

Highly accurate protein structure prediction with AlphaFold (**AlphaFold2**) ---- Nature

Accurate prediction of protein structures and interactions using a three-track neural network (**RoseTTAFold**) ---- Science

outline

1. Preamble
2. AlphaFold2
3. RoseTTAFold

outline

1. Preamble

2. AlphaFold2

3. RoseTTAFold

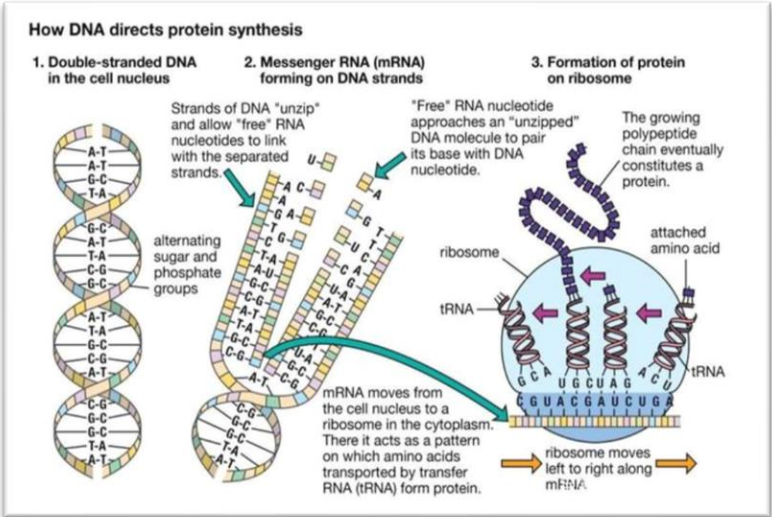
Protein Structure

- The Physical Interactions

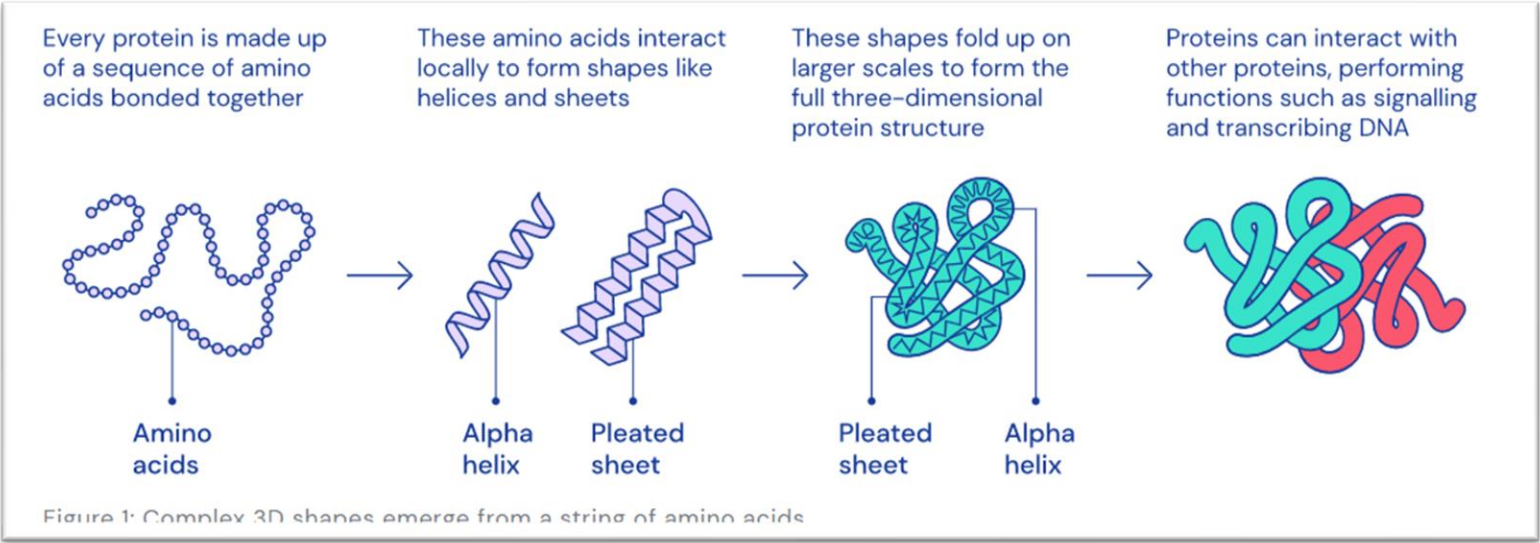
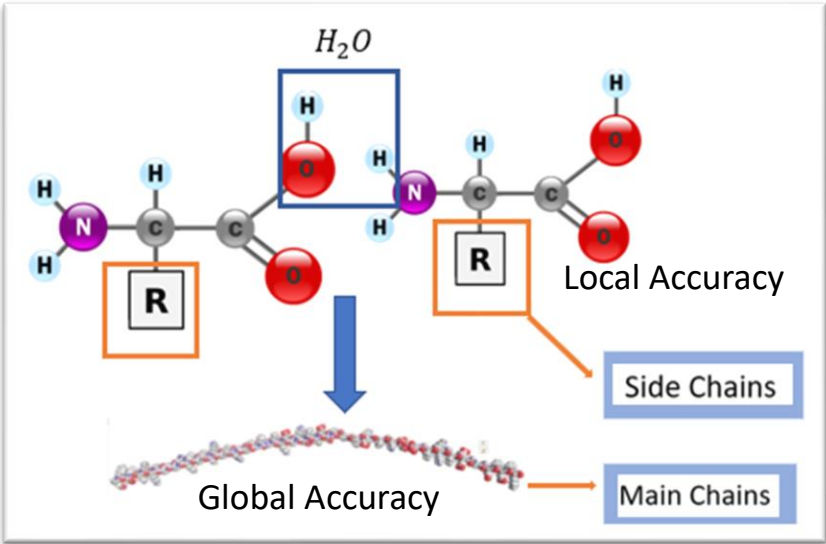
Interaction forces between thousands of atoms of amino acids. (Computational and modelling challenges)

- The Evolutionary History

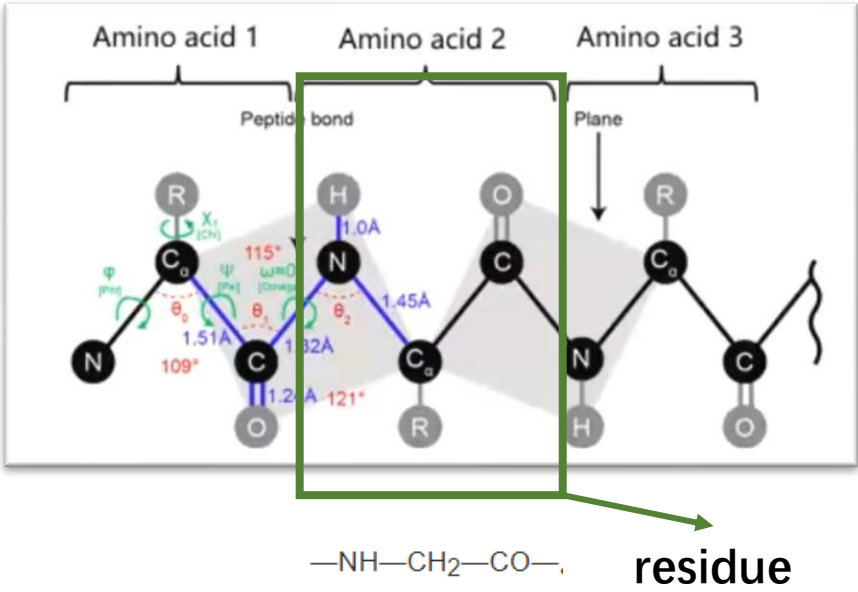
Many protein sequences in biology can be used.
Deep Learning



