UCLA Samueli Computer Science

ScAI Machine Learning Group

AlphaFold: AI Solution for Decade-long Protein Folding Challenge in Biology

Junheng Hao

University of California, Los Angeles (UCLA)

Tuesday, 10/19/2021

Junheng Hao

PhD Candidate, University of California Los Angeles (2017-)
Advisor: Wei Wang, Yizhou Sun
Website: Jeff's Home (haojunheng.com)

Bio

- 5th-year Ph.D. candidate at UCLA co-advised by Yizhou Sun and Wei Wang in UCLA Data Mining Group.
- My research interests include knowledge graph, graph representation learning, KG-empowered applications (NLP, Bioinformatics, recommender systems, etc.).

Past Experiences

- Research Intern, Microsoft Research Redmond, 2021
- PhD Research Intern, IBM, 2020
- Applied Science Intern, Amazon Product Graph, 2019
- Research Intern, NEC Labs America, 2018

Today's Agenda

- Background: Protein Structure Prediction
- AlphaFold v1: CNN
- AlphaFold v2: Transformer/Attention
- Science: Three-stack NN & SE(3)-Transformers
- Discussion: Network Science and Graph in Biology World

Papers

- AlphaFold: Improved protein structure prediction using potentials from deep learning (Published on Nature, Jan 2020)
- AlphaFold2: Highly accurate protein structure prediction with AlphaFold (Published on Nature, July 2021)
- (Optional Reading) Accurate prediction of protein structures and interactions using a three-track neural network (Published on Science, July 2021)

Background: What is protein folding and why is it important?

A decade-long biology challenge for proteins, the building blocks of life in the planet.

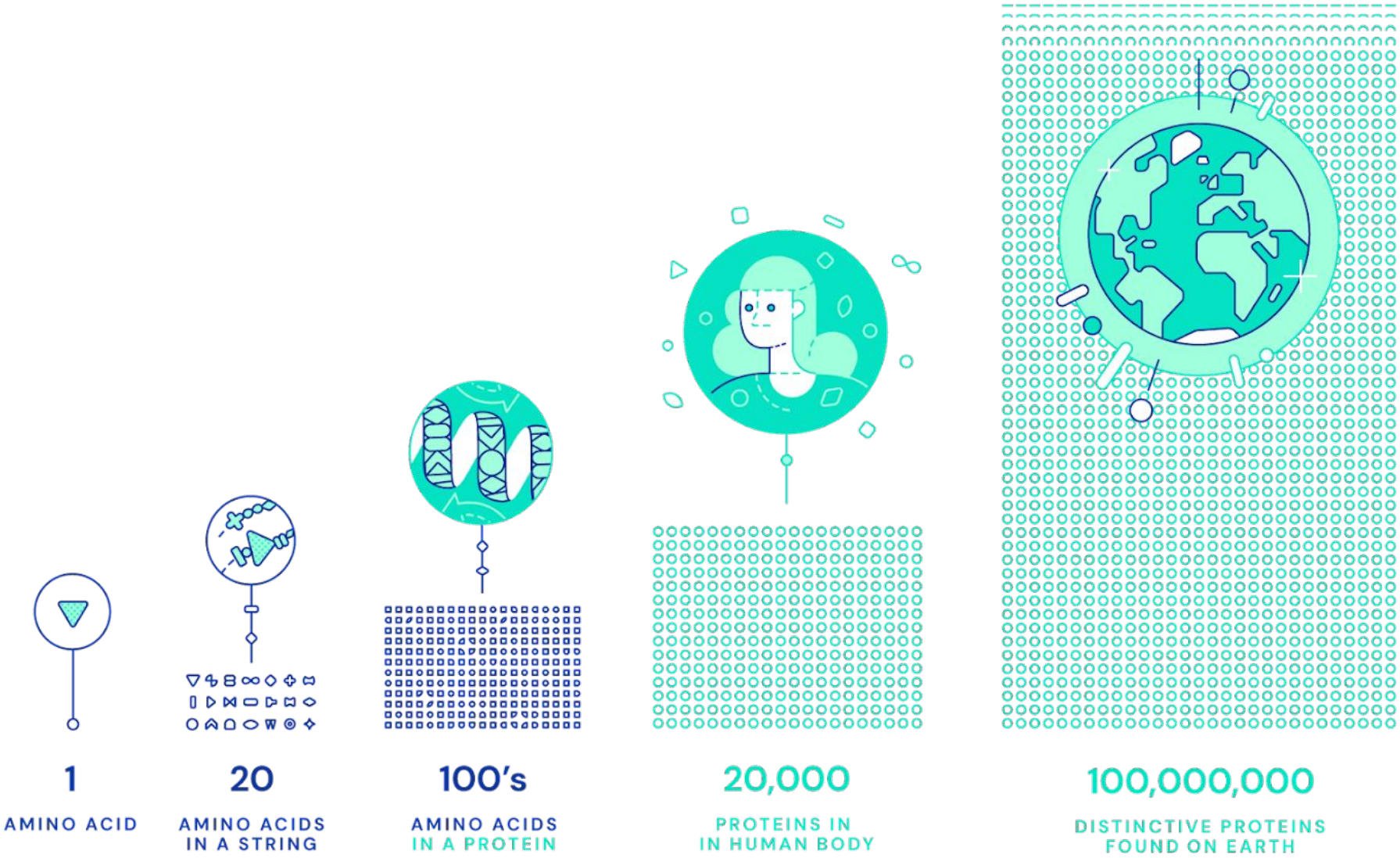
Biology 101: Proteins

UCLA
Engineer Change.

- Proteins are large, complex molecules essential to all of life. Nearly every function that our body performs (e.g. contracting muscles, sensing light, or turning food into energy), relies on proteins, and how they move and change.
- What any given protein can do (largely) depends on its unique 3D structure. Examples are:
  - Notorious "*spike proteins*" which stud coronavirus that allows the virus to enter our cells.
  - Antibody proteins utilized by our immune systems are *Y-shaped*, and form unique hooks.
  - Collagen proteins are shaped like cords, which transmit tension between cartilage, ligaments, bones, and skin.
- The recipes for those proteins, called genes, are encoded in our DNA and generated by Ribosome. Many diseases and deaths for an organism, are fundamentally linked to malformed proteins.
- Proteins are composed of **chains of amino acids** (also referred to as amino acid **residues**). But DNA only contains information about the sequence of amino acids, not how they fold into shape.

Biology 101: Proteins

UCLA
Engineer Change.



Why is Protein Folding Important?

UCLA
Engineer Change.

"I think that we shall be able to get a more thorough understanding of the nature of disease in general by investigating the molecules that make up the human body, including the abnormal molecules, and that this understanding will permit...the problem of disease to be attacked in a more straightforward manner such that new methods of therapy will be developed."

-- Linus Pauling, 1960

Why is Protein Folding Important?

UCLA
Engineer Change.

- Scientists have long been interested in determining the structures of proteins because a protein's form is thought to dictate its function.
- Once a protein's shape is understood, its role within the cell can be guessed at, and scientists can develop drugs that work with the protein's unique shape.
- Traditional methods: Experimental techniques like cryo-electron microscopy, nuclear magnetic resonance and X-ray crystallography
  - *A lot of trial and error, time consuming, high cost*
  - *Tens or hundreds of thousands of dollars per protein*
- Motivation: Biologists are turning to AI methods as an alternative to this long and laborious process for difficult proteins.
- The ability to predict a protein's shape computationally from its genetic code alone could no doubt help accelerate research.

X-ray crystallography

UCLA
Engineer Change.

- Huge cost: Hundreds of thousands of dollars and about one years in duration for one protein → Only 170,000 protein folding structures have been identified

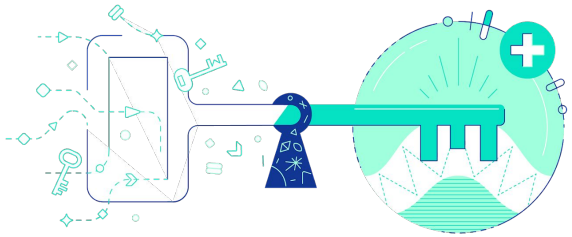
Protein Folding: Take-away

UCLA
Engineer Change.

Uniqueness: The sequence usually map 1-to-1 to a 3D structure.

Problem: Huge number ways and possibilities to fold.

Cost: X-ray crystallography costs $120,000 and takes 1 year.



Function: 3D structure determines its function. Misfold → disease

Dataset: **200M** proteins with sequences but only **170K** with available 3D structures.

Protein Folding: Promising Applications

UCLA
Engineer Change.

Near-term

DNA → Function: Learn unknown function of genes encoded in DNA

Disease: Understand the cause of disease as results of misfolded proteins.

Treatment: Design proteins to fix other misfolded proteins.

Other applications: Agriculture, Supplements and biomaterials.

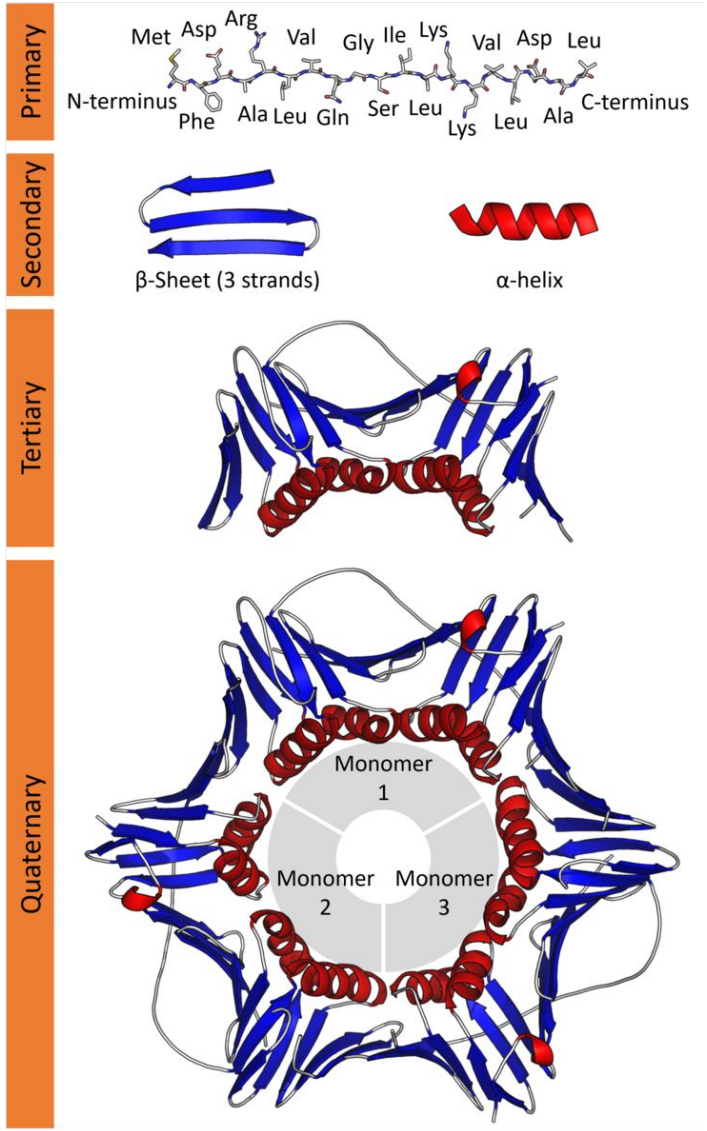
Long-term

Physics-based stimulation of biological systems

Biological and artificial life

Four Levels of Protein Structures

UCLA
Engineer Change.

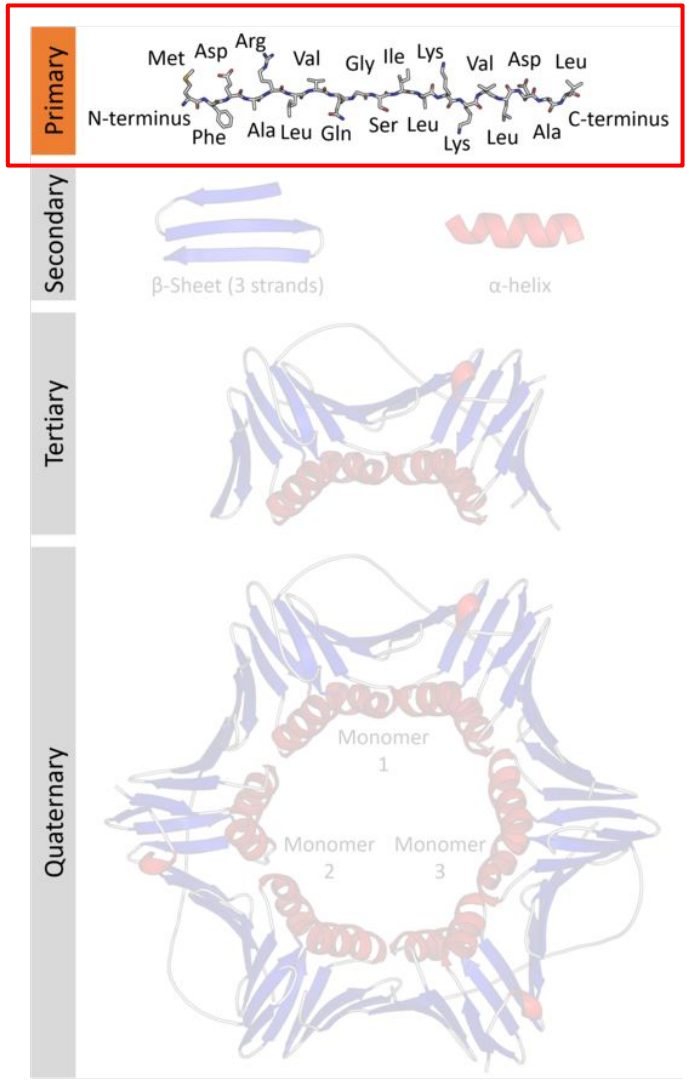


← **Level 1: What we mostly (and easily) know about!**

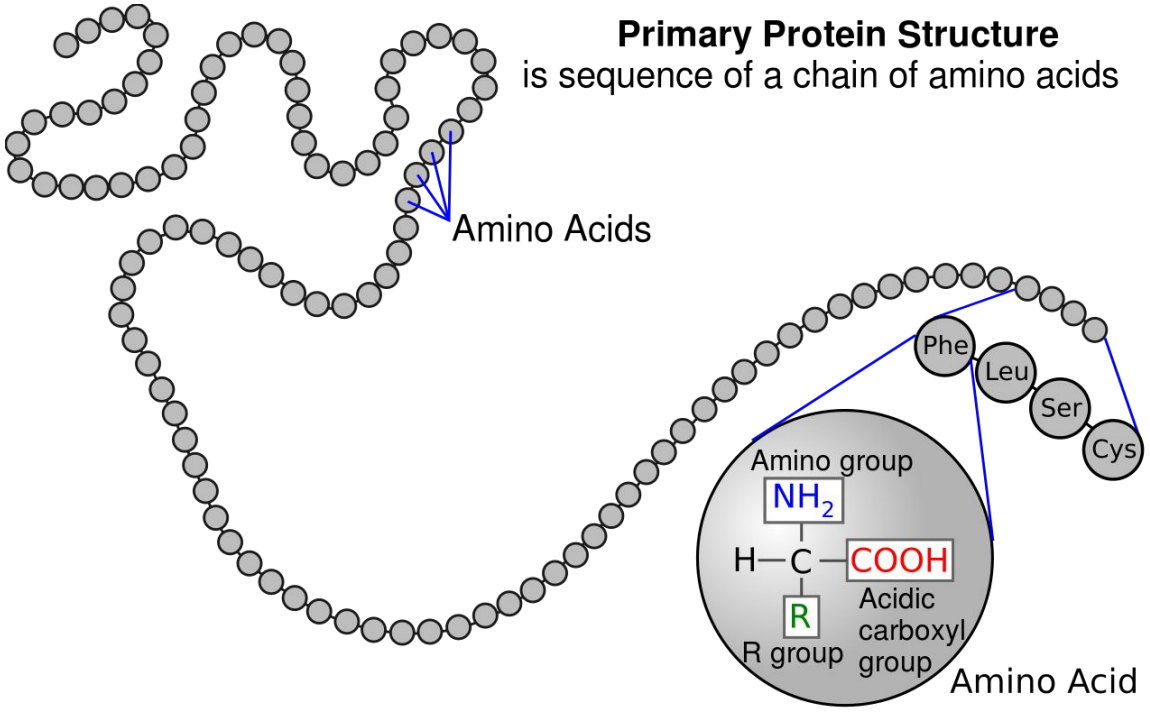
← **Level 3: What we mostly care about! The Folding!**

Credit:
[1]https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/a/orders-of-protein-structure
[2]https://en.wikipedia.org/wiki/Protein_structure

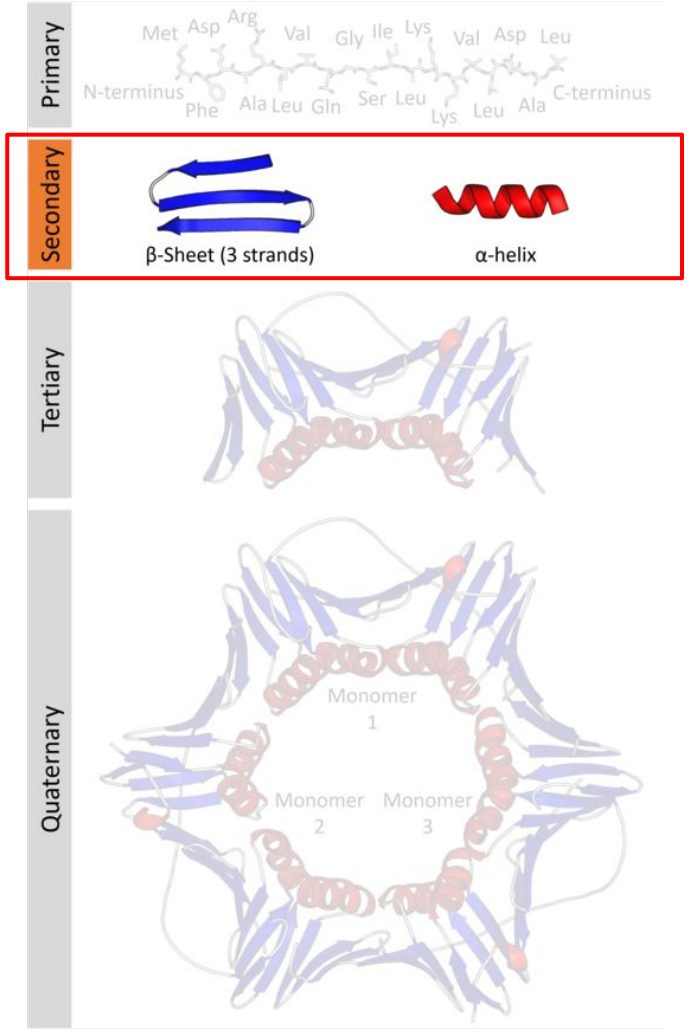
Four Levels of Protein Structures: Primary



