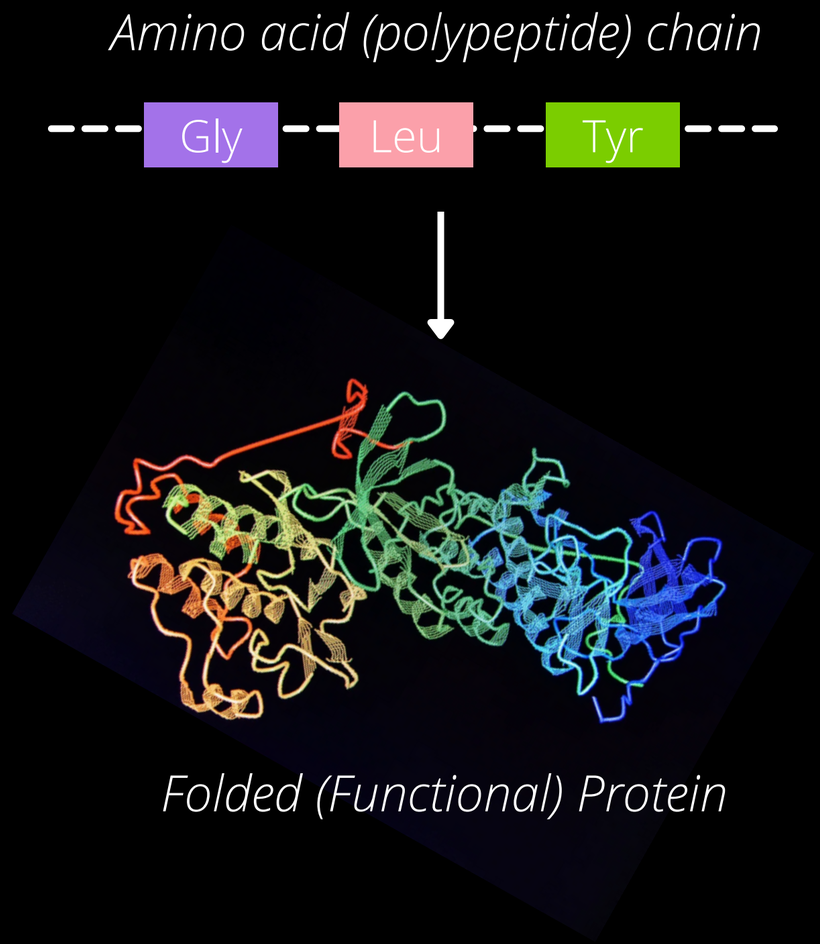
By Andrew W. Senior et al @ DeepMind

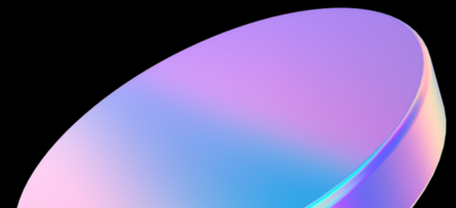
Nature, 2020

ANDRE YE | FEBRUARY 2021

How proteins fold is an important but complex problem.

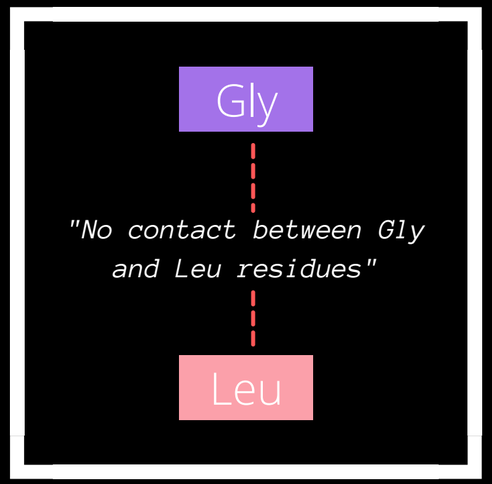
- Protein function is dependent on shape.
- If protein shape is known, it can be engineered to serve a function.
- Protein structures are highly complex and difficult to predict.



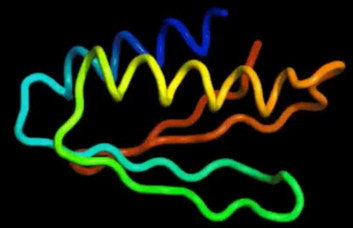


Recent protein modeling is a two-part process.