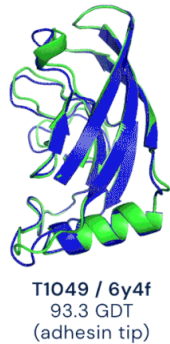
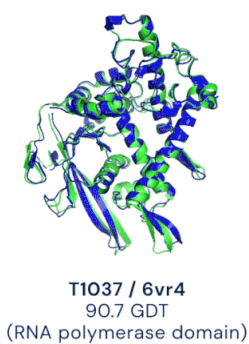
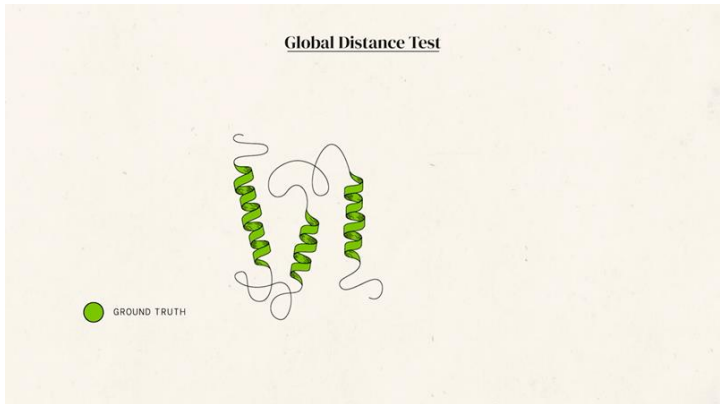
the Process of Protein Production



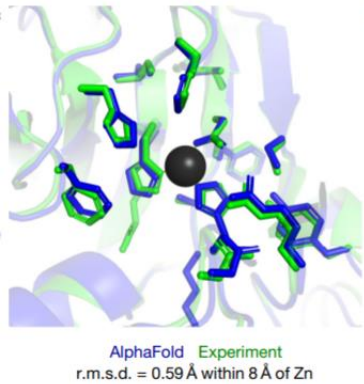
the Four Structures of Proteins



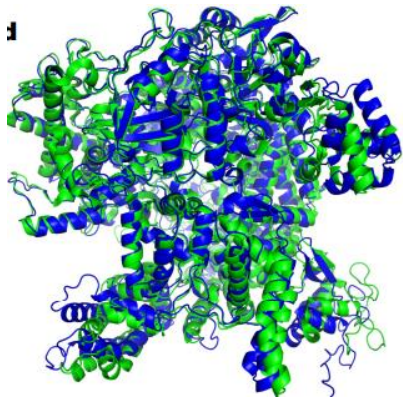
The Performance of Alphafolds and RoseTTAFold



● Experimental result
● Computational prediction



Side Chains

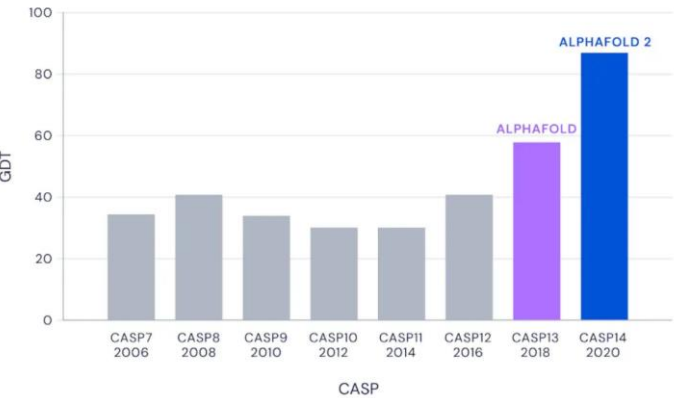


a 2,180-residue single chain

Accuracy: **GDT** (Global Distance Test), A GDT score of around 90 is considered to be competitive with human experimental methods.