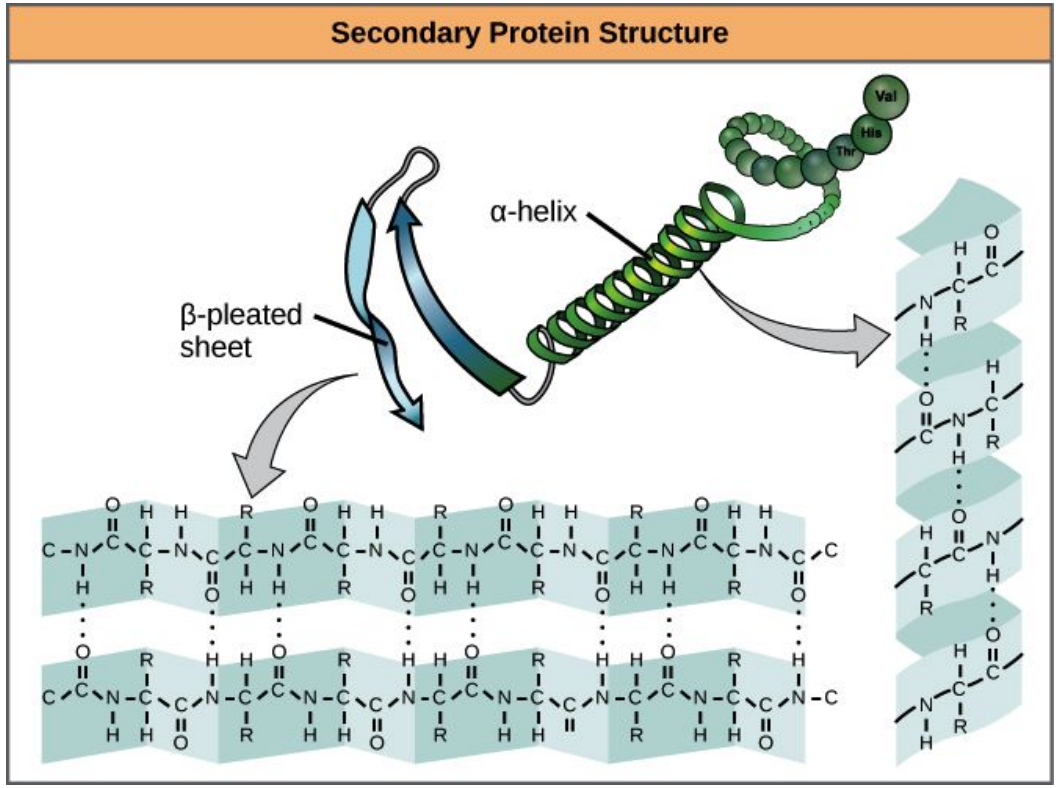
- A sequence of amino acids with the alphabet = "ARNDCQEGHILKMFPSTWYV")
- Connected by **Peptide Bond** -CO-NH-



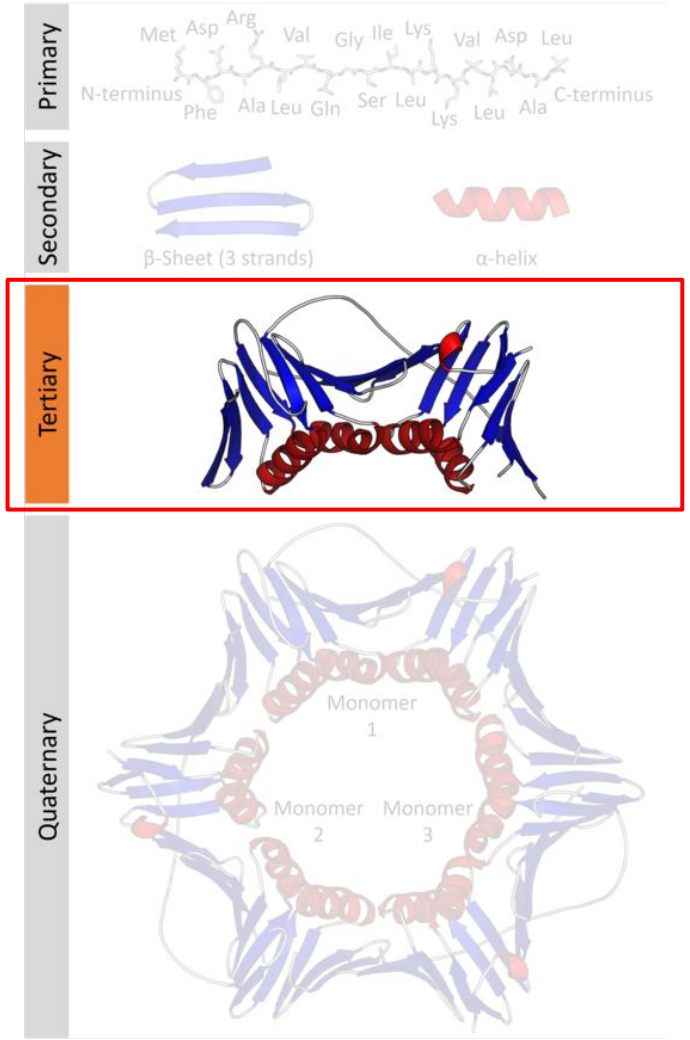
Four Levels of Protein Structures: Secondary



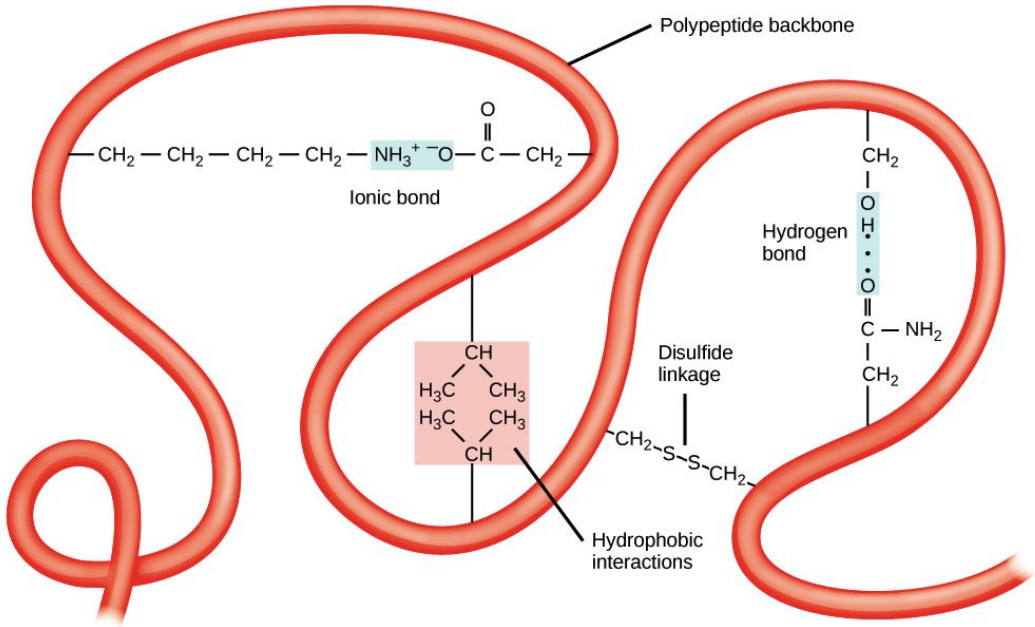
- Typical substructures: helices and sheets
- By Hydrogen Bonds



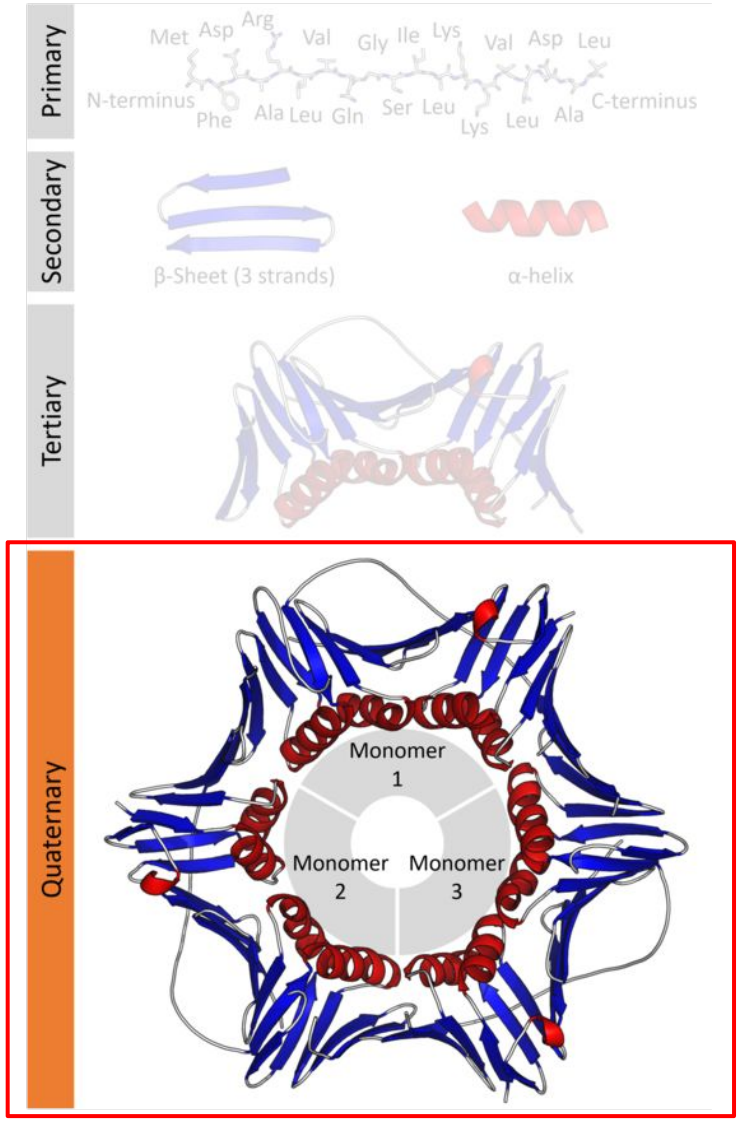
Four Levels of Protein Structures: Tertiary



- The overall three-dimensional structure of a polypeptide.
- Typically require deep knowledge about stereochemistry and more advanced expertise.
- **This is the level of prediction where AlphaFold (and AlphaFold 2) focus.**



Four Levels of Protein Structures: Quaternary

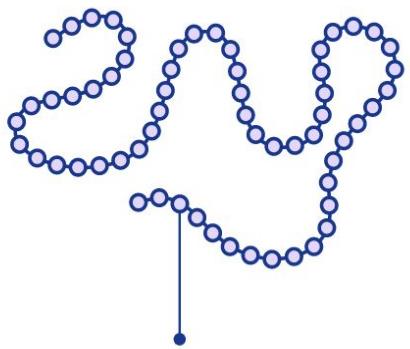


- Many proteins are made up of a single polypeptide chain and have only three levels of structure (the ones we've just discussed).
- However, some proteins are made up of multiple polypeptide chains, also known as subunits. When these subunits come together, they give the protein its **quaternary** structure.

Protein Structures: High-level summary

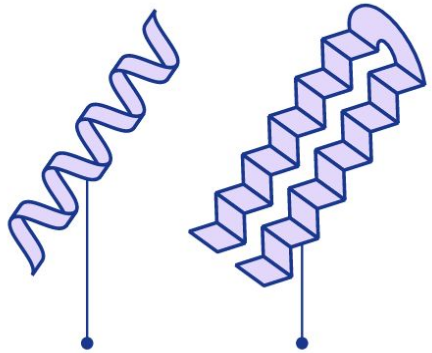
UCLA
Engineer Change.

Every protein is made up of a sequence of amino acids bonded together



Amino acids

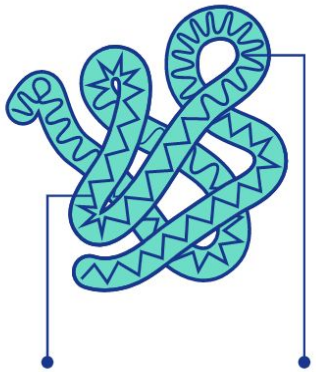
These amino acids interact locally to form shapes like helices and sheets



Alpha helix

Pleated sheet

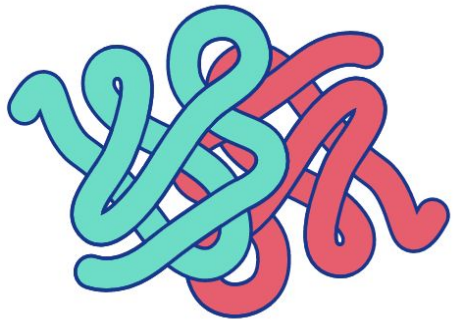
These shapes fold up on larger scales to form the full three-dimensional protein structure



Pleated sheet

Alpha helix

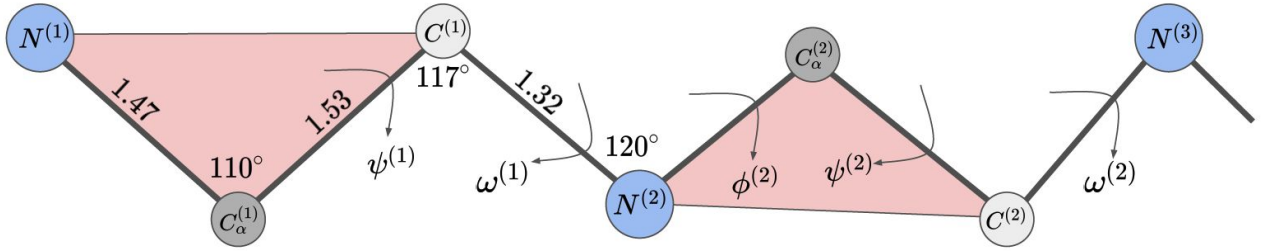
Proteins can interact with other proteins, performing functions such as signalling and transcribing DNA

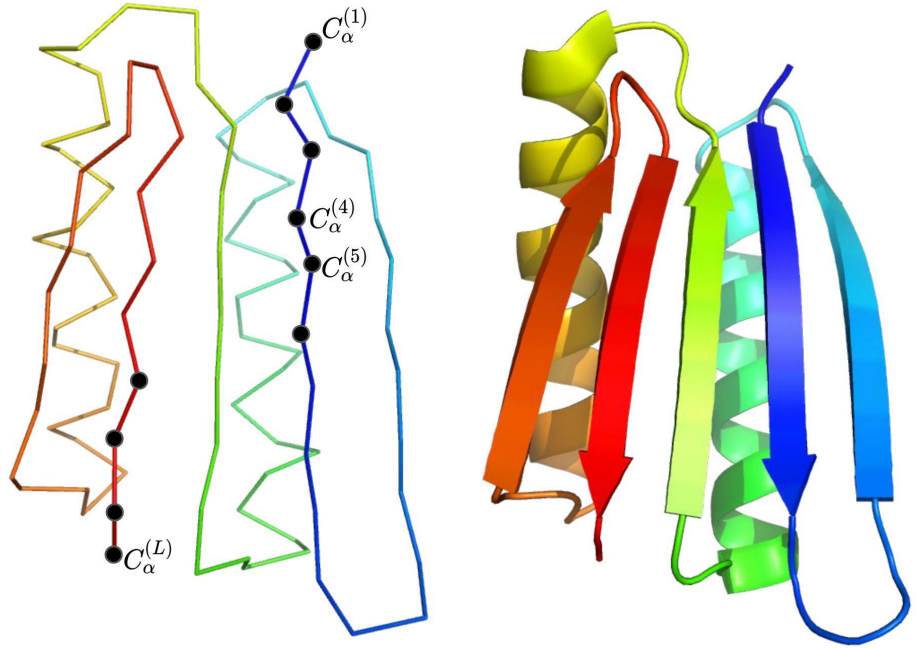


Protein Backbone Geometry

UCLA
Engineer Change.