Polypeptide Chain



Predicted Folded Protein



An animation of the gradient descent method predicting a structure for CASP13 target T1008

1 Predict relationships between residues.

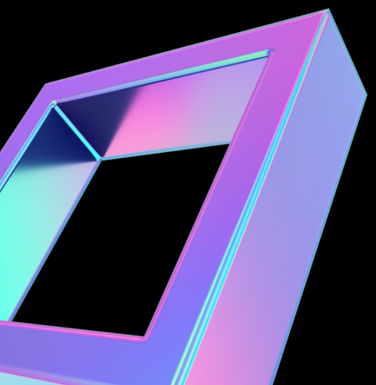
2 Find the structure that optimizes these relationships.
"Fragment Assembly"

Previous attempts relied on predicting residue contact.

- The model returns whether two residues should be in contact.
- This method is **not** information-rich.

AlphaFold relies on predicting the distance between residues.

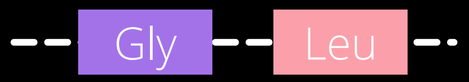
- The model returns the predicted distance between any two residues.
- This method is information-rich.



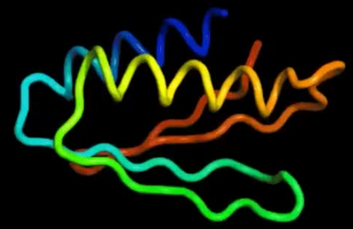
Putting it all together...

The AlphaFold System

Polypeptide Chain



Predicted Folded Protein



An animation of the gradient descent method predicting a structure for CASP13 target T1008

1 Predict the distance between residues.

2 Optimize the proposed structure given distances.

AlphaFold can model both template and free forms.

Proteins are split into domains, or significant parts of proteins.
Domains fold independently.

Template-Based Modelling (TBM)

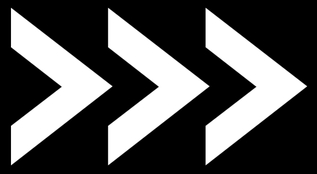
The structure of a domain with a similar sequence is known.

Make modifications to an existing template structure.

Free-Modelling (FM)

There is no homologous (similar) structure known.

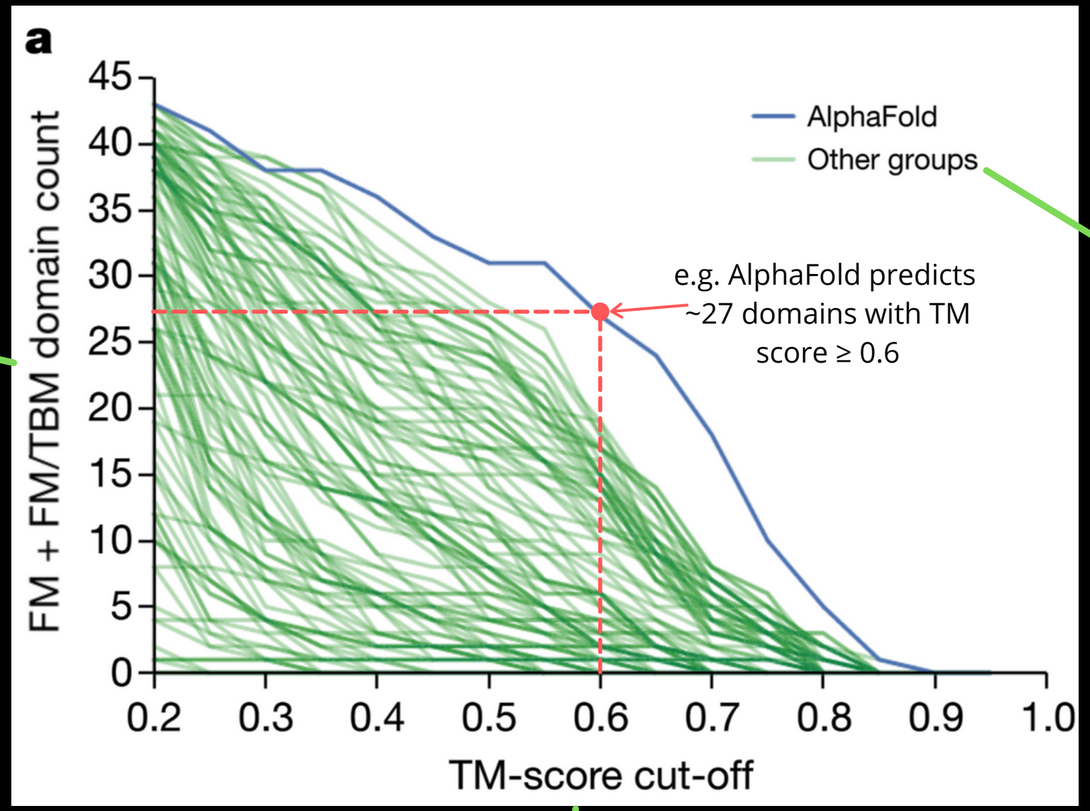
Generate a completely novel structure.