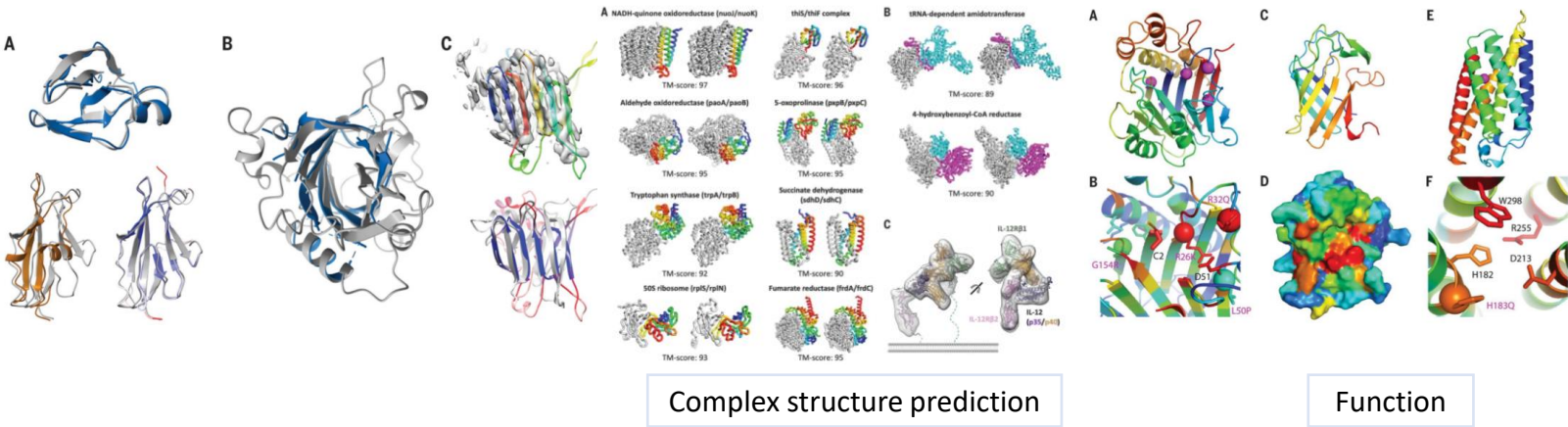
RMSE

Median Free-Modelling Accuracy



CSAP: The CASP assessment is a blind test

The Result of Alphafold2



The Result of RoseTTAFold

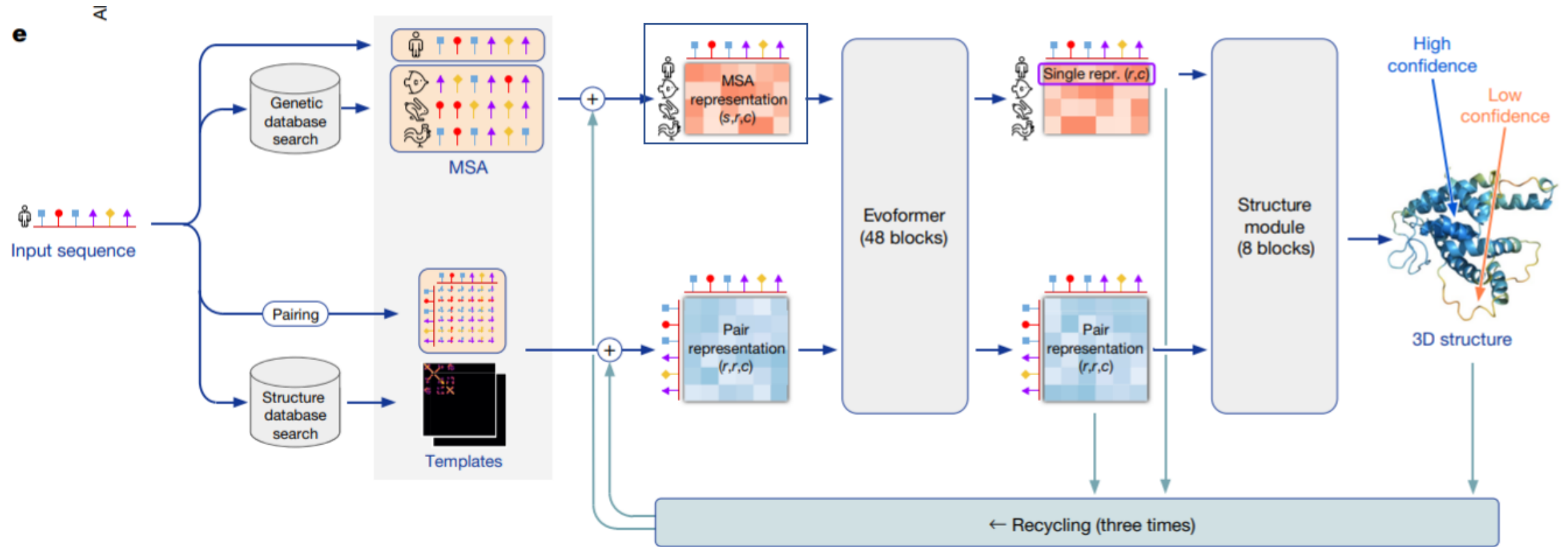
outline

1. Preamble

2. AlphaFold2

3. RoseTTAFold

Alphafolds (end-end)

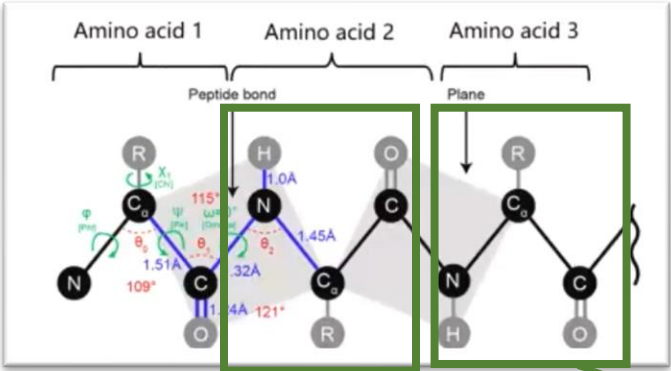


Model Architecture of Alphaflods

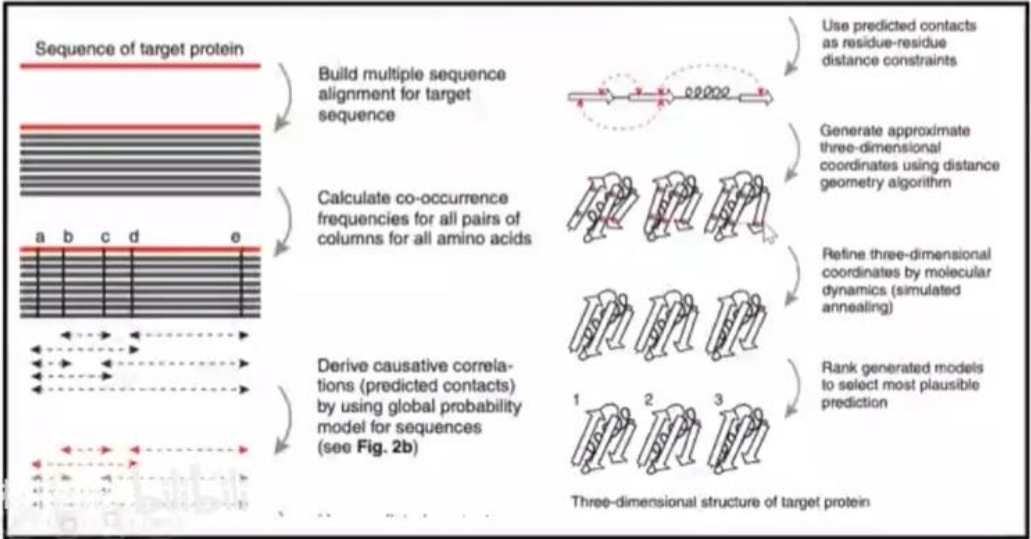
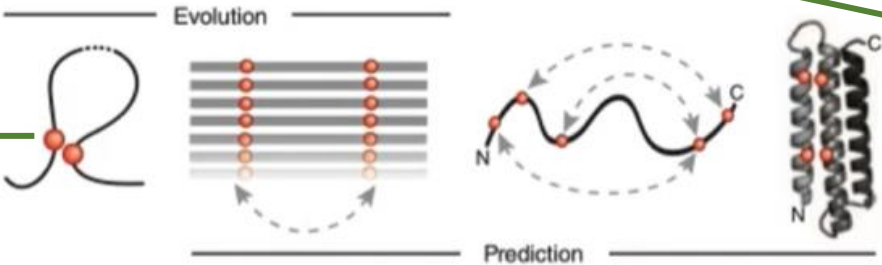
s , number of sequences
 r , number of residues
 c , number of channels