A protein backbone is a repeating sequence (linear chain) of 3 atoms: nitrogen, carbon, and another carbon, namely $\underbrace{N^{(1)}, C_\alpha^{(1)}, C^{(1)}}, \underbrace{N^{(2)}, C_\alpha^{(2)}, C^{(2)}}, \ldots, \underbrace{N^{(L)}, C_\alpha^{(L)}, C^{(L)}}$

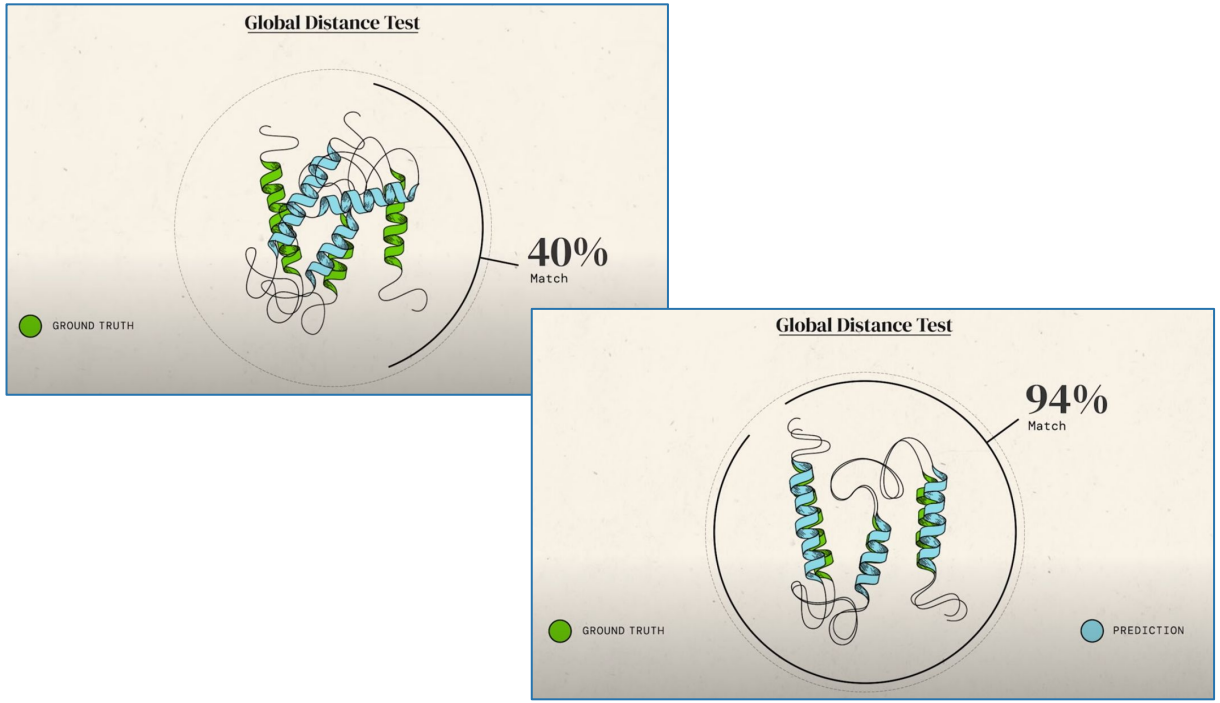


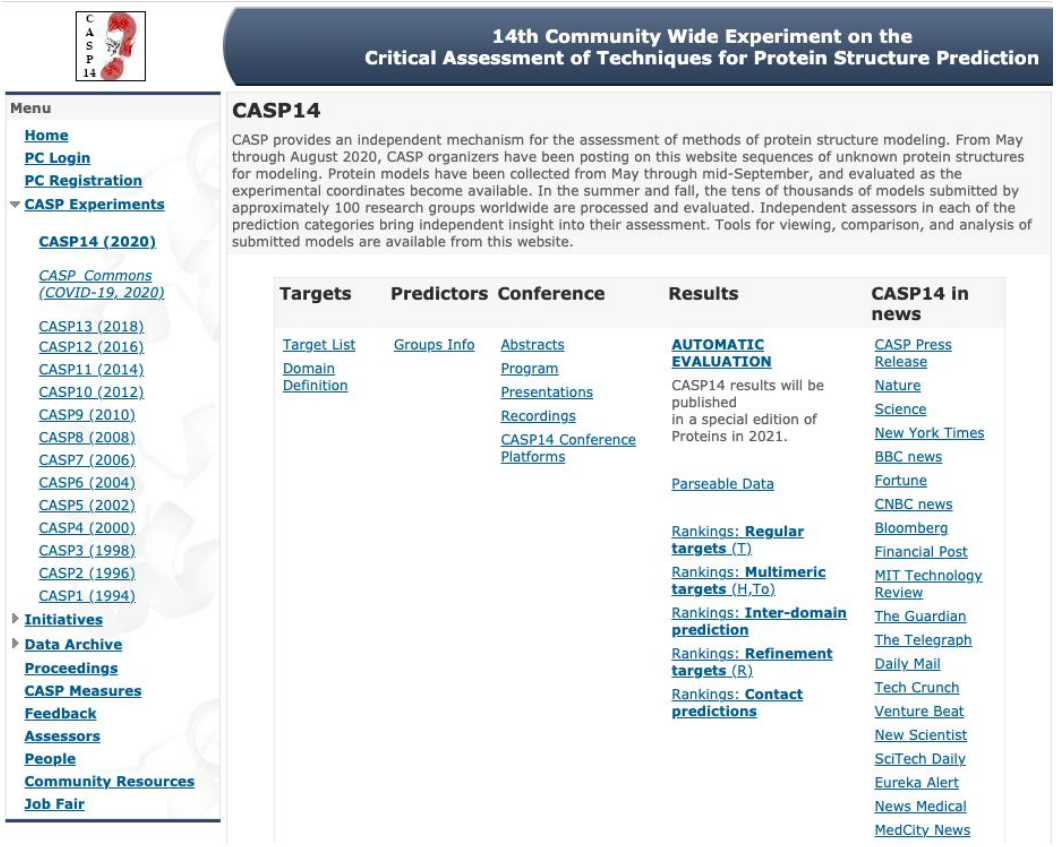


CASP: “Kaggle” on Protein Structure Prediction

UCLA
Engineer Change.

- Critical Assessment of protein Structure Prediction [Main Page]
  - Known as “Protein Structure Prediction Center”
- Evaluation
  - Global Distance Test (GDT)
  - TM-Score, RMSD





Dataset: What do we know about proteins?

UCLA
Engineer Change.

- **Sequence databases → 200M+**
  - UniRefA (JackHMMER)
  - BFD (HHblits)
  - MGnify clusters (JackHMMER)
- **Structural databases → Around 170K**
  - PDB (training)
  - PDB70 clustering (hhsearch)

References:

[1] Berman et al., Nature Structural Biology (2003) doi:10.1038/nsb1203-980

[2] Mitchell et al., Nucleic Acids Research (2019) doi:10.1093/nar/gkz1035

[3] Potter et al., Nucleic Acids Research (2018) doi:10.1093/nar/gky448

[4] Steinegger et al., BMC Bioinformatics (2019) doi:10.1186/s12859-019-3019-7

[5] Steinegger et al., Nature Methods (2019) doi:10.1038/s41592-019-0437-4

[6] Suzek et al., Bioinformatics (2015) doi:10.1093/bioinformatics/btu739

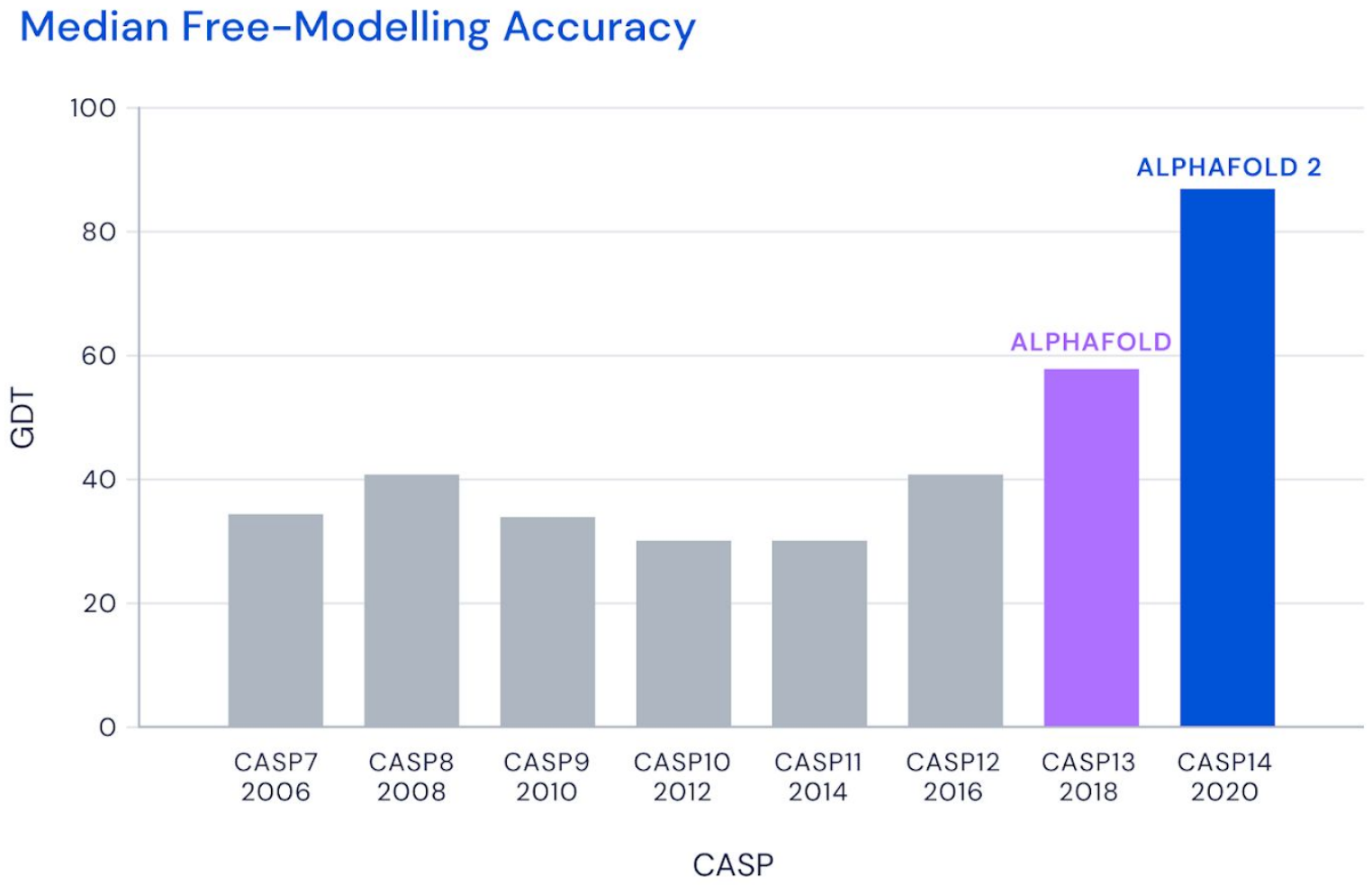
Visualization: PyMol https://pymol.org/2/

AlphaFold v1: Improved protein structure prediction using potentials from deep learning

One CNN-supported Protein Folding Model

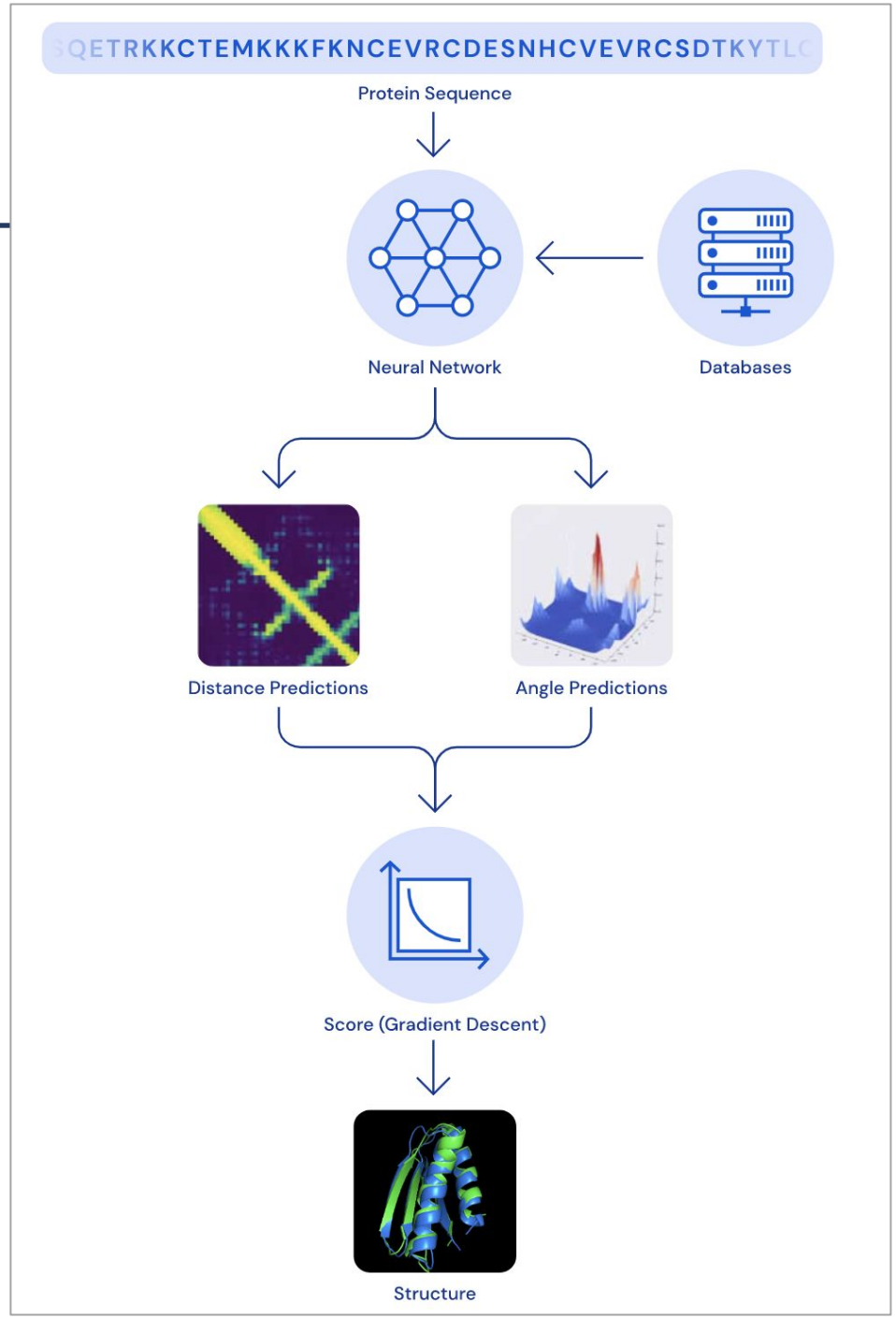
AlphaFold and AlphaFold on CAPS14 Challenge

UCLA
Engineer Change.

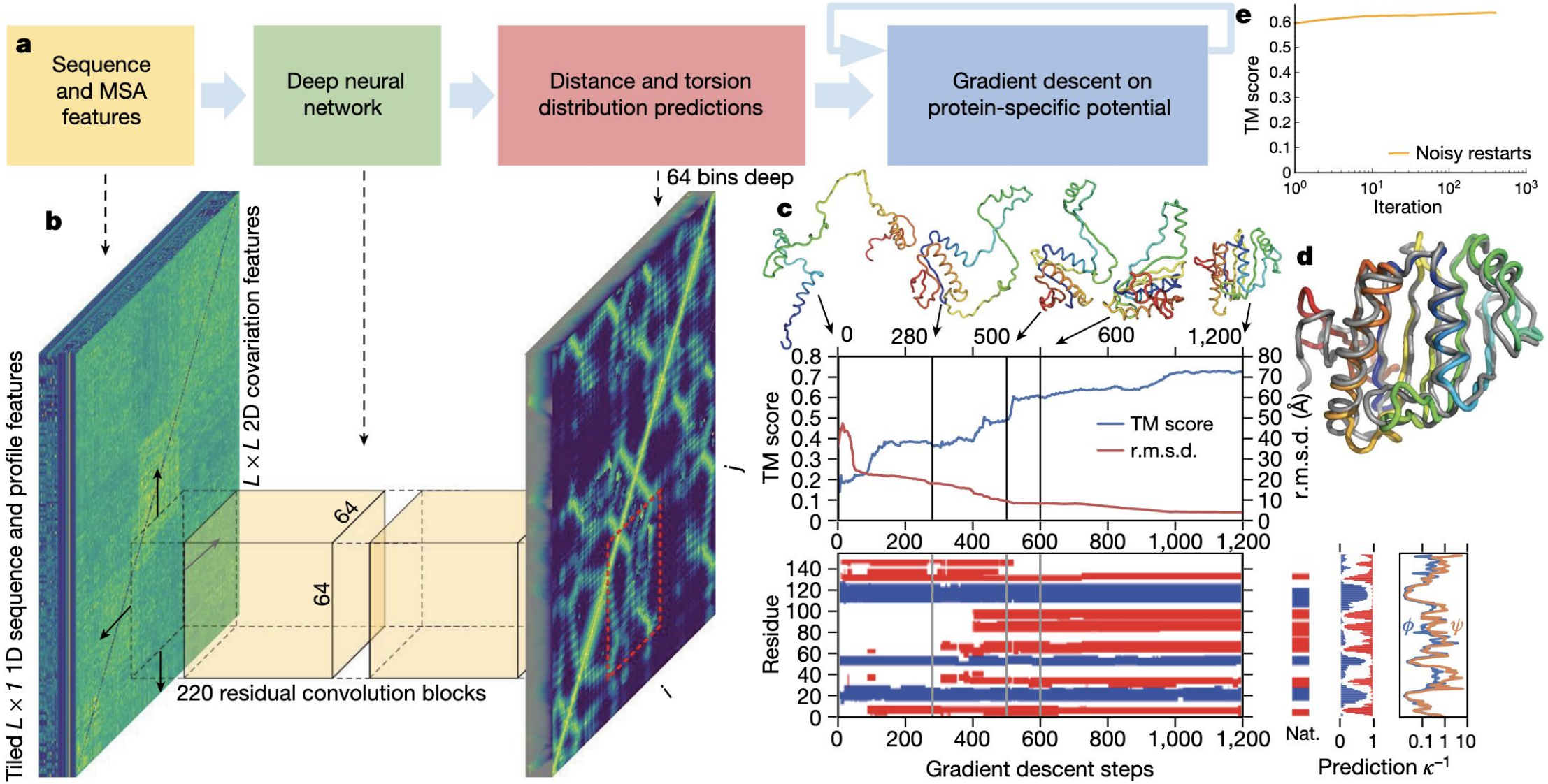


AlphaFold v1: Schematic Architecture

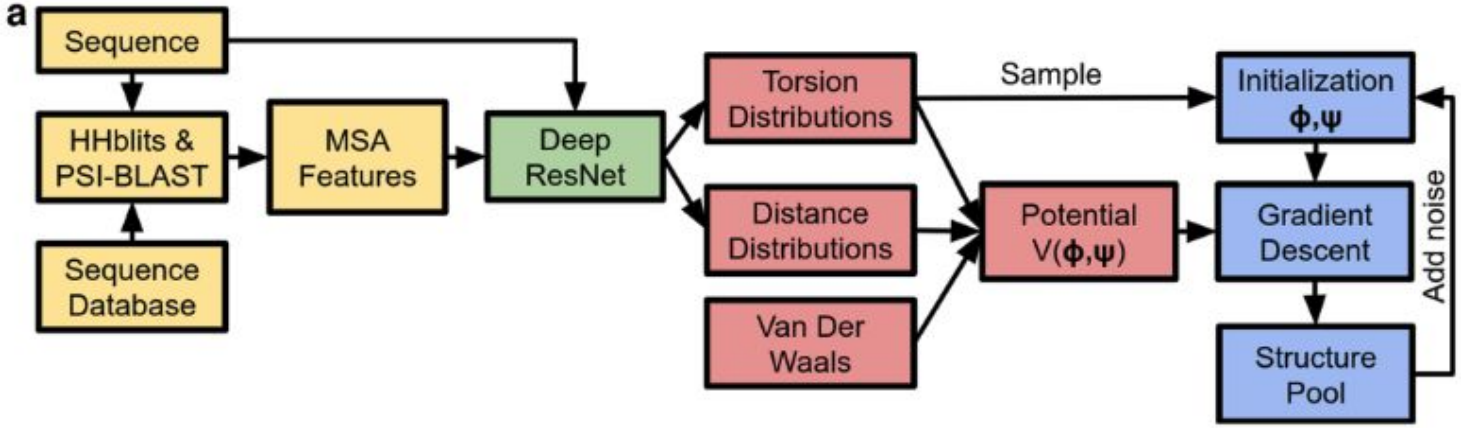
- Residual CNN as core model to predict distance and angle to create final structure output
- Using Multiple Sequence Alignment (MSA) from databases for feature generation