Result: AlphaFold is astoundingly accurate.

"FM + FM/TBM domain count":

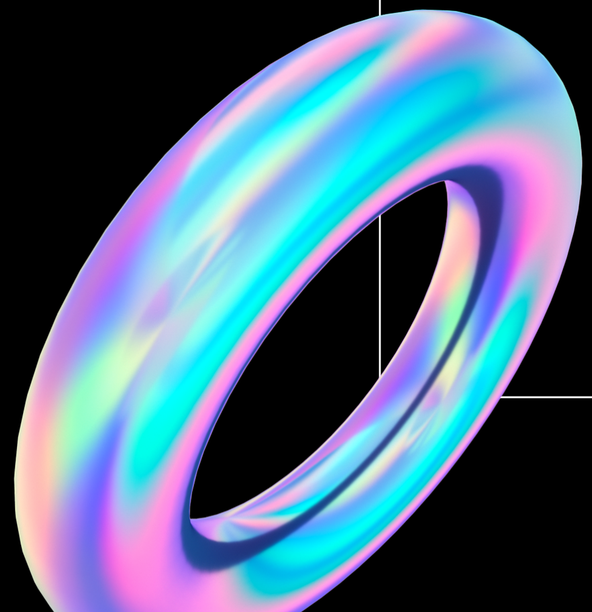
- Measures how many domains can be predicted with a *TM score ≥ the cutoff*.
- Considers both TBM and FM domains
- Bigger is better

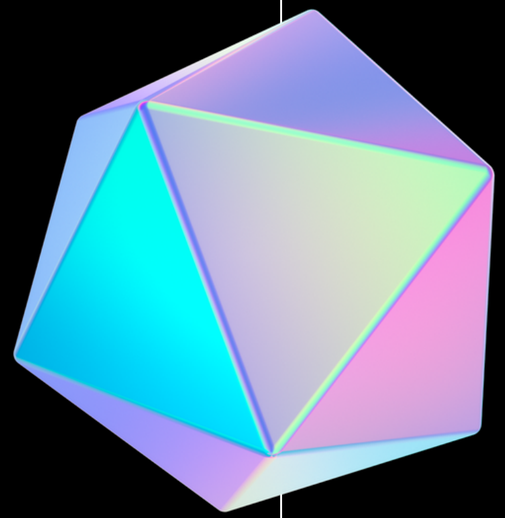


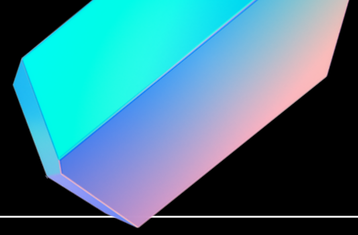
"Other groups": referring to other efforts or solutions to this problem