MSA

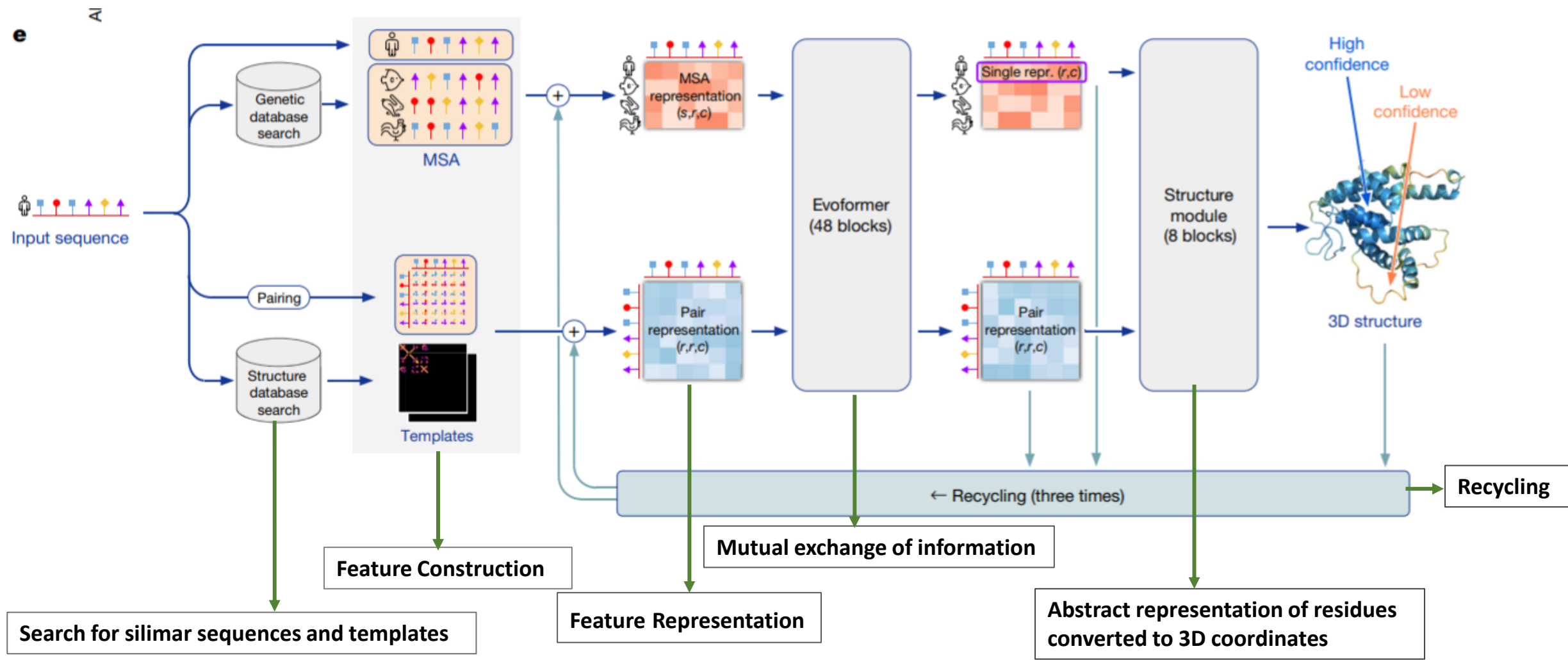
multiple sequence alignments



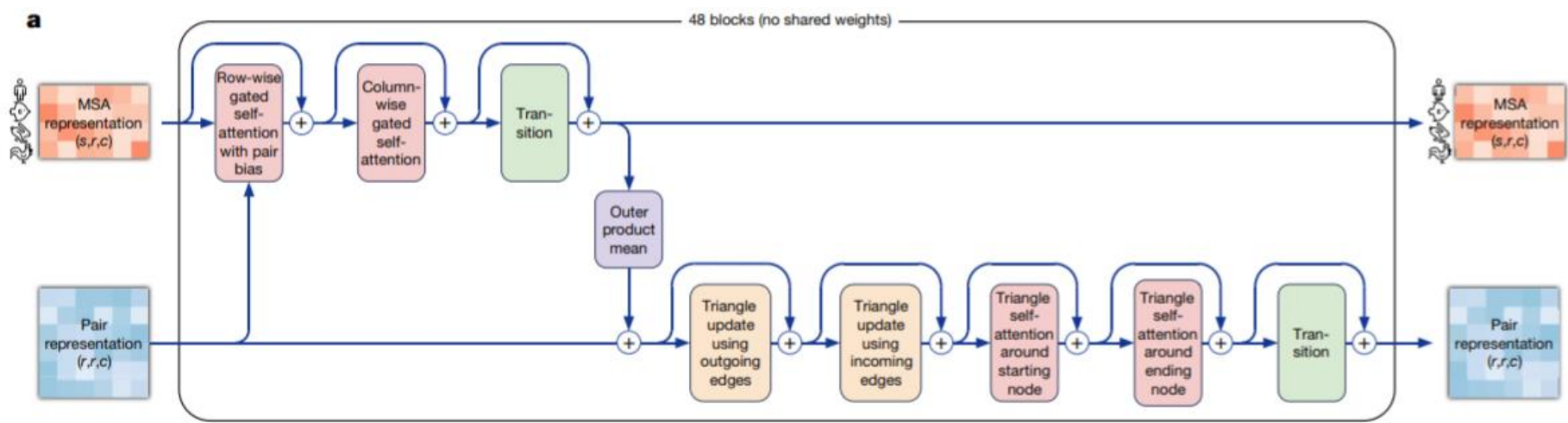
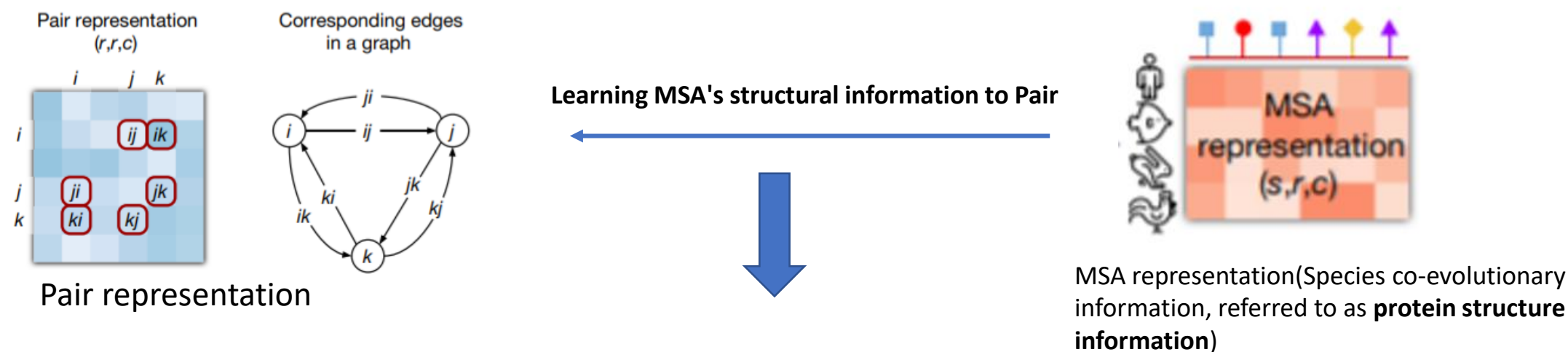
Residue Pair