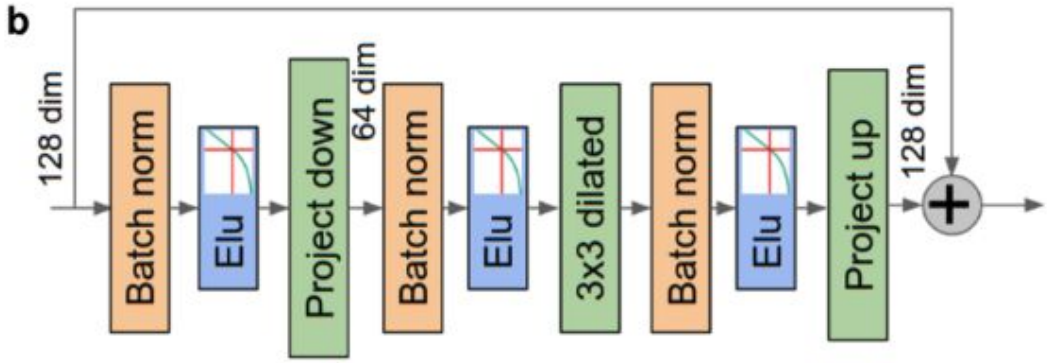


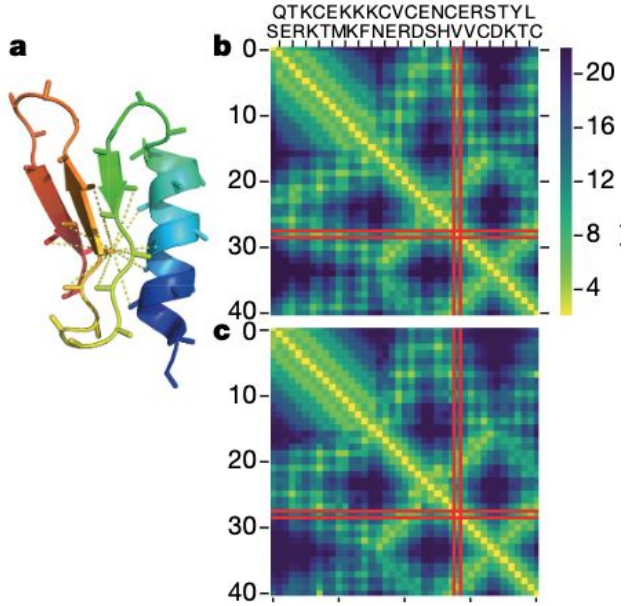
AlphaFold v1: Model Overview

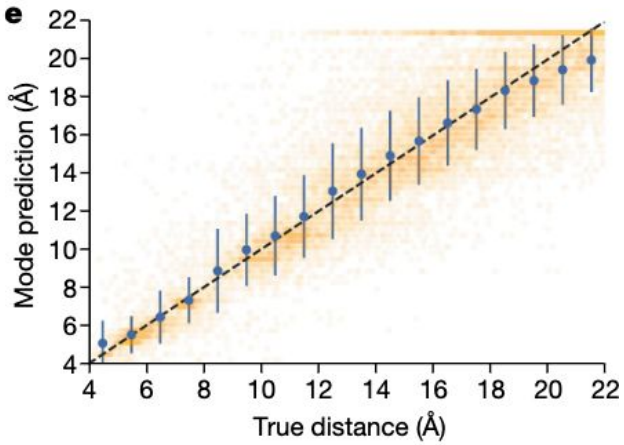


AlphaFold v1: Model Details



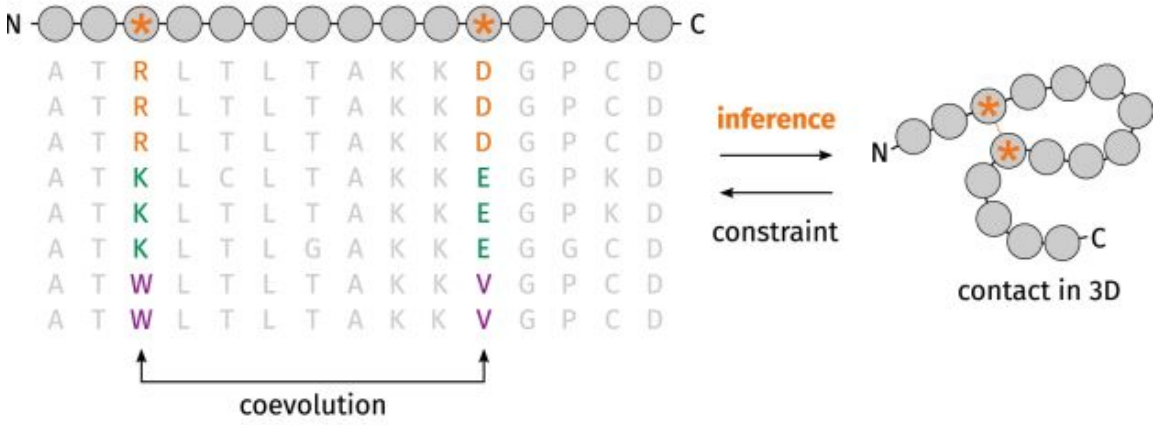


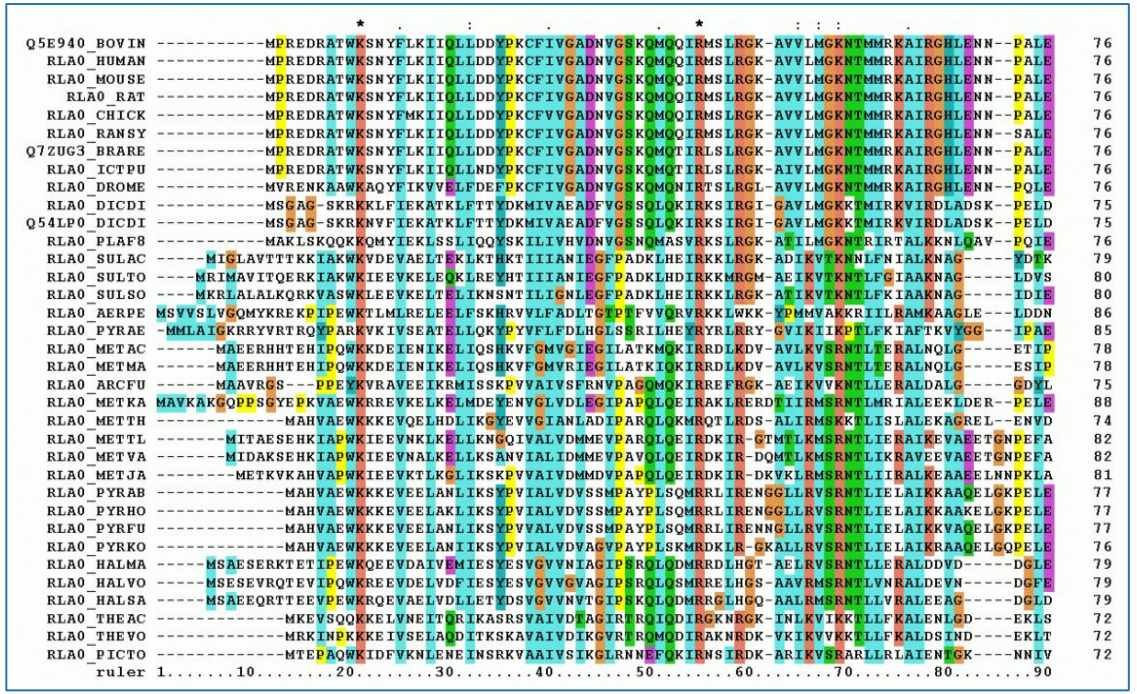




Multiple Sequence Alignment (MSA)

- Refer to the process or the result of sequence alignment of three or more biological sequences
- In AlphaFold, MSA is used to generate feature maps.
- Important indicator for structure information

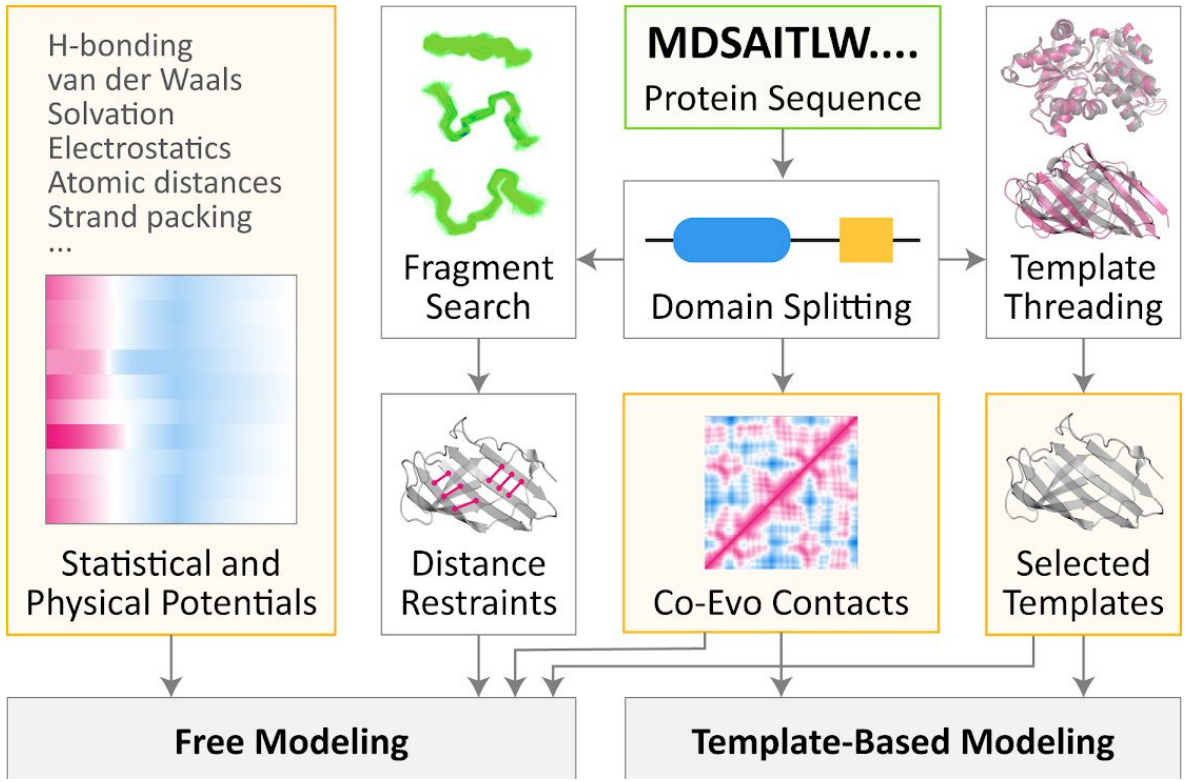


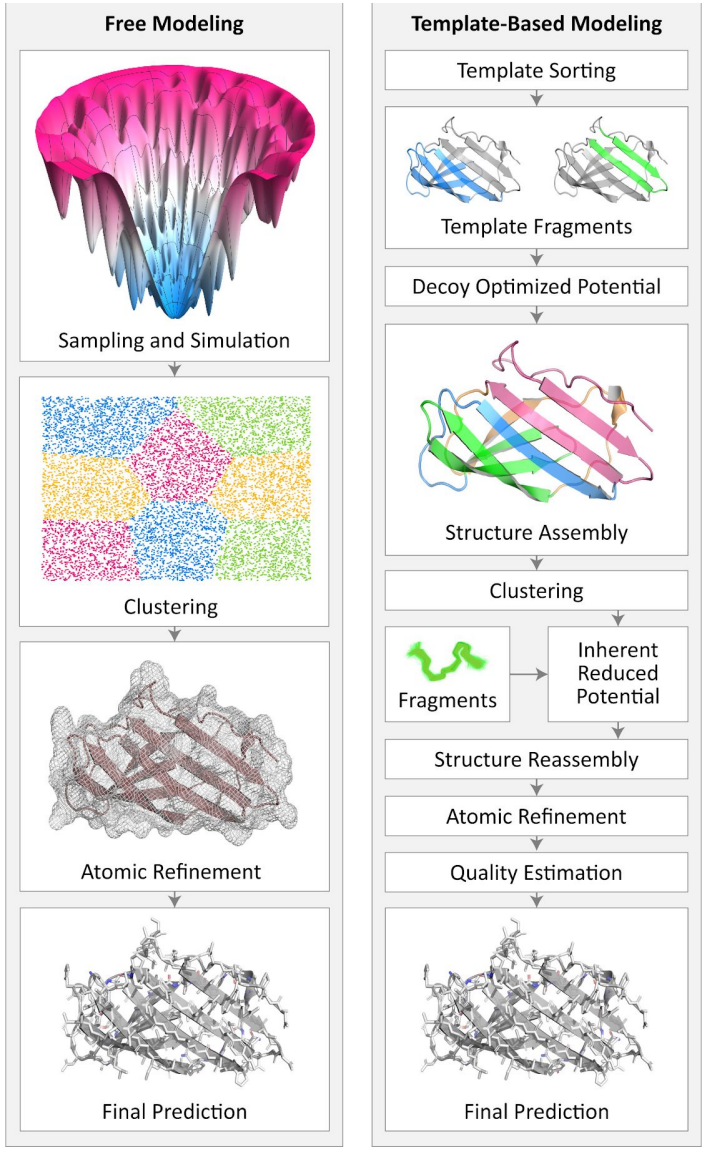


Protein Folding: Conventional Pipeline

UCLA
Engineer Change.

Example of Feature processing pipeline

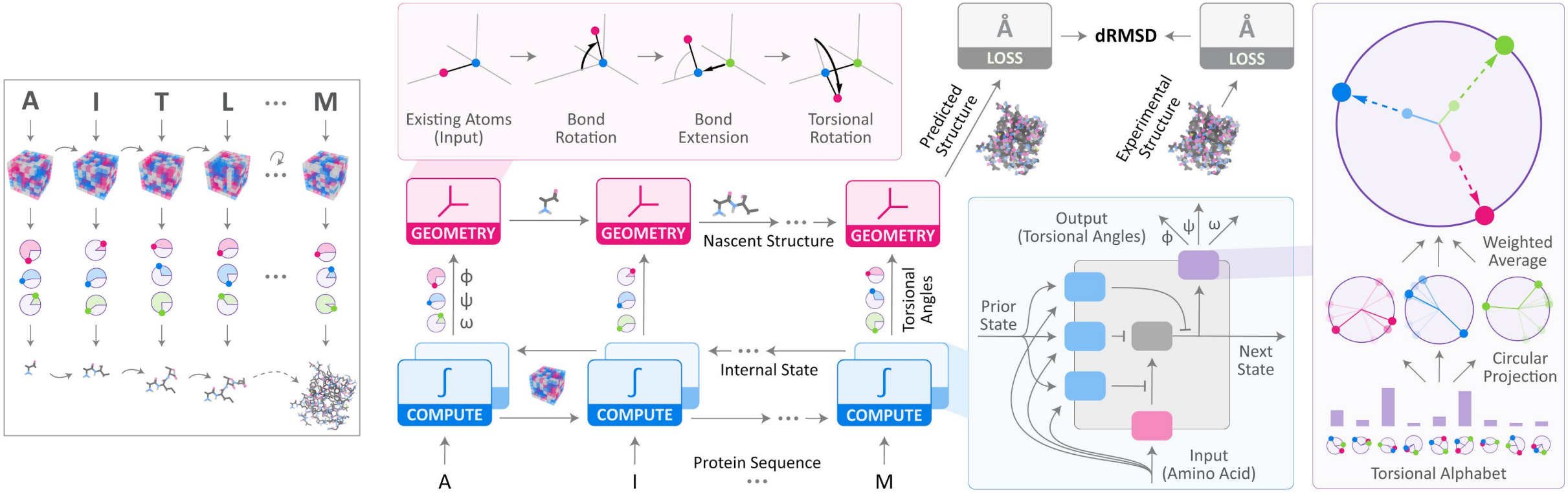




Another RNN Method for Protein Folding

UCLA
Engineer Change.