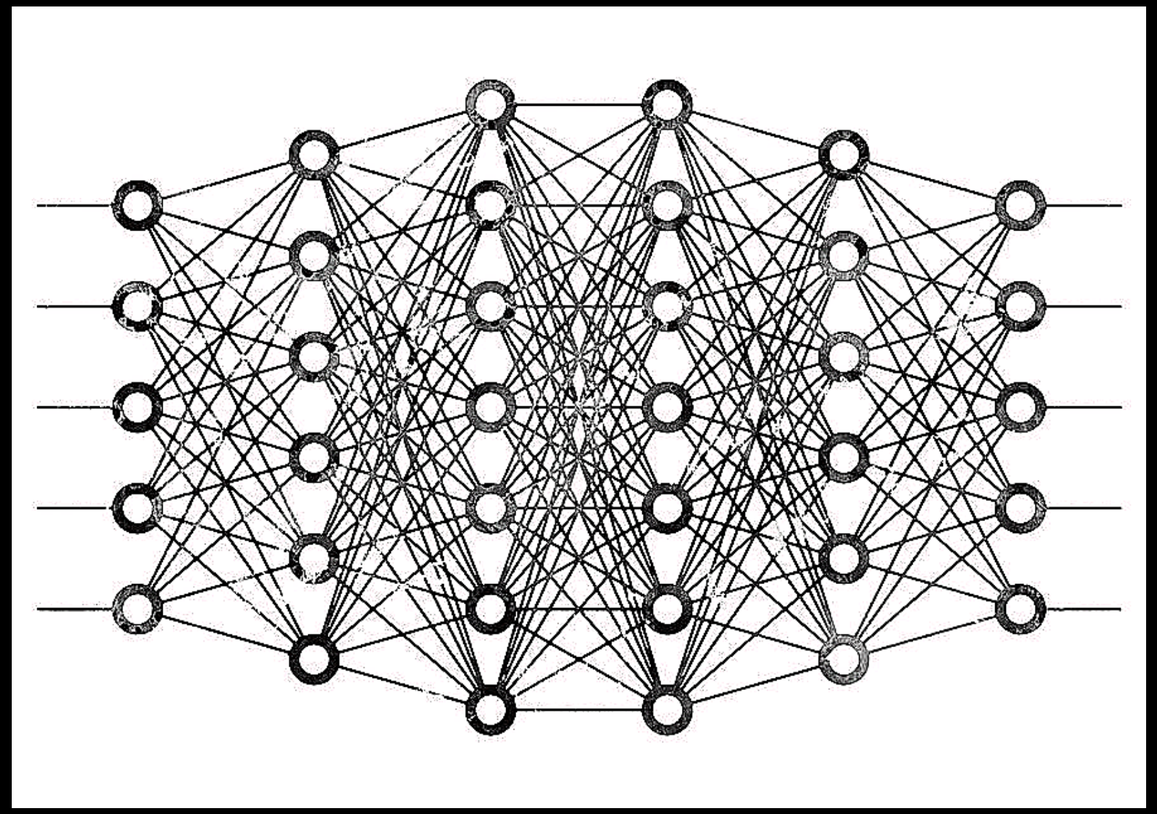
"TM-score": between 0 and 1, indicating how close the predicted structure is to the ground truth

Thanks for listening!