Alphafolds

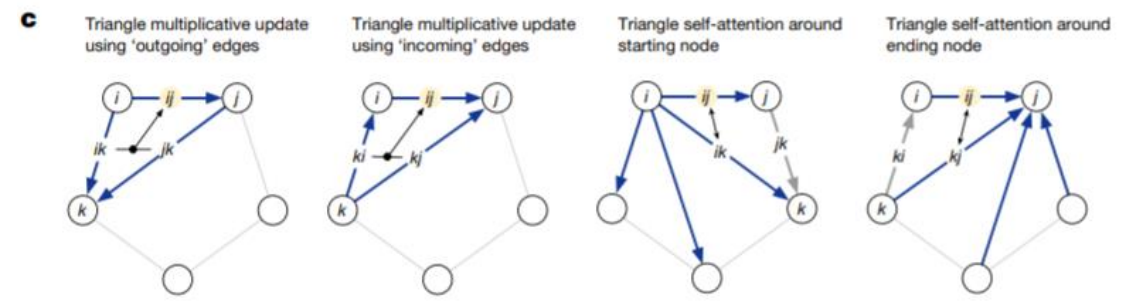
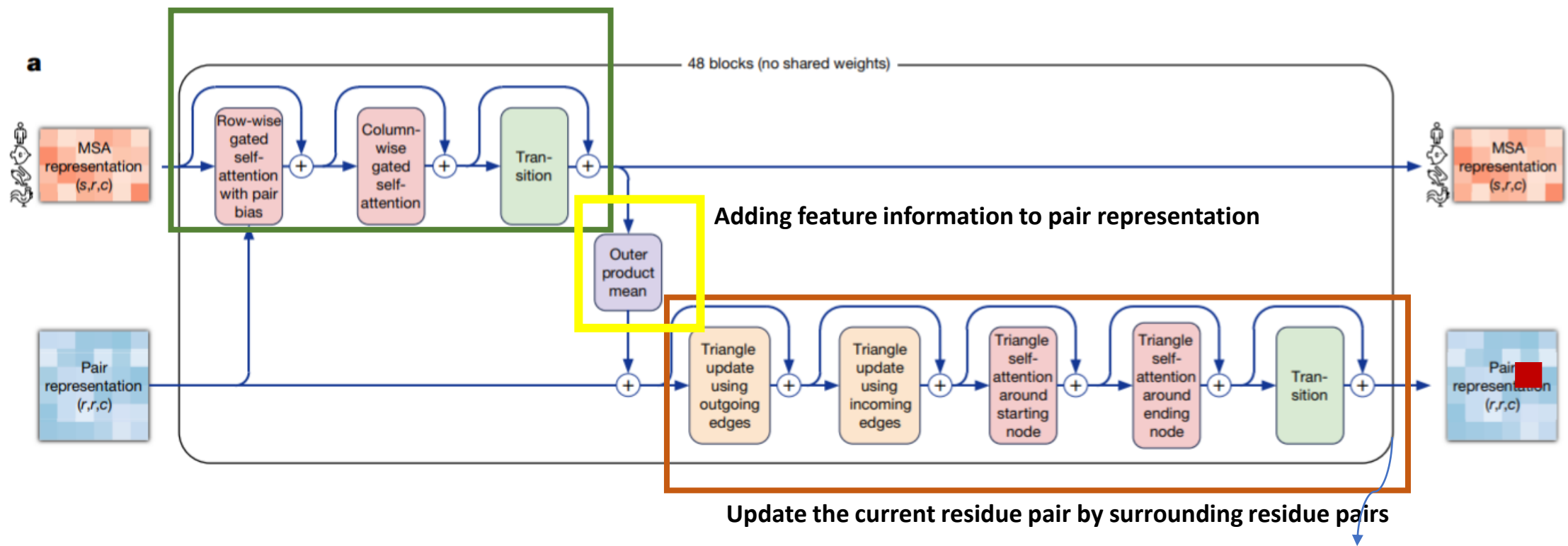


Alphafolds - Evoformer module

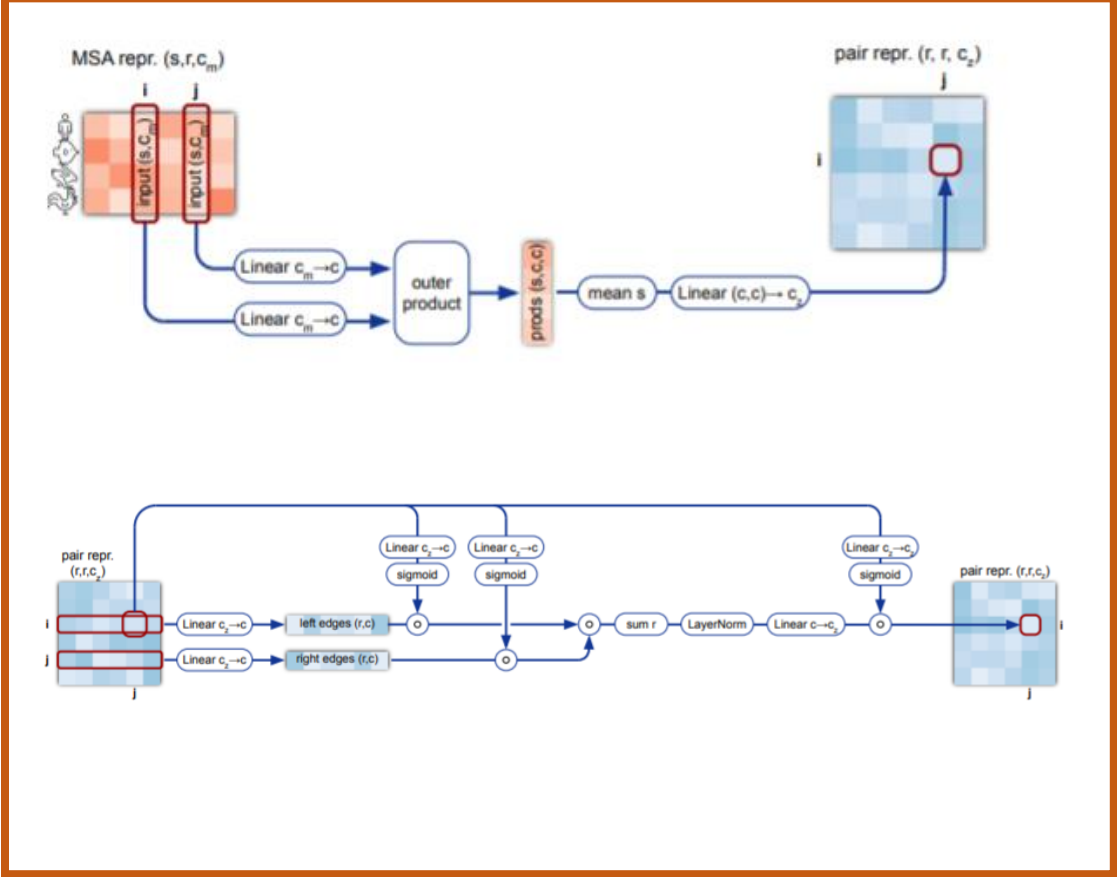
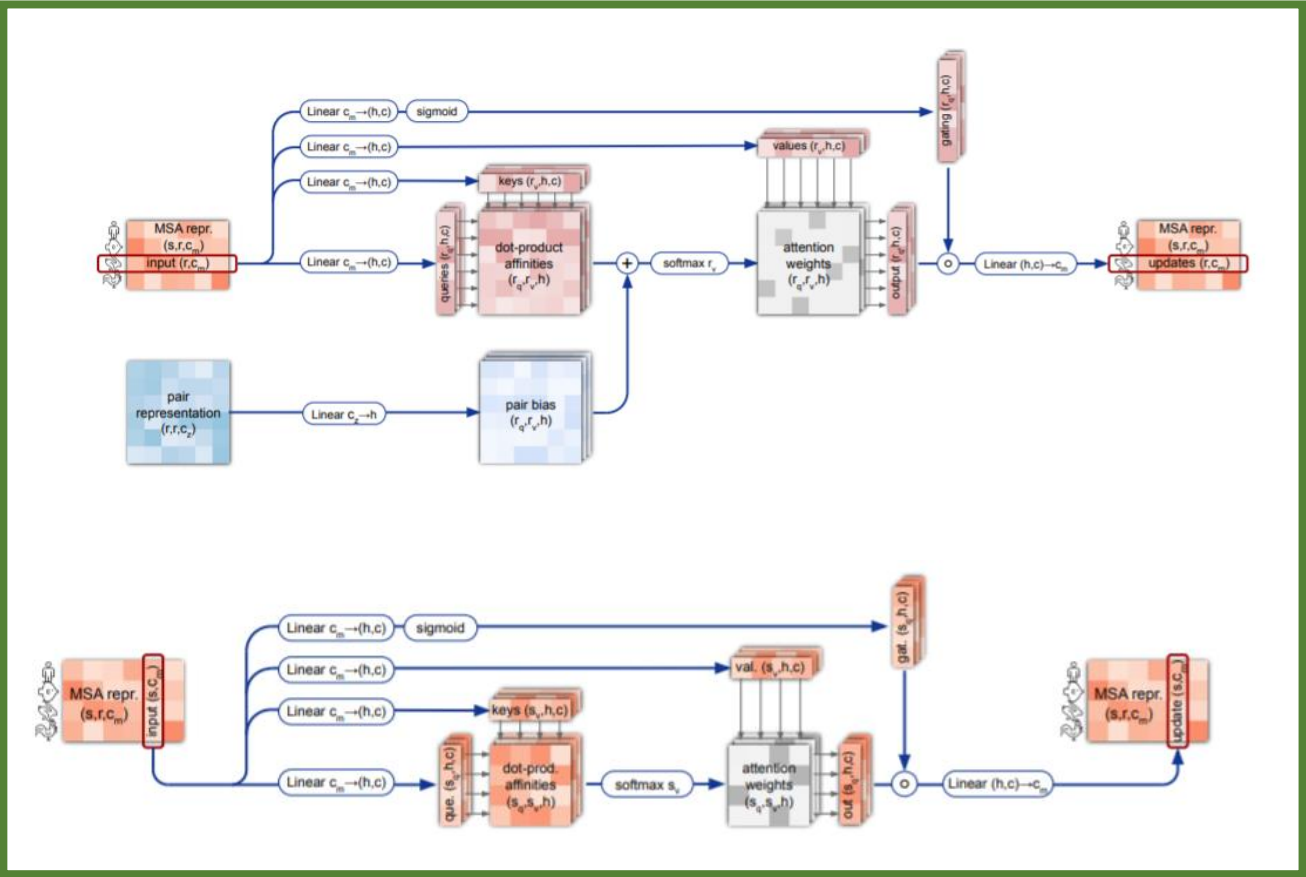