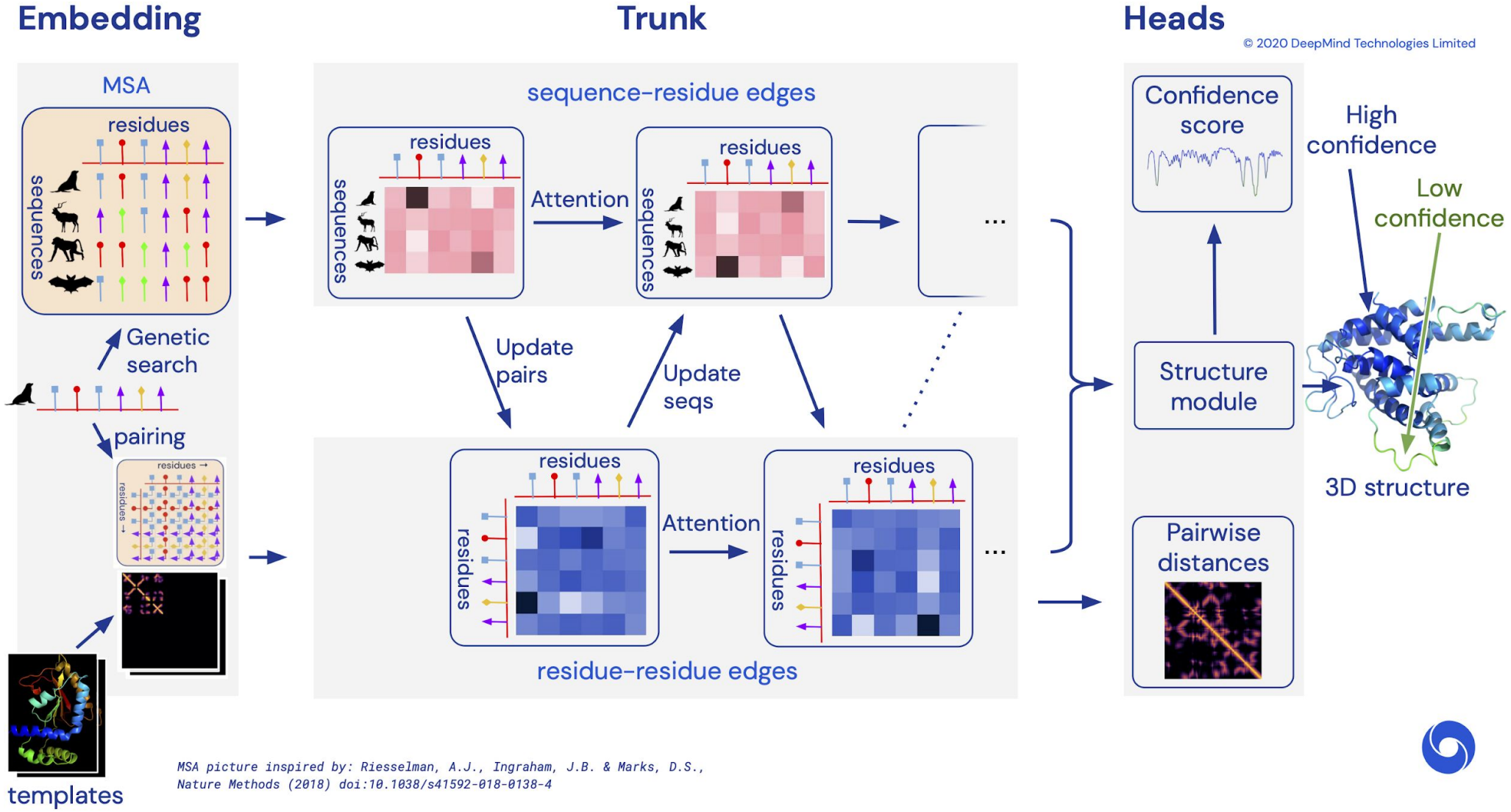
- End-to-End Differentiable Learning of Protein Structure, by Cell Systems



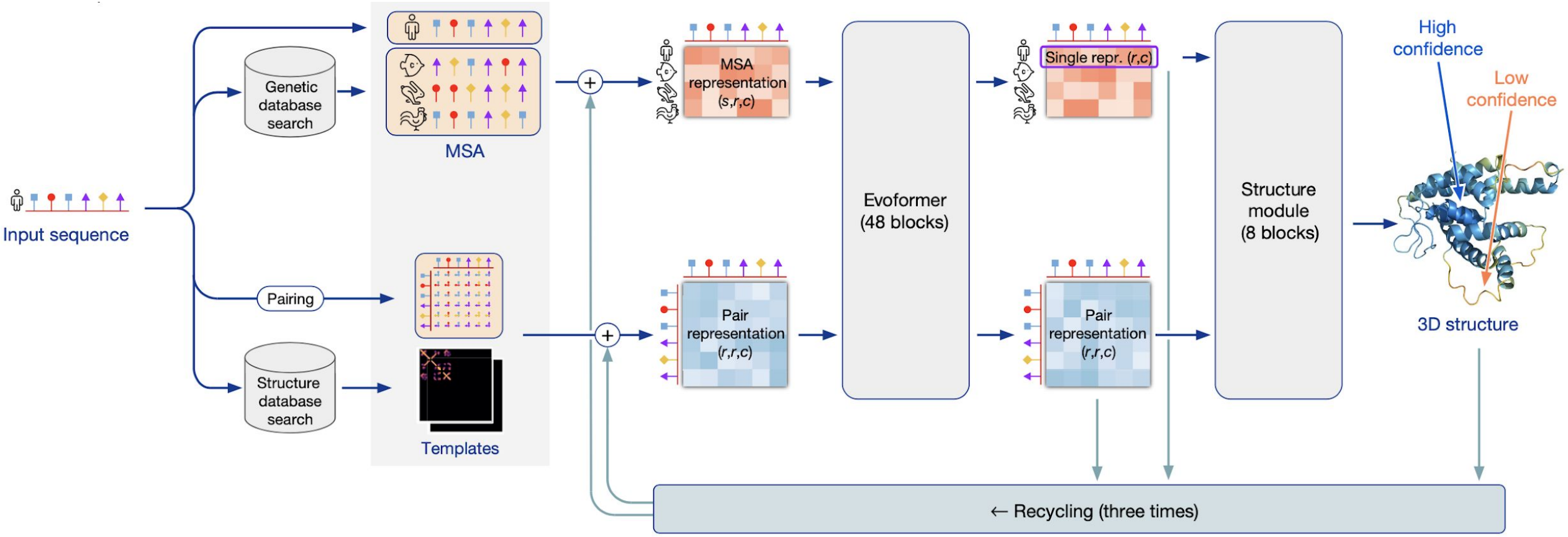
AlphaFold v2: Highly accurate protein structure prediction with AlphaFold

Where attention mechanism replace CNN and produce a breakthrough on the folding prediction

AlphaFold v2: Glance View

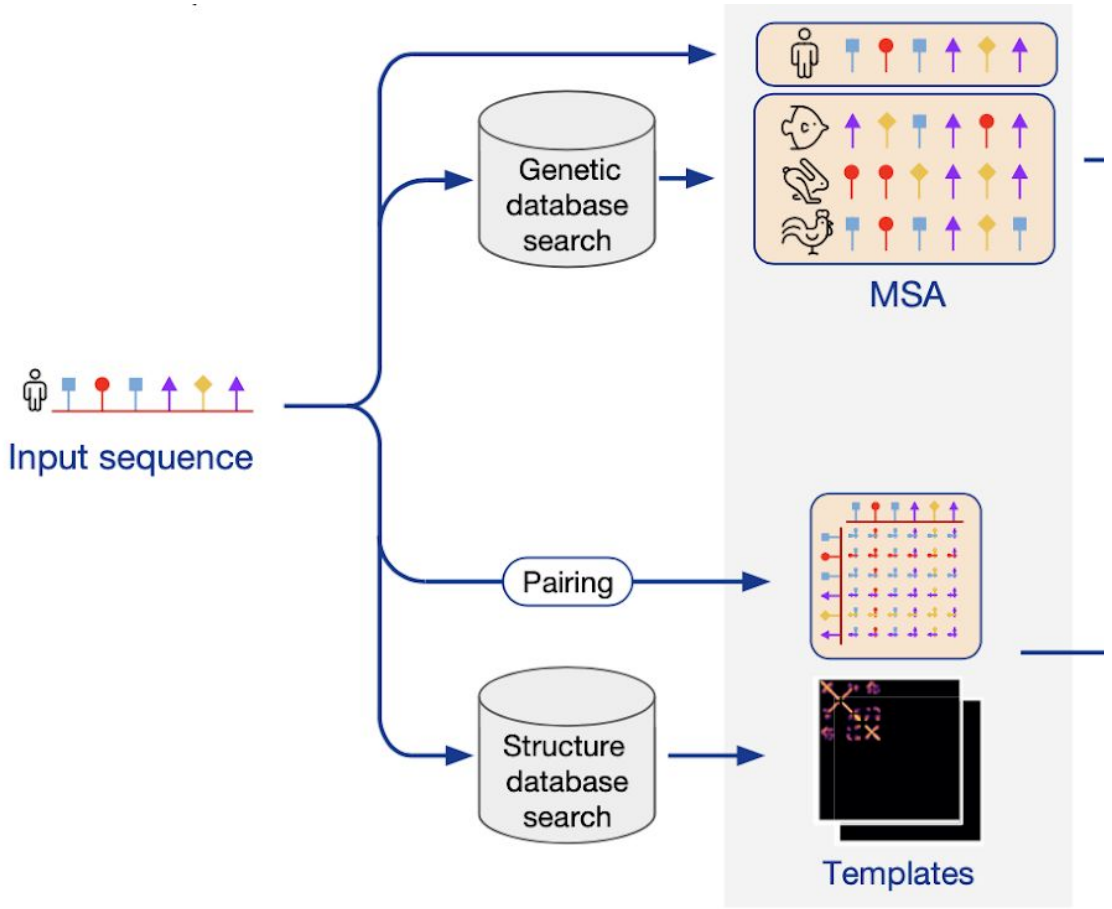


AlphaFold v2: Glance View (Clearer Version)

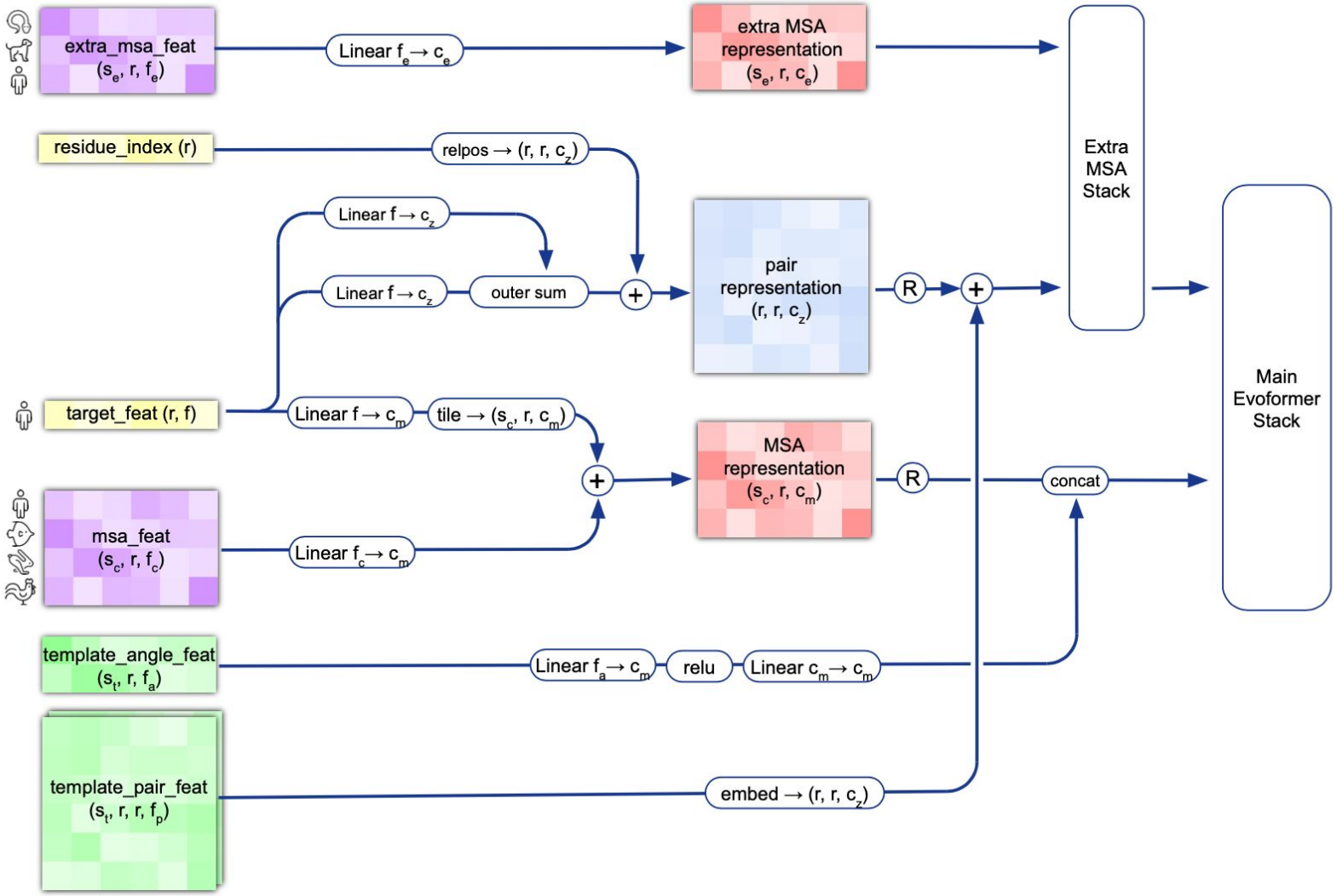


AlphaFold2: Input

- Not significantly different from AlphaFold v1, or even other models
- Input sequence, and leveraging some known knowledge
- MSA (**sequence-residue** from genetic database), in the shape of ***(s,r,c)***
- Templates (**residue-residue**, structure database from known proteins),), in the shape of ***(r,r,c)***

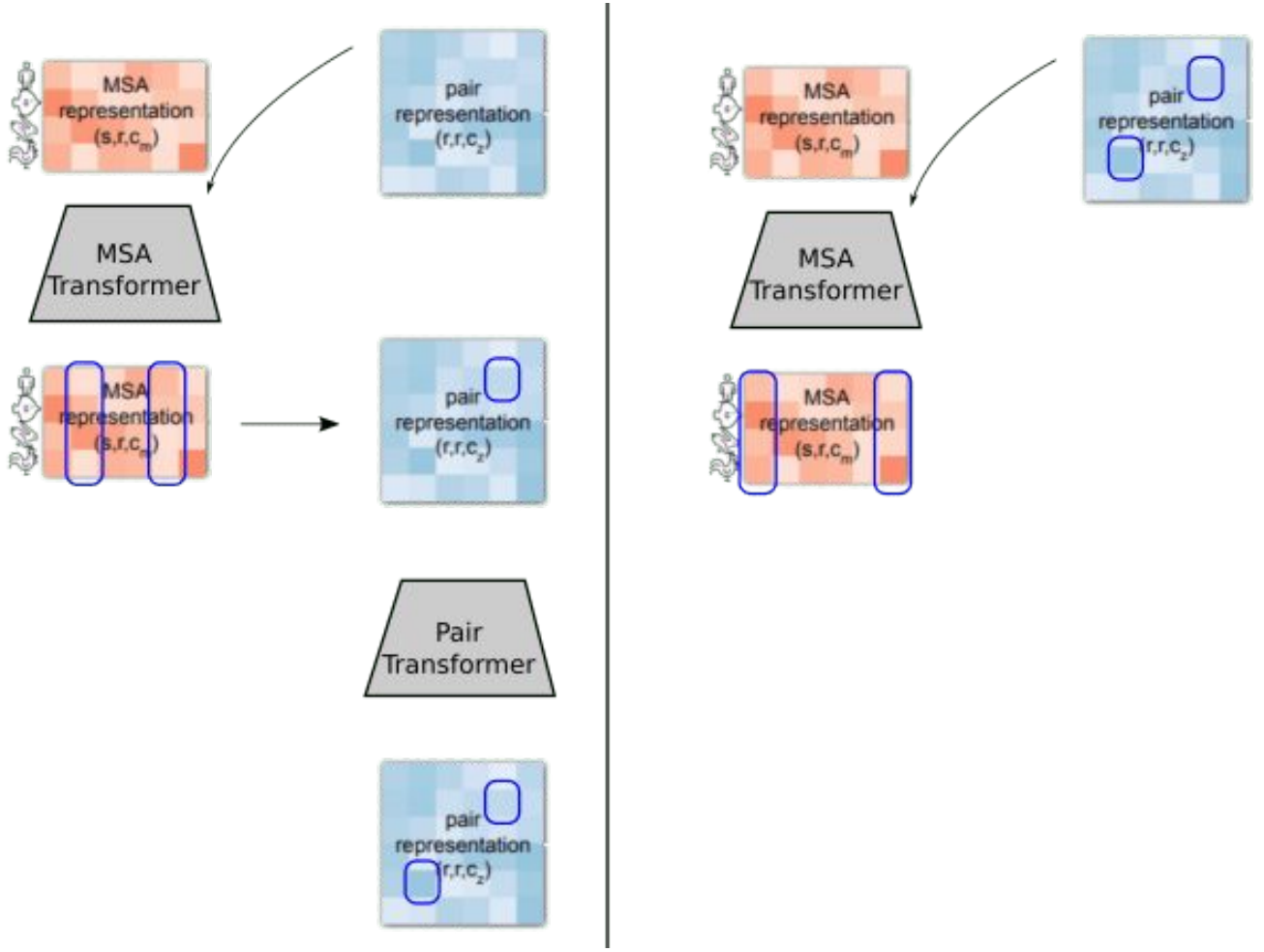


AlphaFold2: Input (Complete Version)



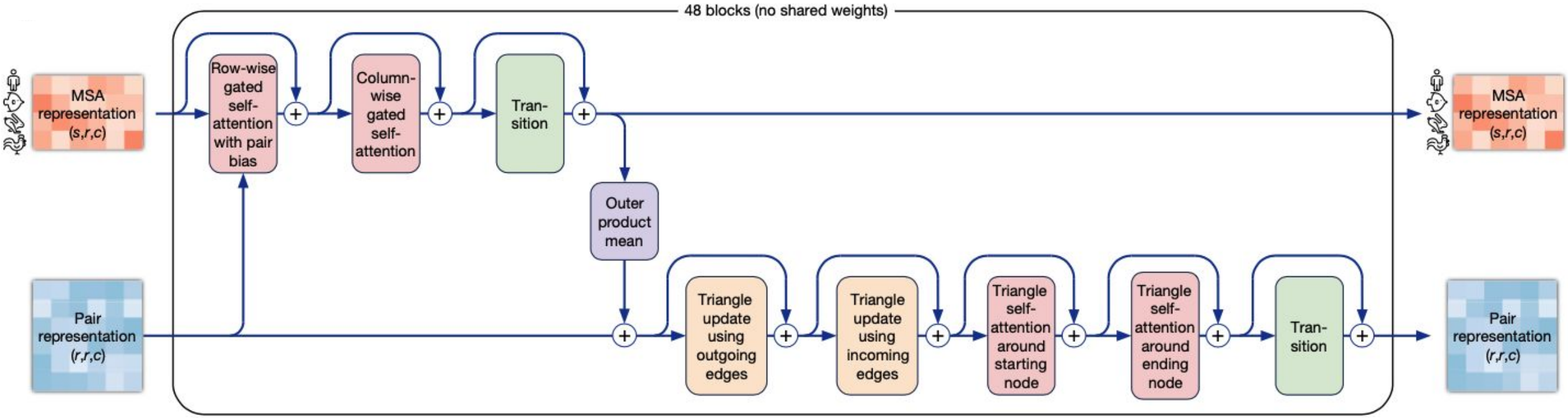
Evoformer: Evolutionary Transformer?