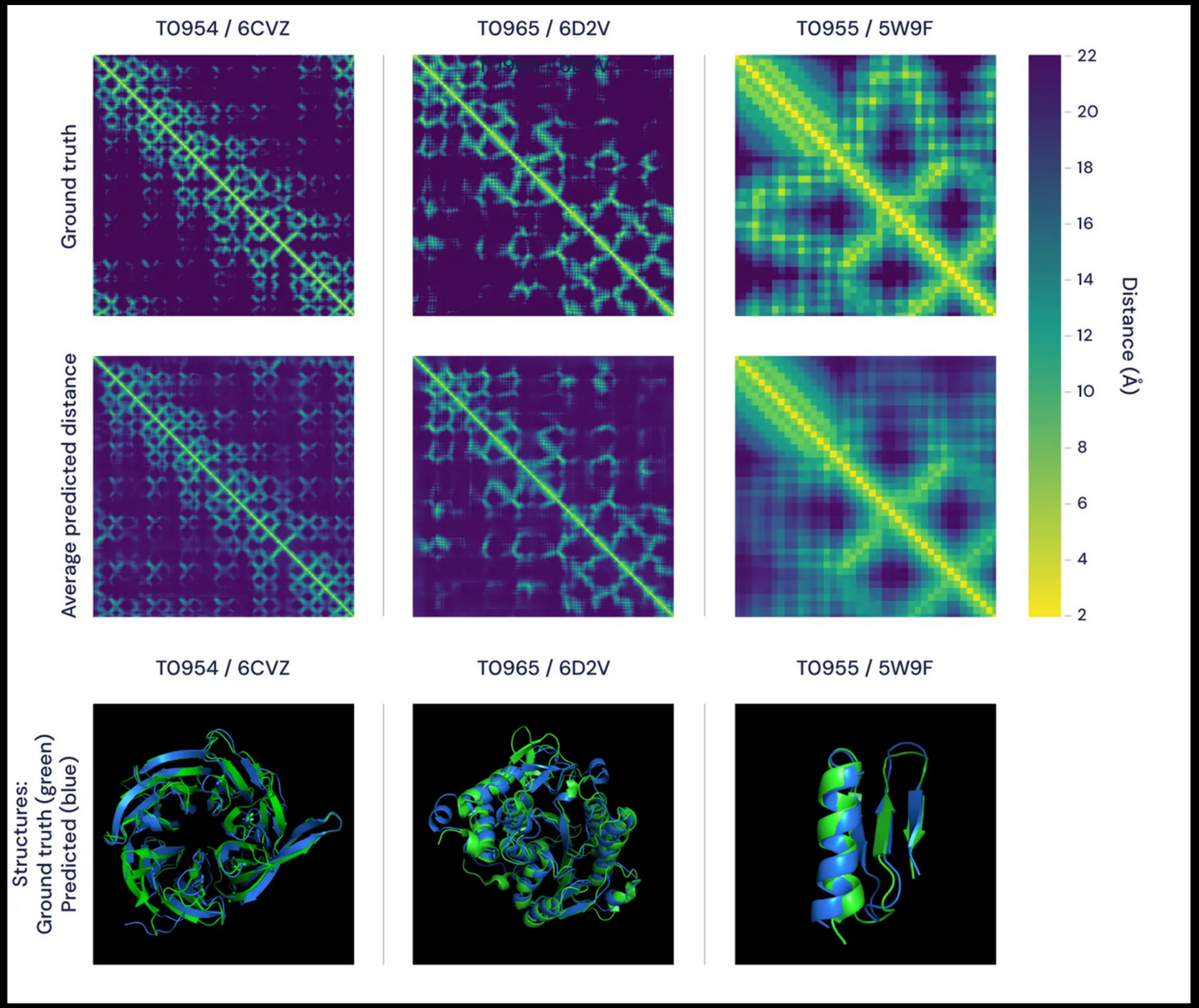


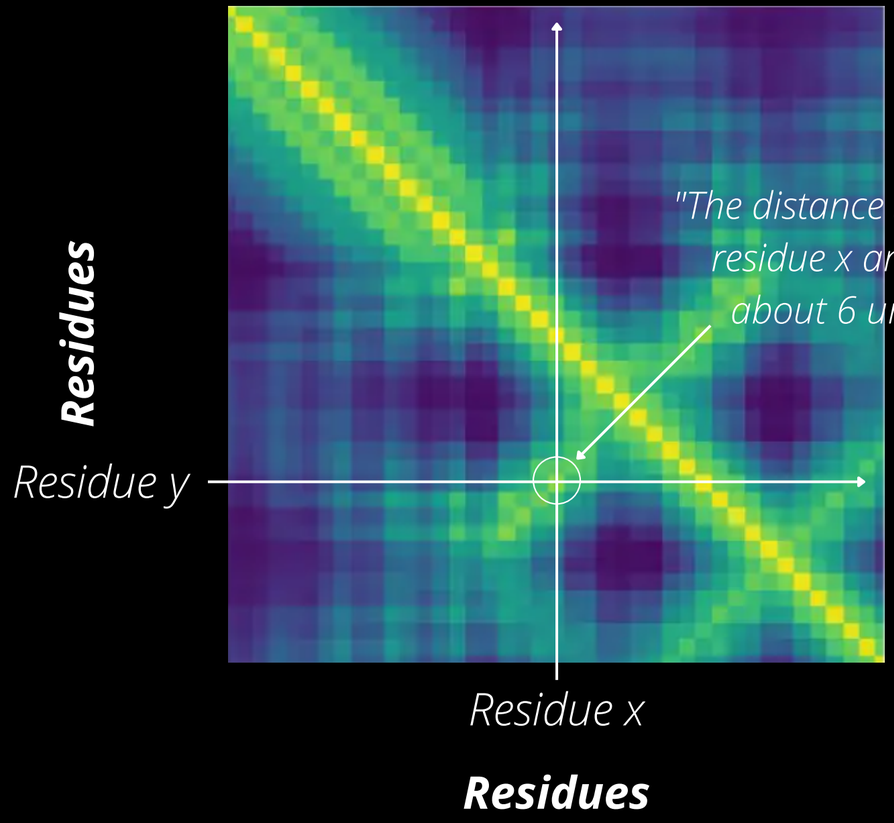




Deep neural networks learn the dynamics of chemistry to predict distances.

- Highly complex algorithms capable of learning highly complex phenomena.
- Trained on a dataset to predict distances between residues given the amino acid sequence.


T0954 / 6CVZ
T0965 / 6D2V
T0955 / 5W9F
Ground truth
Average predicted distance
22
20
18
16
14
12
10
8
6
4
2
Distance (Å)
T0954 / 6CVZ
T0965 / 6D2V
T0955 / 5W9F
Structures:
Ground truth (green)
Predicted (blue)


Residues
Residue y
"The distance between residue x and y is about 6 units."
Residue x
Residues