Alphafolds - Evoformer module

Extraction of protein feature from MSA



Alphafolds - Detailed Evoformer module



nature portfolio

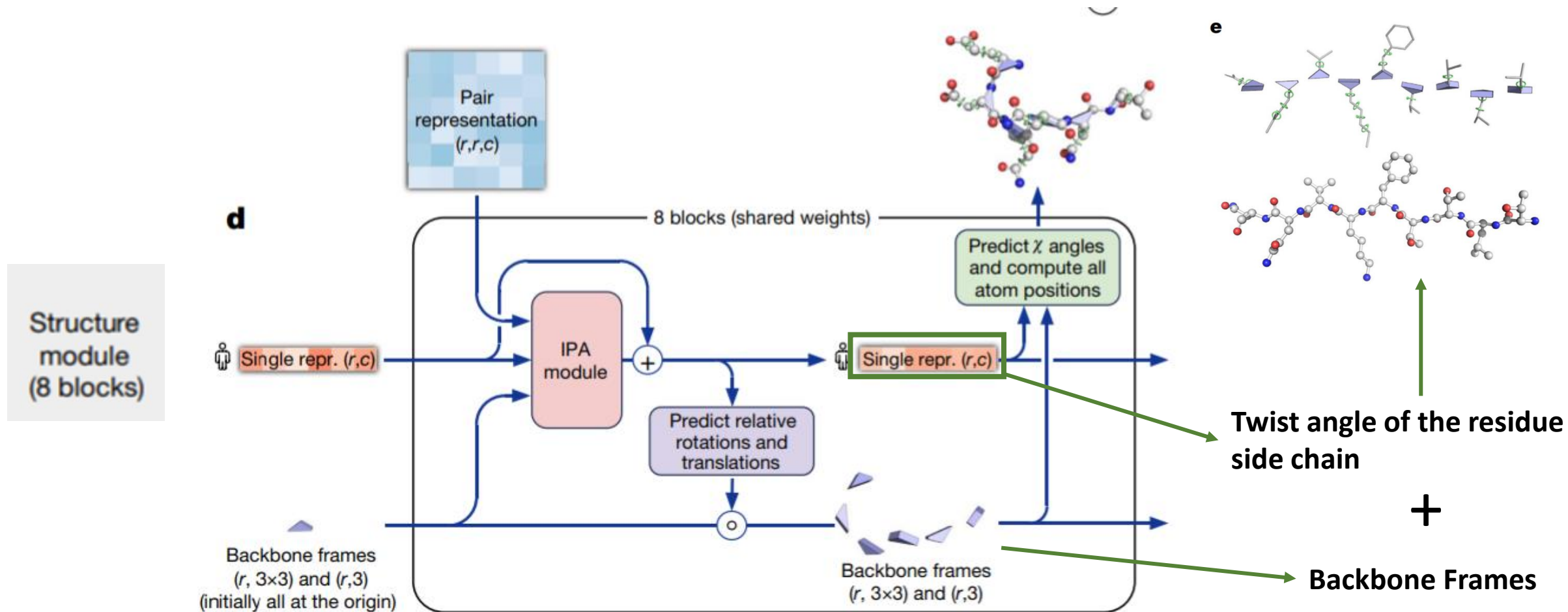
<https://doi.org/10.1038/s41586-021-03819-2>

Supplementary information

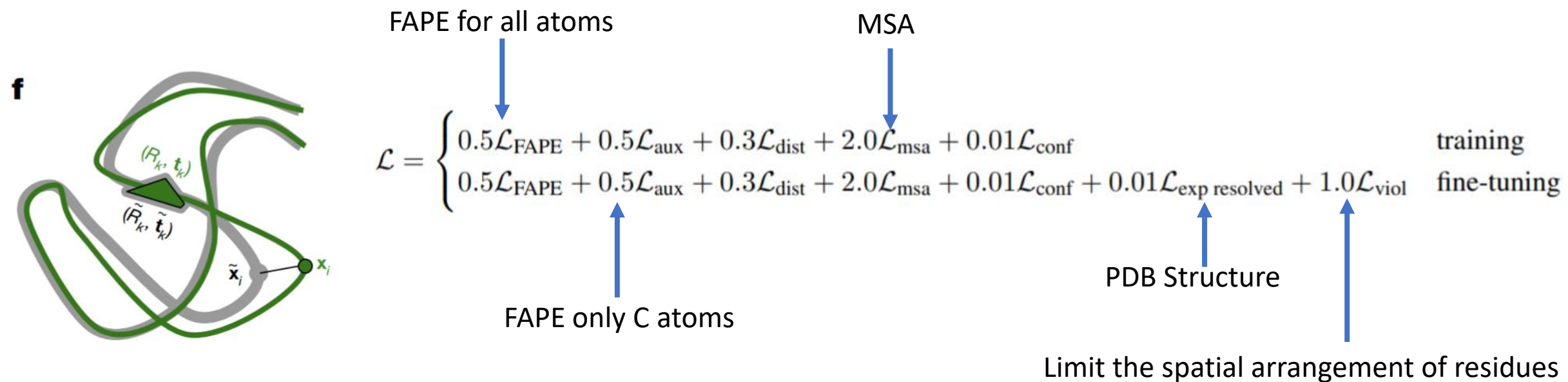
Highly accurate protein structure prediction with AlphaFold