- Central idea: AlphaFold2 leverages the current structural hypothesis to improve the assessment of the multiple sequence alignment, which in turns leads to a new structural hypothesis, back and forth at every cycle.
- two transformers (a “two-tower architecture”), with one clear communication channel.



Evoformer Block

- Information flow in one Evoformer Block. A total of 48 Evo blocks are used.



Evoformer Stack: Algorithm Workflow

Algorithm 6 Evoformer stack

def EvoformerStack($\{\mathbf{m}_{si}\}, \{\mathbf{z}_{ij}\}, N_{\text{block}} = 48, c_s = 384$) :

for all $l \in [1, \ldots, N_{\text{block}}]$ **do**

 # MSA stack

$\{\mathbf{m}_{si}\}$ += DropoutRowwise$_{0.15}$(MSARowAttentionWithPairBias($\{\mathbf{m}_{si}\}, \{\mathbf{z}_{ij}\}$))

$\{\mathbf{m}_{si}\}$ += MSAColumnAttention($\{\mathbf{m}_{si}\}$)

$\{\mathbf{m}_{si}\}$ += MSATransition($\{\mathbf{m}_{si}\}$)

 # Communication

$\{\mathbf{z}_{ij}\}$ += OuterProductMean($\{\mathbf{m}_{si}\}$)

 # Pair stack

$\{\mathbf{z}_{ij}\}$ += DropoutRowwise$_{0.25}$(TriangleMultiplicationOutgoing($\{\mathbf{z}_{ij}\}$))

$\{\mathbf{z}_{ij}\}$ += DropoutRowwise$_{0.25}$(TriangleMultiplicationIncoming($\{\mathbf{z}_{ij}\}$))

$\{\mathbf{z}_{ij}\}$ += DropoutRowwise$_{0.25}$(TriangleAttentionStartingNode($\{\mathbf{z}_{ij}\}$))

$\{\mathbf{z}_{ij}\}$ += DropoutColumnwise$_{0.25}$(TriangleAttentionEndingNode($\{\mathbf{z}_{ij}\}$))

$\{\mathbf{z}_{ij}\}$ += PairTransition($\{\mathbf{z}_{ij}\}$)

end for

 # Extract the single representation

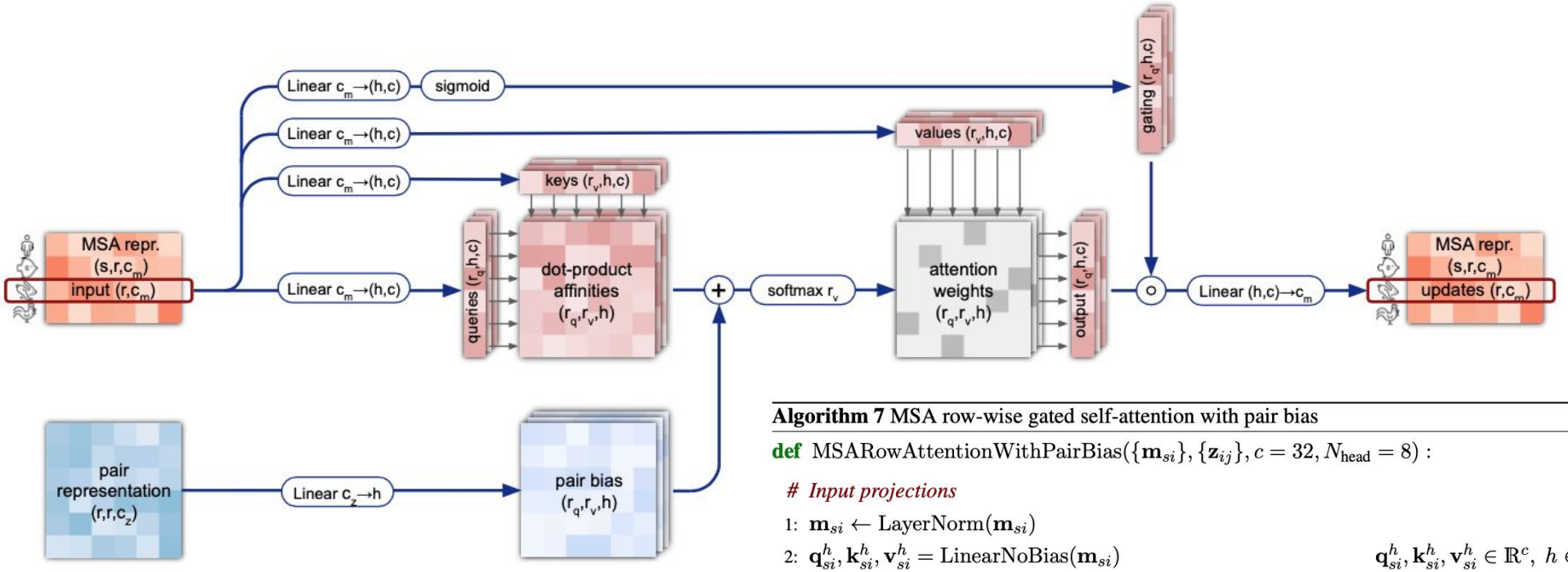
$\mathbf{s}_i = \text{Linear}(\mathbf{m}_{1i})$ $\quad \mathbf{s}_i \in \mathbb{R}^{c_s}$

return $\{\mathbf{m}_{si}\}, \{\mathbf{z}_{ij}\}, \{\mathbf{s}_i\}$

AlphaFold2's MSA Transformer

- The attention is "factorized" in "row-wise" and "column-wise" components.
- MSA Transformer first computes attention in the horizontal direction, allowing the network to identify which pairs of amino acids are more related; and then in the vertical direction, determining which sequences are more informative.
- MSA Transformer's row-wise (horizontal) attention mechanism incorporates information from the "pair representation".
- Gated attention applied.

AlphaFold2's MSA Row-wise Gated Attention



Algorithm 7 MSA row-wise gated self-attention with pair bias

def $\text{MSARowAttentionWithPairBias}(\{\mathbf{m}_{si}\}, \{\mathbf{z}_{ij}\}, c = 32, N_{\text{head}} = 8)$:

 # Input projections

$\mathbf{m}_{si} \leftarrow \text{LayerNorm}(\mathbf{m}_{si})$

$\mathbf{q}^h_{si}, \mathbf{k}^h_{si}, \mathbf{v}^h_{si} = \text{LinearNoBias}(\mathbf{m}_{si})$ $\quad \mathbf{q}^h_{si}, \mathbf{k}^h_{si}, \mathbf{v}^h_{si} \in \mathbb{R}^c,\ h \in \{1, \ldots, N_{\text{head}}\}$

$b^h_{ij} = \text{LinearNoBias}(\text{LayerNorm}(\mathbf{z}_{ij}))$

$\mathbf{g}^h_{si} = \text{sigmoid}\left(\text{Linear}(\mathbf{m}_{si})\right)$ $\quad \mathbf{g}^h_{si} \in \mathbb{R}^c$

 # Attention

$a^h_{sij} = \text{softmax}_j\left(\frac{1}{\sqrt{c}}\, \mathbf{q}^{h\top}_{si} \mathbf{k}^h_{sj} + b^h_{ij}\right)$

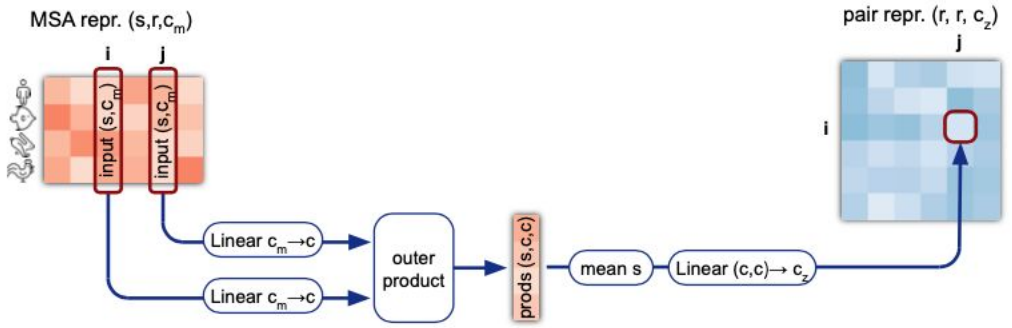
$\mathbf{o}^h_{si} = \mathbf{g}^h_{si} \odot \sum_j a^h_{sij} \mathbf{v}^h_{sj}$

 # Output projection

$\tilde{\mathbf{m}}_{si} = \text{Linear}\left(\text{concat}_h(\mathbf{o}^h_{si})\right)$ $\quad \tilde{\mathbf{m}}_{si} \in \mathbb{R}^{c_m}$

return $\{\tilde{\mathbf{m}}_{si}\}$

Evoformer: MSA Stack to Pair Stack



Supplementary Figure 5 | Outer product mean. Dimensions: s: sequences, r: residues, c: channels.

Algorithm 10 Outer product mean

def $\text{OuterProductMean}(\{\mathbf{m}_{si}\}, c = 32)$:

1: $\mathbf{m}_{si} \leftarrow \text{LayerNorm}(\mathbf{m}_{si})$

2: $\mathbf{a}_{si}, \mathbf{b}_{si} = \text{Linear}(\mathbf{m}_{si})$ $\quad \mathbf{a}_{si}, \mathbf{b}_{si} \in \mathbb{R}^{c}$

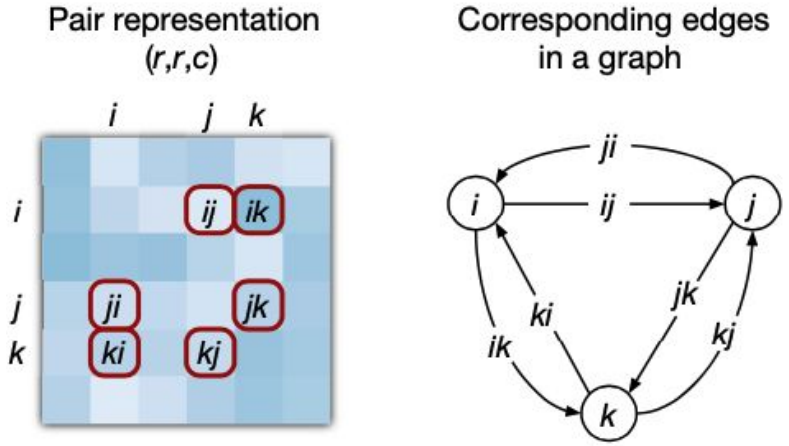
3: $\mathbf{o}_{ij} = \text{flatten}\left(\text{mean}_s(\mathbf{a}_{si} \otimes \mathbf{b}_{sj})\right)$ $\quad \mathbf{o}_{ij} \in \mathbb{R}^{c \cdot c}$

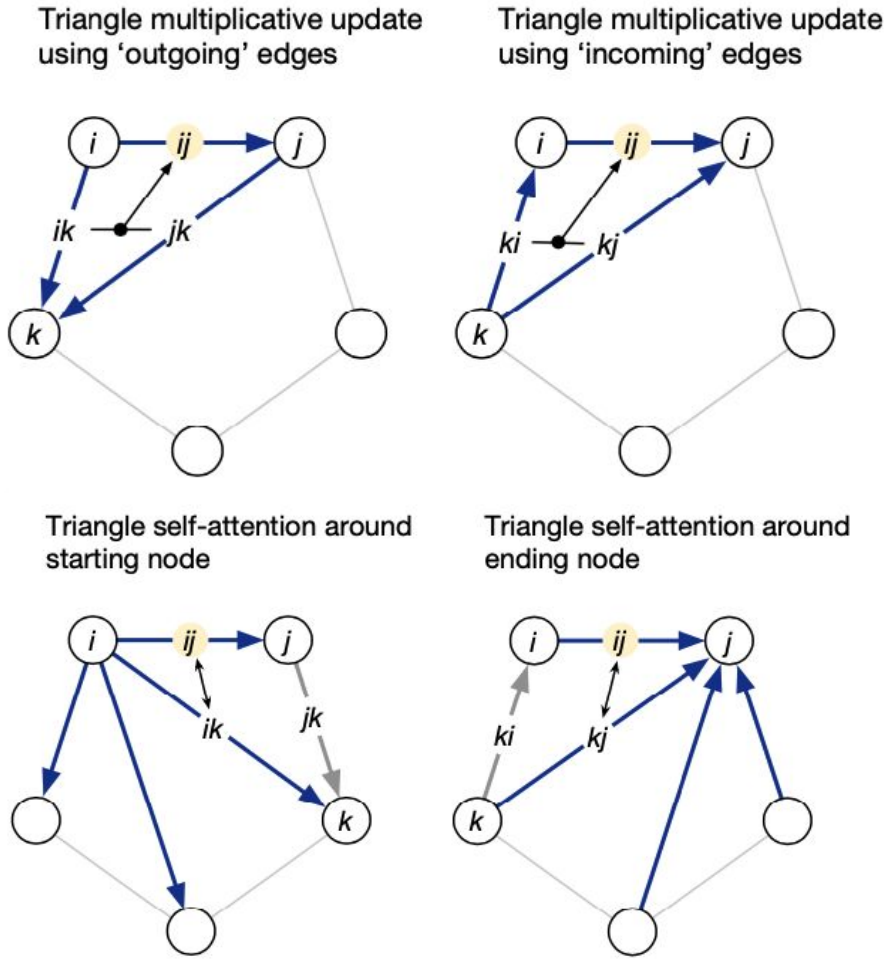
4: $\mathbf{z}_{ij} = \text{Linear}(\mathbf{o}_{ij})$ $\quad \mathbf{z}_{ij} \in \mathbb{R}^{c_z}$

5: **return** $\{\mathbf{z}_{ij}\}$

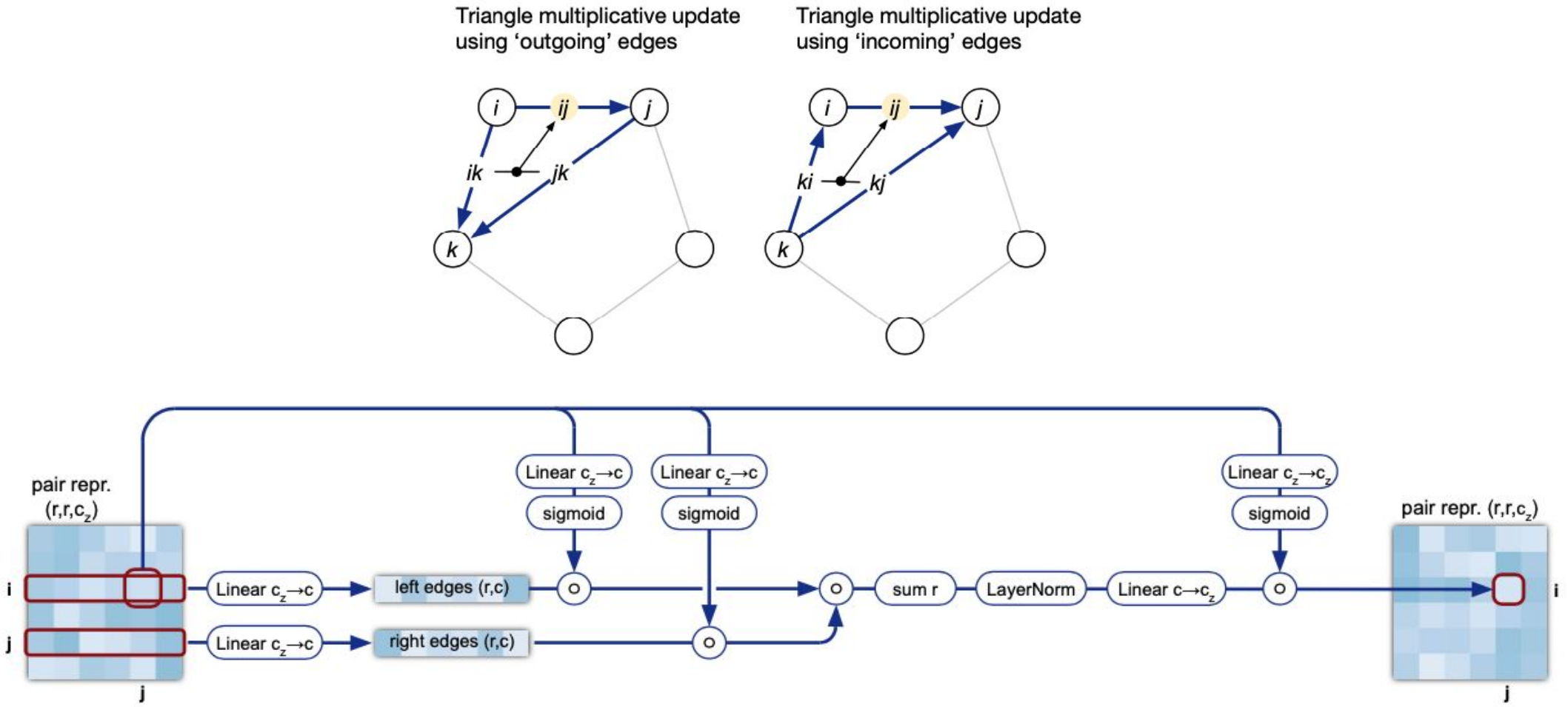
AlphaFold2's Pair Transformer

- Attention is arranged in terms of triangles of residues. □ Triangular attention