In the format provided by the authors and unedited

Alphafolds – Structure modules



Alphafolds- Loss Function



Frame aligned point error (FAPE)

Green, predicted structure;
grey, true structure;
(Rk, tk), frames;
xi , atom positions.

Alphafolds- Models

Model	initial training	first fine-tuning		second fine-tuning				
	1	1.1	1.2	1.1.1	1.1.2	1.2.1	1.2.2	1.2.3
Parameters initialized from	Random	Model 1	...	Model 1.1	...	Model 1.2
Number of templates N_{templ}	4	4	0	4	...	0
Sequence crop size N_{res}	256	384
Number of sequences N_{seq}	128	512
Number of extra sequences $N_{\text{extra_seq}}$	1024	5120	1024	5120	...	1024
Initial learning rate	10^{-3}	$5 \cdot 10^{-4}$
Learning rate linear warm-up samples	128000	0
Structural violation loss weight	0.0	1.0
“Experimentally resolved” loss weight	0.0	0.01
Training samples ($\cdot 10^6$)	9.2	1.1	1.7	0.3	0.6	1.4	1.1	2.4
Training time	6d 6h	1d 10h	2d 3h	20h	1d 13h	4d 1h	3d	5d 12h

↓
Templates are used

↓
Templates are not used

AlphaFold2- Limitations

- 1. Prediction of multi-protein complexes**
- 2. Protein interactions with DNA, RNA and small molecules**
- 3. Precise position of amino acid side chains**