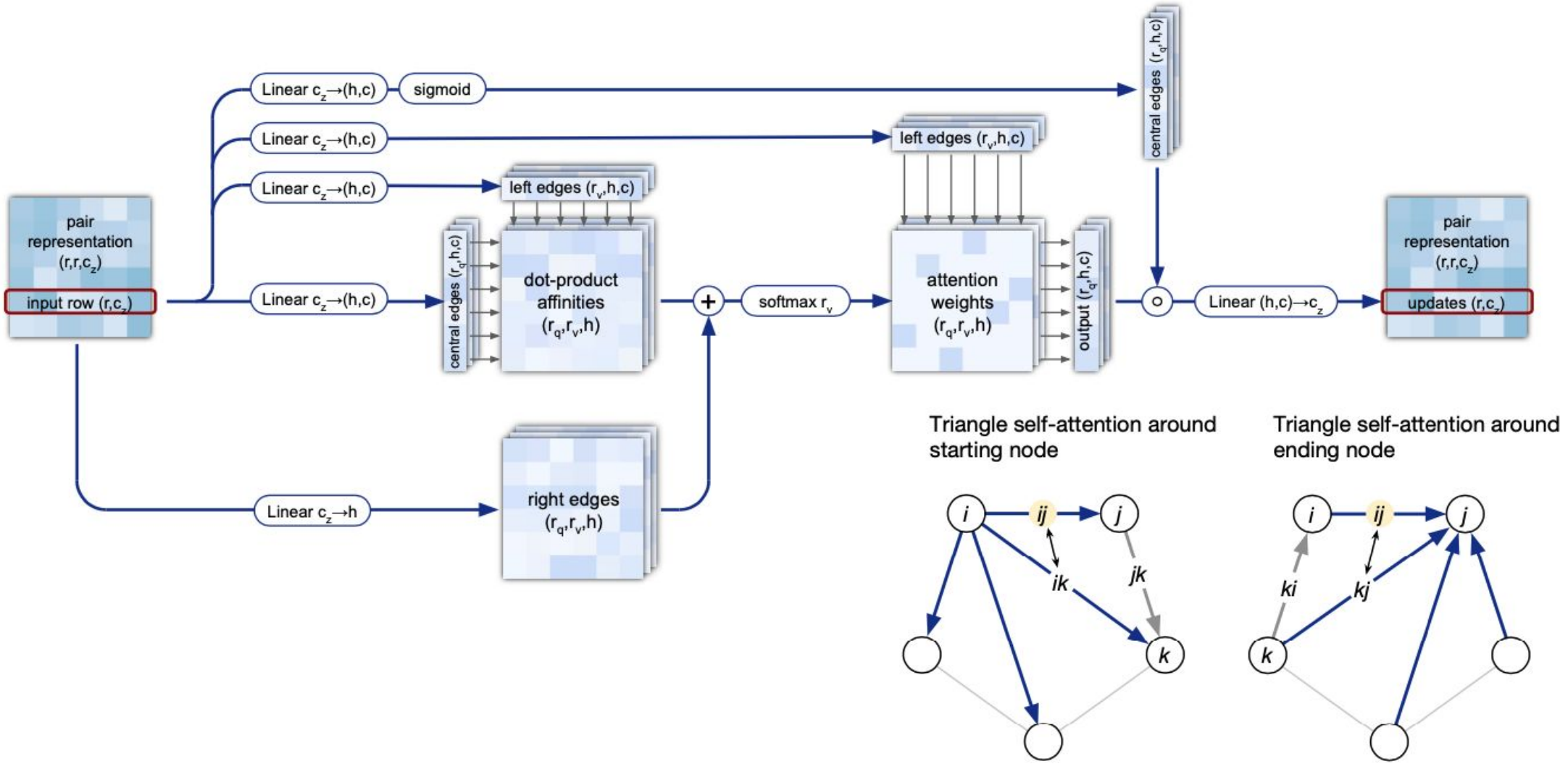


Pair Transformer: Triangular Multiplicate Update



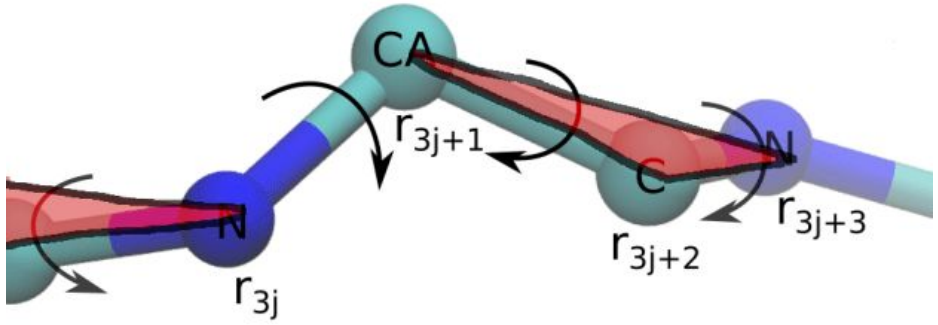
Supplementary Figure 6 | Triangular multiplicative update using "outgoing" edges. Dimensions: r: residues, c: channels.

Pair Transformer: Triangular Self-Attention

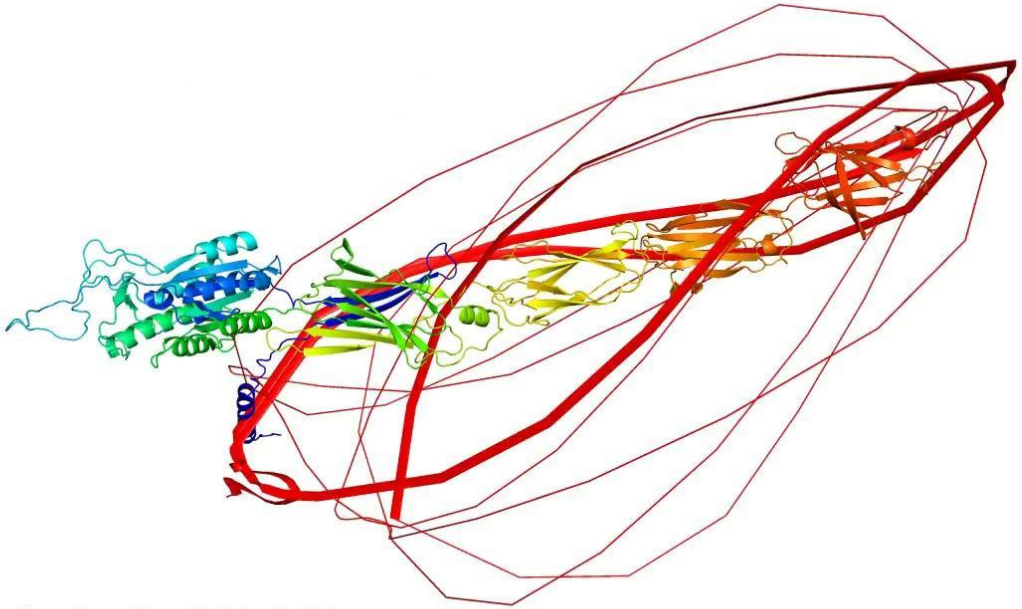


AlphaFold2: Structure module

- The structure module considers the protein as a “residue gas”, a floating backbone.
- Every amino acid is modelled as a triangle, representing the three atoms of the backbone.
- These triangles float around in space and are moved by the network to form the structure.
- These transformations are parametrized as “affine matrices”.
- At every step of the iterative process, AlphaFold 2 produces a set of affine matrices that displace and rotate the residues in space. *There are potential structural violations in stereochemistry.*



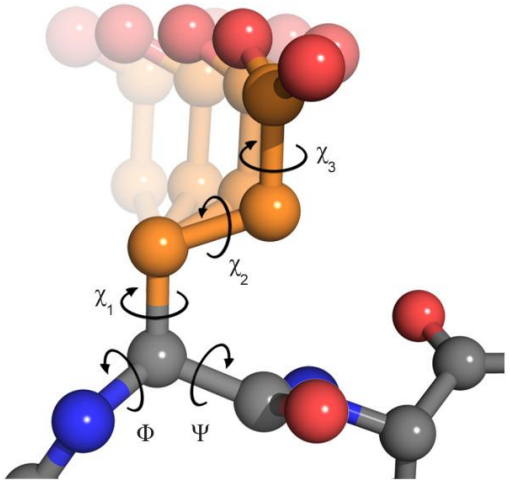
$$\mathbf{M} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

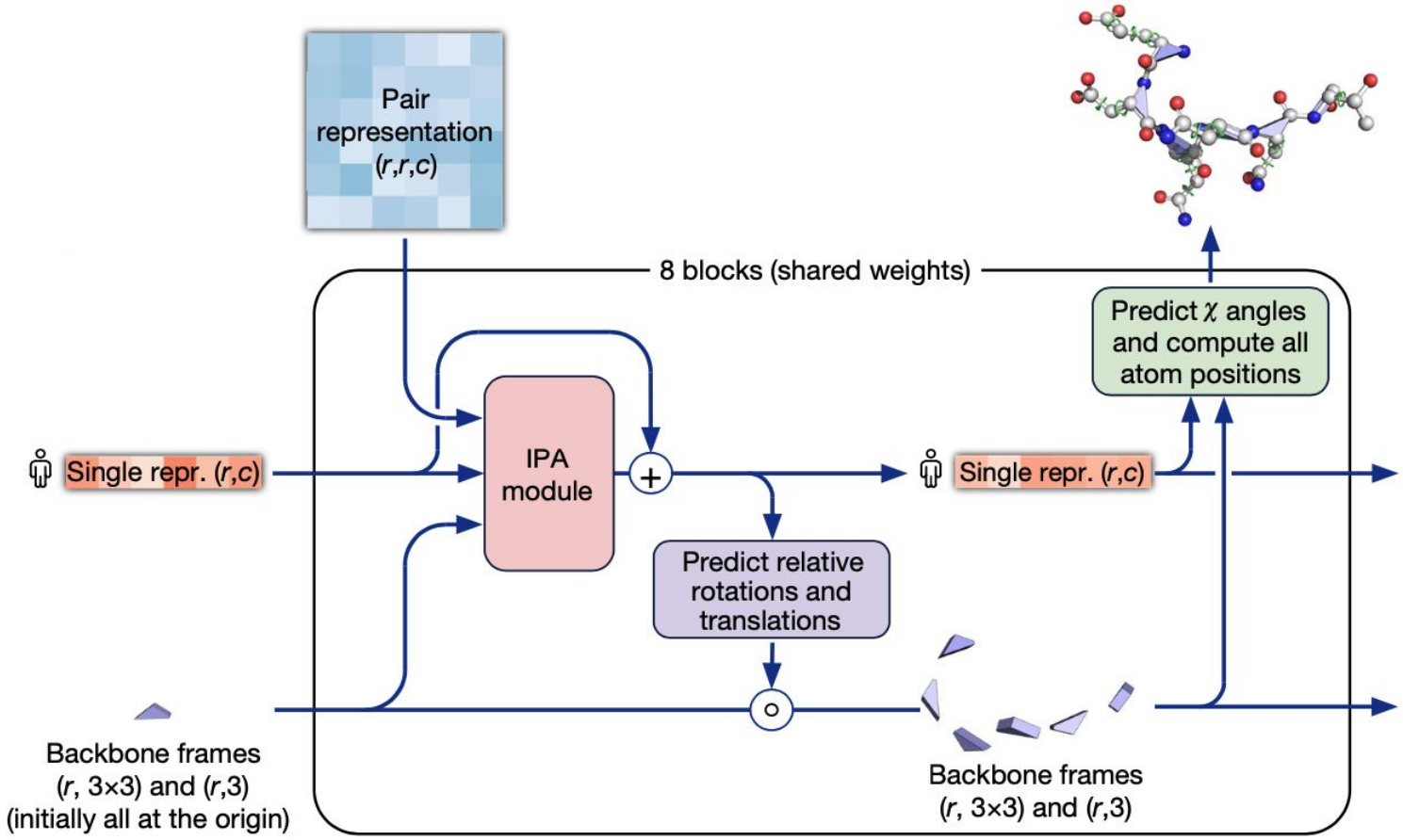


Recycling iteration 2, block 37
Secondary structure assigned from the final prediction

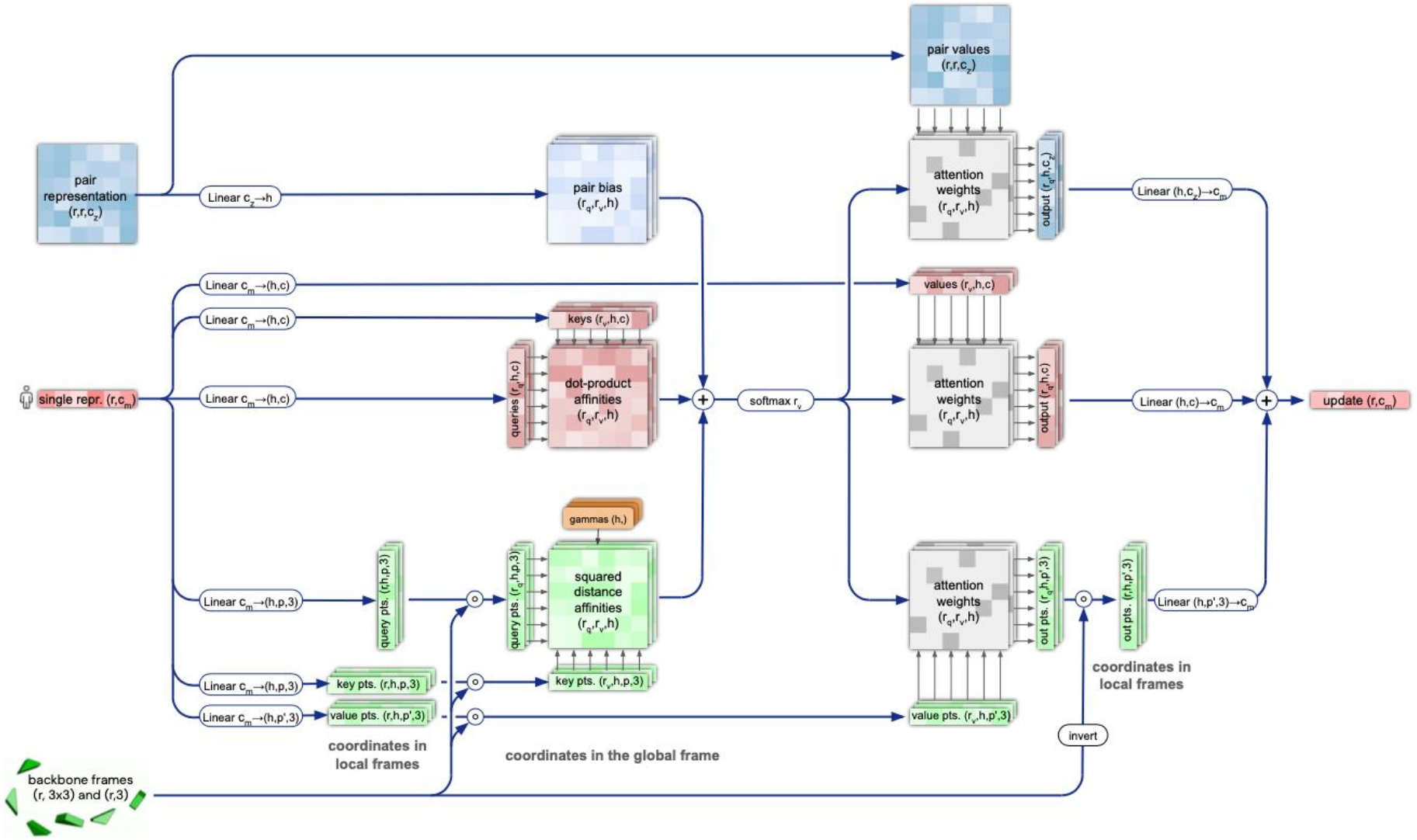
AlphaFold2: Structure module

- Contains one module named Invariant point attention (IPA)