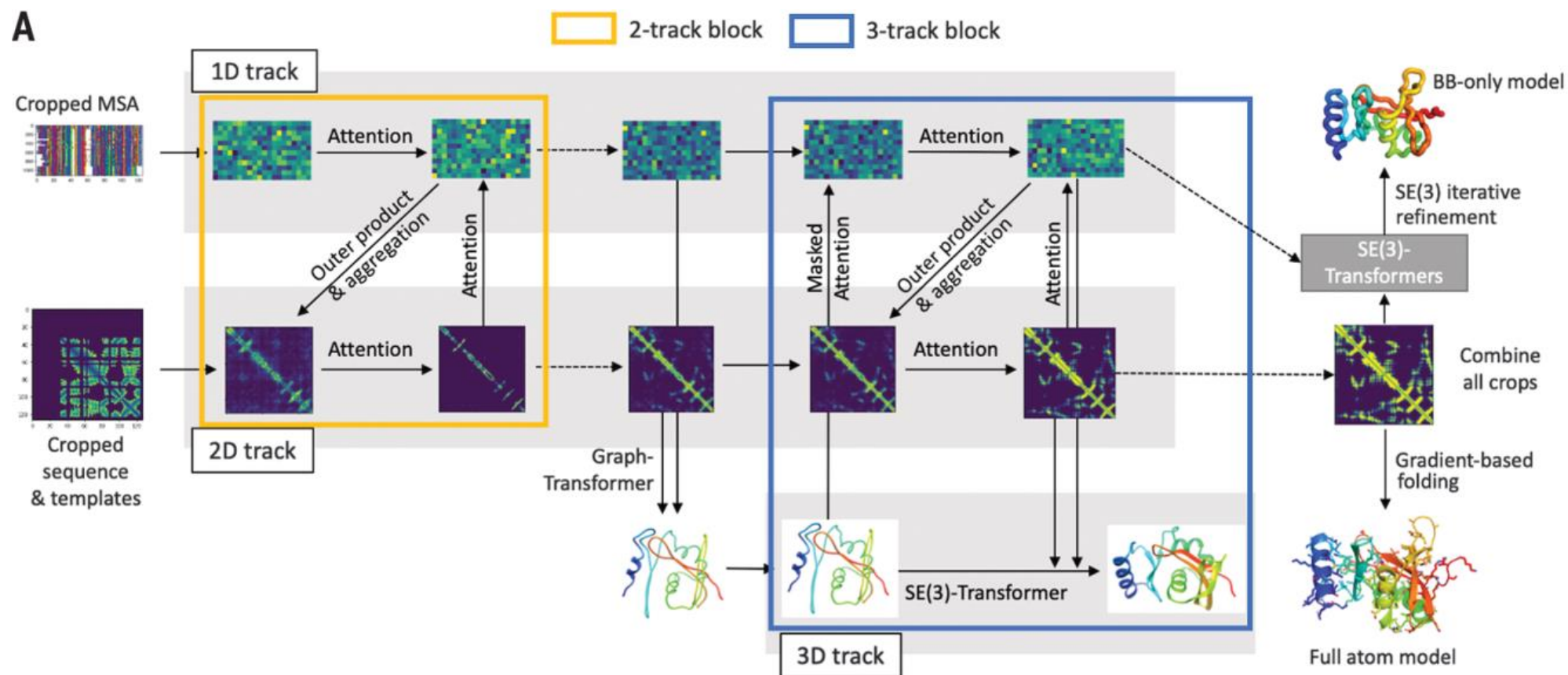
outline

1. Preamble

2. AlphaFold2

3. RoseTTAFold

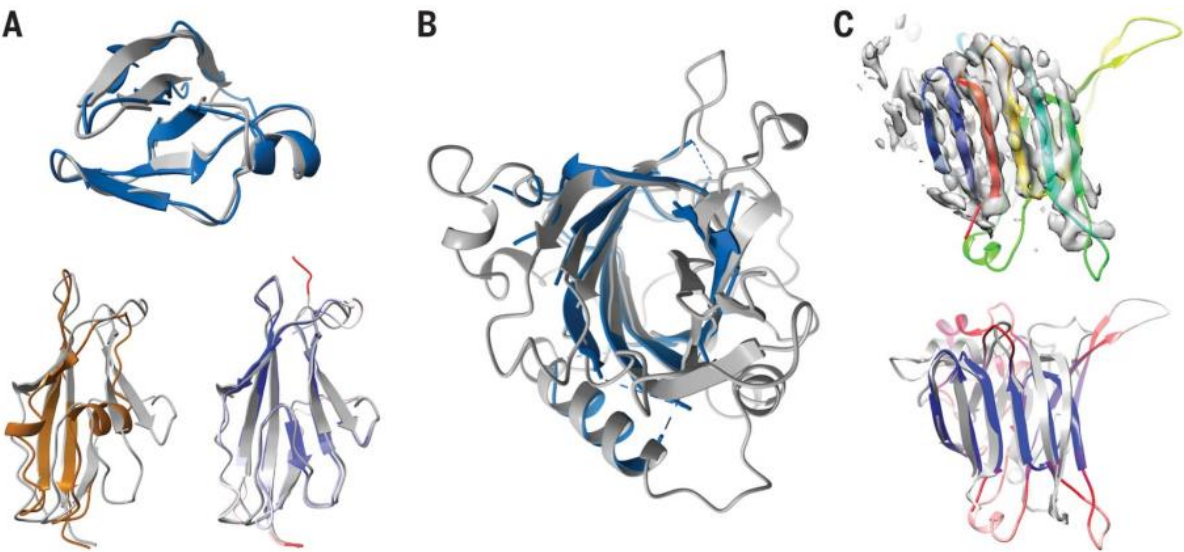
RoseTTAFold



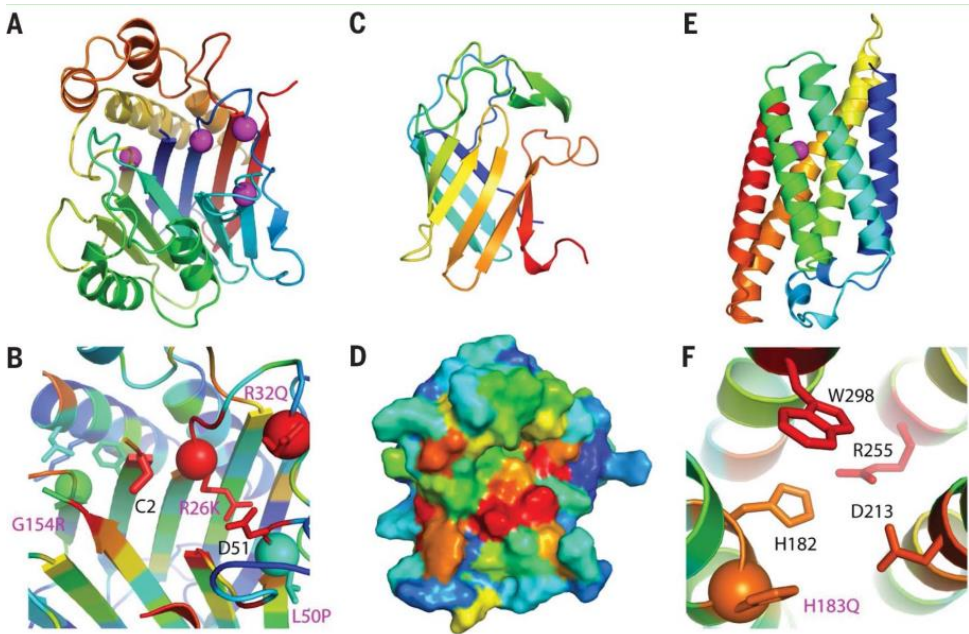
RoseTTAFold architecture with 1D, 2D, and 3D attention tracks.

Multiple connections between tracks allow the network to simultaneously learn relationships within and between **sequences, distances, and coordinates** (see methods and fig. S1 for details).

RoseTTAFold

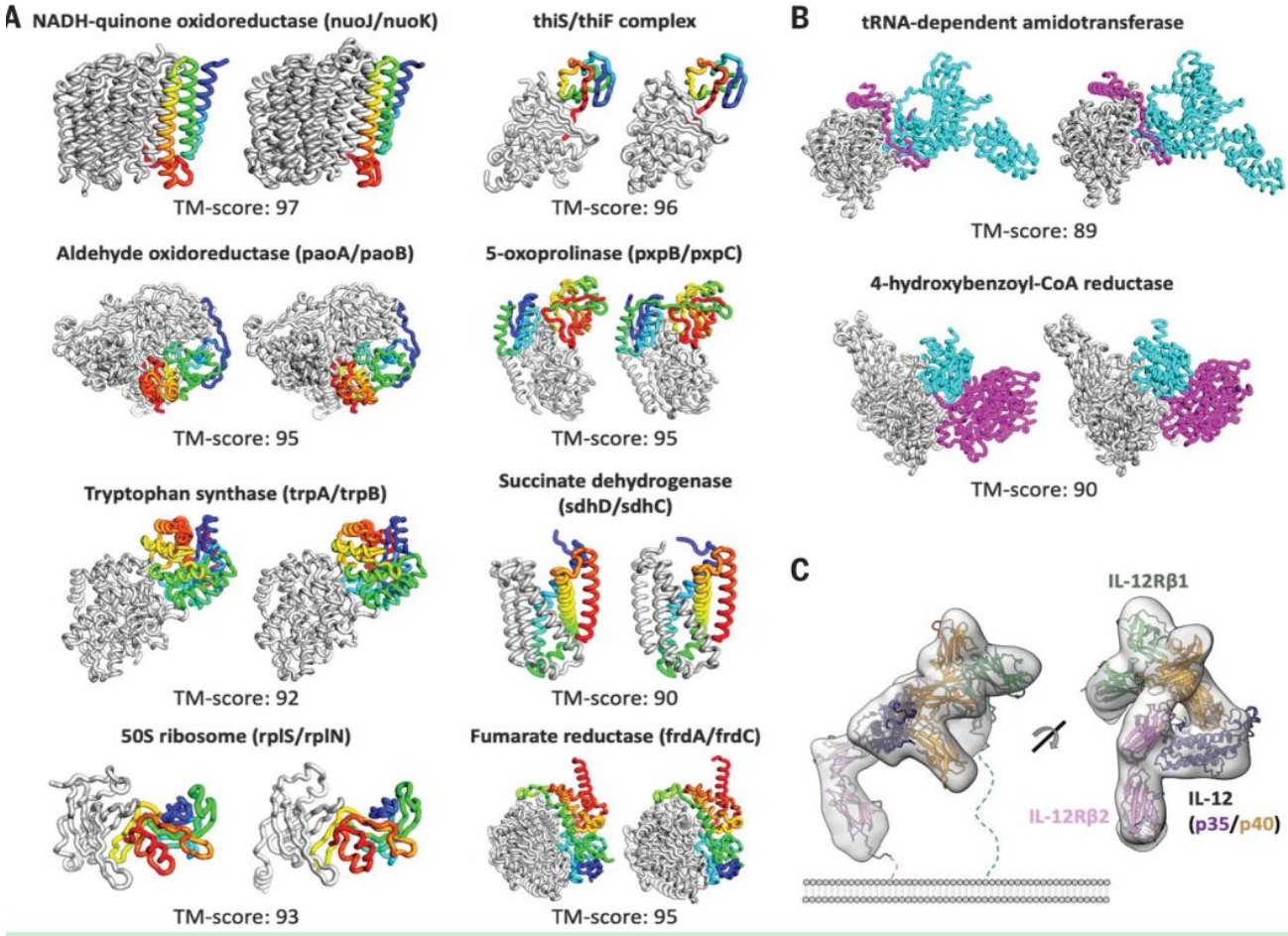


Enabling structure determination with RoseTTAFold



RoseTTAFold models provide insights into function

RoseTTAFold



Complex structure prediction using RoseTTAFold