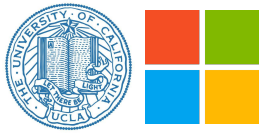




AlphaFold2: IPA Module



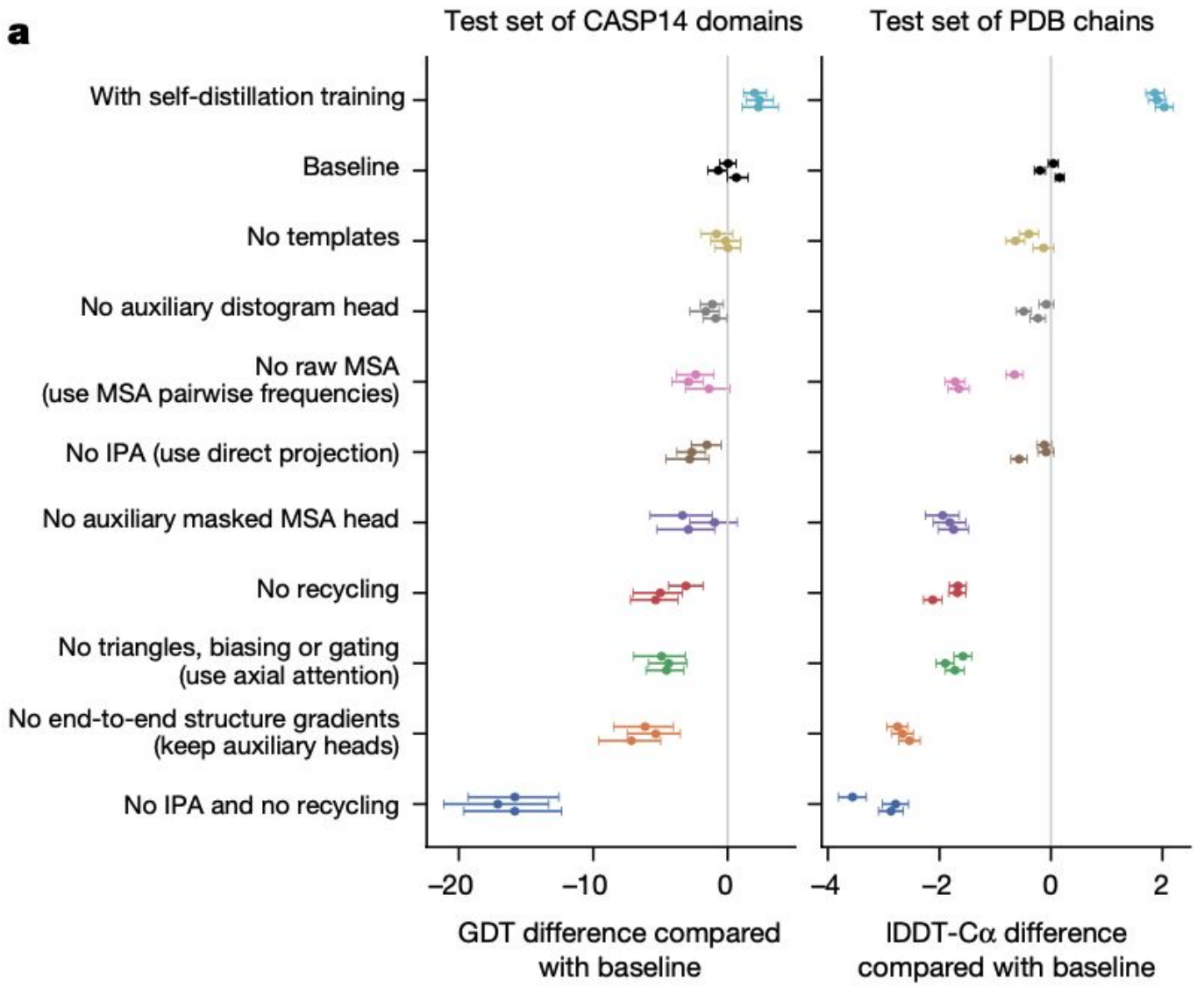
AlphaFold2: Quick Fact of Training Losses



- Specific structural loss which is called FAPE (Frame Aligned Point Error)
- Auxiliary loss: MSA Masking
  - The model is given a multiple sequence alignment with some symbols “masked out” and asked to predict these symbols. → Self-supervision
- **Self-distillation**
  - In this approach, they took a model trained exclusively on the *PDB (full structure details available) and* predicted the structures of ~300k diverse protein sequences obtained from *Uniclust (no structure available).*
  - They then retrained the full model, incorporating a small random sample of these structures (a high-confidence subset) at every training cycle.
  - They claim this allows the model to leverage the large amount of unlabeled data available in protein sequence repositories.
- Other tricks…

AlphaFold2: Tons of Engineering and Design

Ablation study of multiple variants



Summary

- AI can contribute to basic scientific discovery, with the hope of making real-world impact, such as AlphaFold(2) in the realm of protein biology.
- A tool like AlphaFold might help rare disease researchers predict the shape of a protein of interest rapidly and economically.
- Physical insights are built into the network structure, instead of just data preprocessing or feature selection and curation.
- However, AlphaFold(2), similar to many computational biology model, are not verified nor experimented in “wet lab” and still skeptical to many biologists and pharmaceutical industry.

AlphaFold v2: Protein Structure Database, Source Code and Demo

Run AlphaFold2 on Google Colab

AlphaFold2 Protein Database

Demo (ACE2-HUMAN): https://alphafold.ebi.ac.uk/entry/Q9BYF1

Protein	Angiotensin-converting enzyme 2
Gene	ACE2
Source organism	Homo sapiens go to search
UniProt	Q9BYF1 go to UniProt
Experimental structures	63 structures in PDB for Q9BYF1 go to PDBe-KB
Biological function	(Microbial infection) Non-functional as a receptor for human coronavirus SARS-CoV-2. go to UniProt

3D viewer

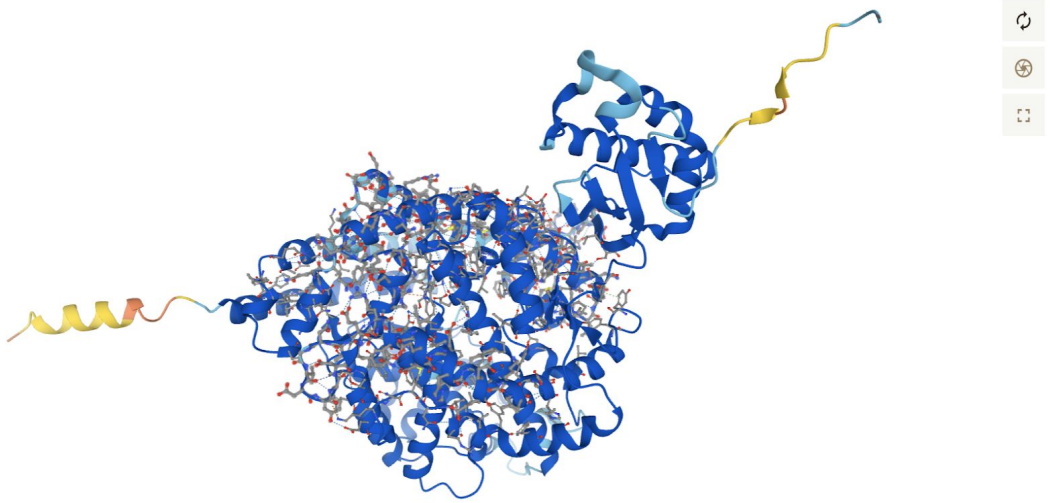
Model Confidence:

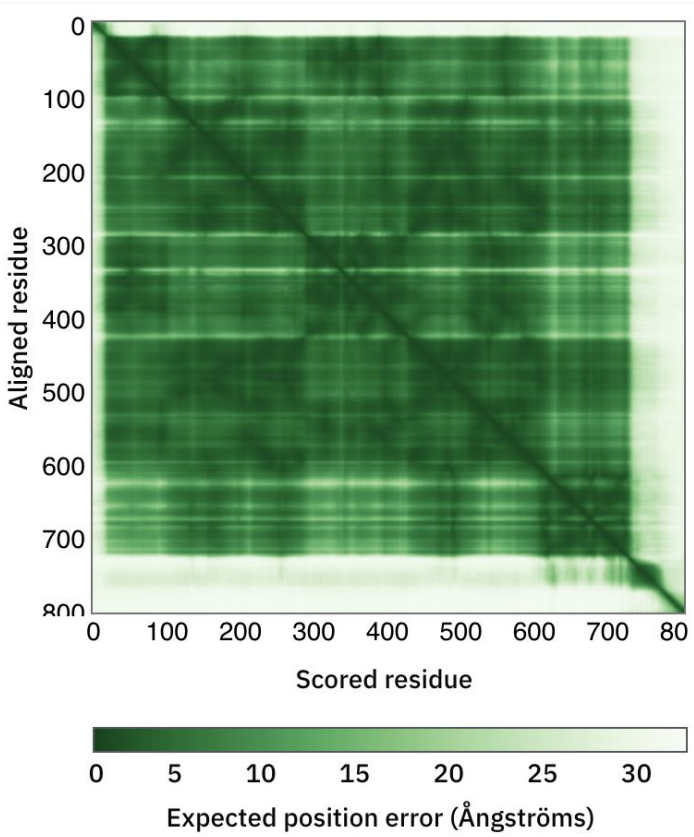
- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

Sequence of AF-Q9BYF1-F1 1: Angiotensi... A

```
MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQNMNNAGDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLNTILNTMS
TIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLYEEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDYSRGQLIEDVEHTFEEIKPLYEHLHAYVRAKL
MNAYPSYISPIGCLPAHLLGDMWGRFWTNLYSLTVPFGQKPNIDVTDAMVDQAWDAQRIFKEAEKFFVSVGLPNMTQGFWENSMLTDPGNVQKAVCHPTAWDLGKGDFRILMCTKVTMDDFLTA
```





Source Code and AlphaFold on Google Colab

Source Code: https://github.com/deepmind/alphafold/

Original AlphaFold Colab: https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb

AlphaFold2 and advanced version (*not* authored by Google/DeepMind): https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb

https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/beta/AlphaFold2_advanced.ipynb

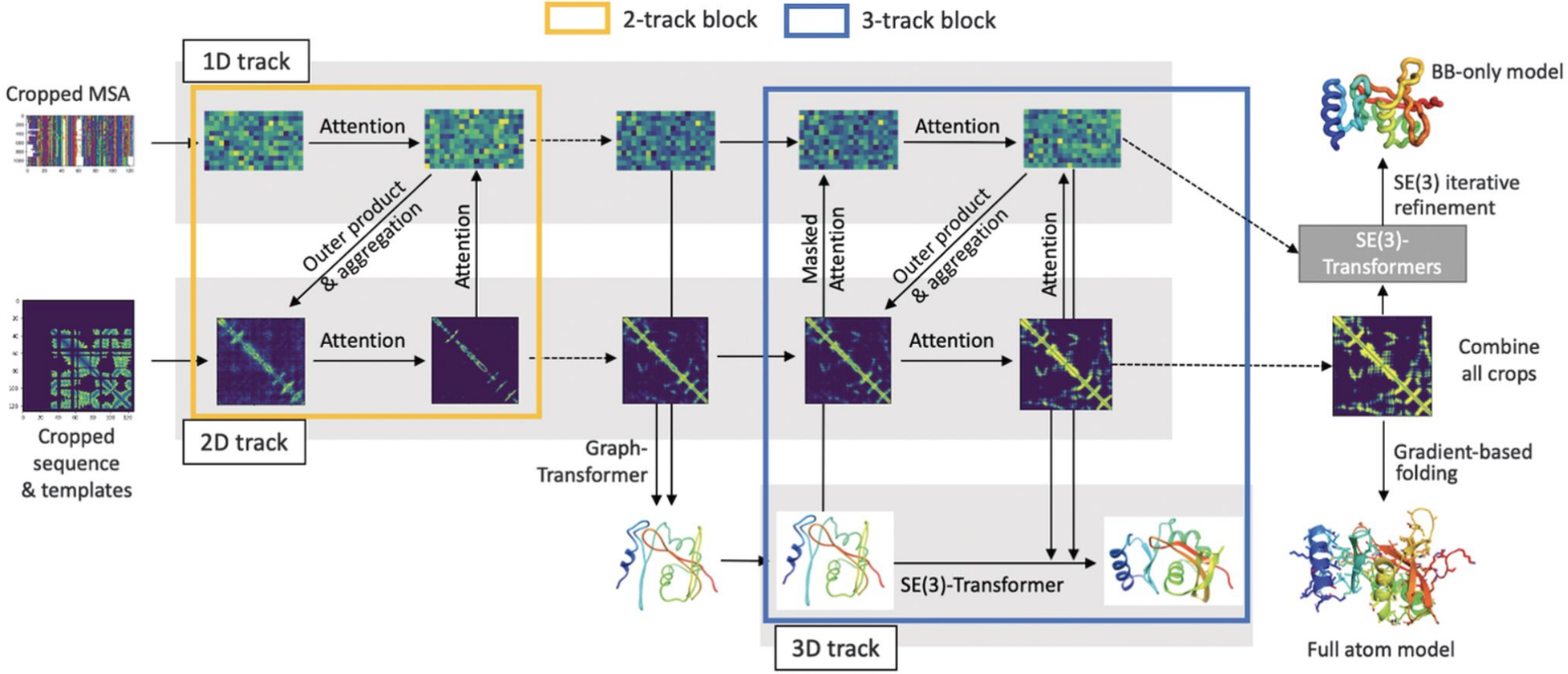
More Colab notebooks: https://github.com/sokrypton/ColabFold/

OpenFold2:

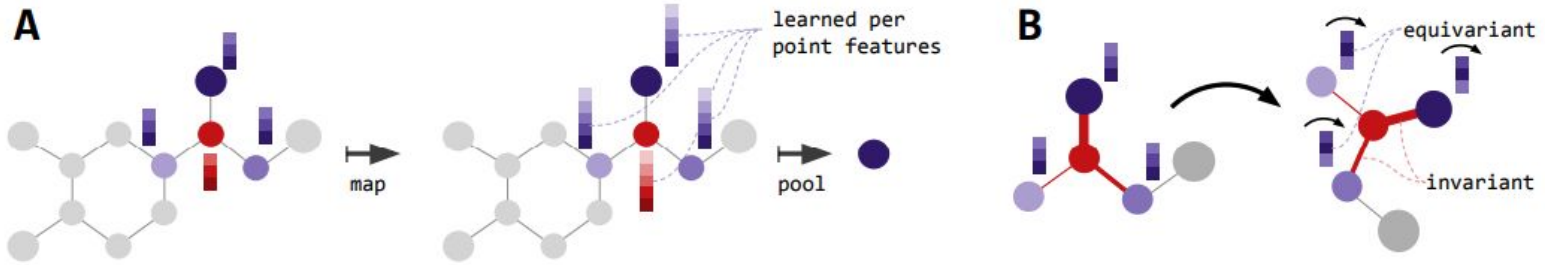
Run AlphaFold2 on Google Colab

New Paper on Science: RoseTTAFold

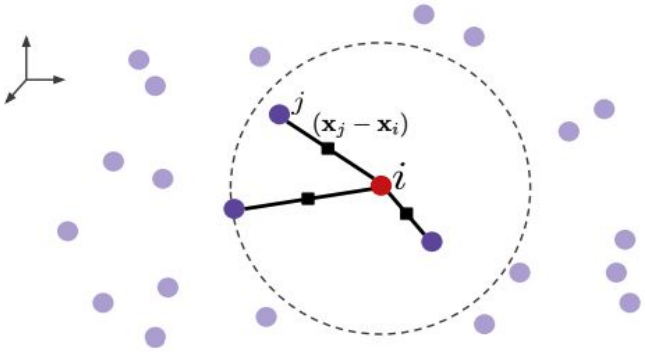
- Accurate prediction of protein structures and interactions using a three-track neural network
- Accuracy approaching closely on DeepMind's
- Claimed the model enables rapid generation of accurate protein-protein complex models



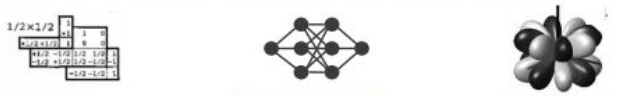
SE(3)-Transformers [Paper]



Step 1: Get nearest neighbours and relative positions



Step 2: Get SO(3)-equivariant weight matrices



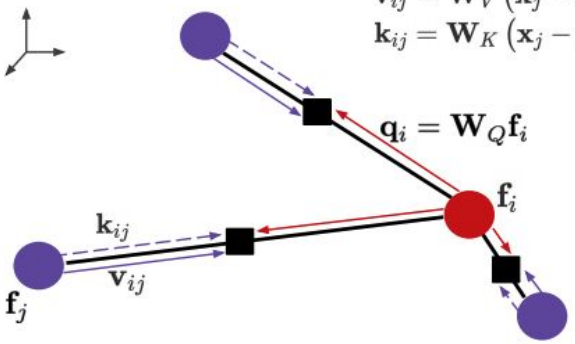
Clebsch-Gordon Coeff. $\mathbf{Q}_{Jm}^{\ell k}$ | Radial Neural Network $\varphi_J^{\ell k}(||x||)$ | Spherical Harmonics $Y_{Jm}\left(\frac{x}{||x||}\right)$

Matrix W consists of blocks mapping between degrees

$$\mathbf{W}(x) = \mathbf{W}\left(\left\{\mathbf{Q}_{Jm}^{\ell k}, \varphi_J^{\ell k}(||x||), Y_{Jm}\left(\frac{x}{||x||}\right)\right\}_{J,m,\ell,k}\right)$$

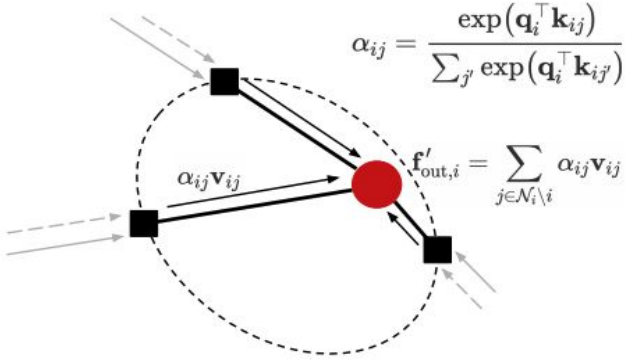
Step 3: Propagate queries, keys, and values to edges

$$\mathbf{v}_{ij} = \mathbf{W}_V\left(\mathbf{x}_j - \mathbf{x}_i\right)\mathbf{f}_j$$
$$\mathbf{k}_{ij} = \mathbf{W}_K\left(\mathbf{x}_j - \mathbf{x}_i\right)\mathbf{f}_j$$



Step 4: Compute attention and aggregate

$$\alpha_{ij} = \frac{\exp\left(\mathbf{q}_i^\top \mathbf{k}_{ij}\right)}{\sum_{j'} \exp\left(\mathbf{q}_i^\top \mathbf{k}_{ij'}\right)}$$

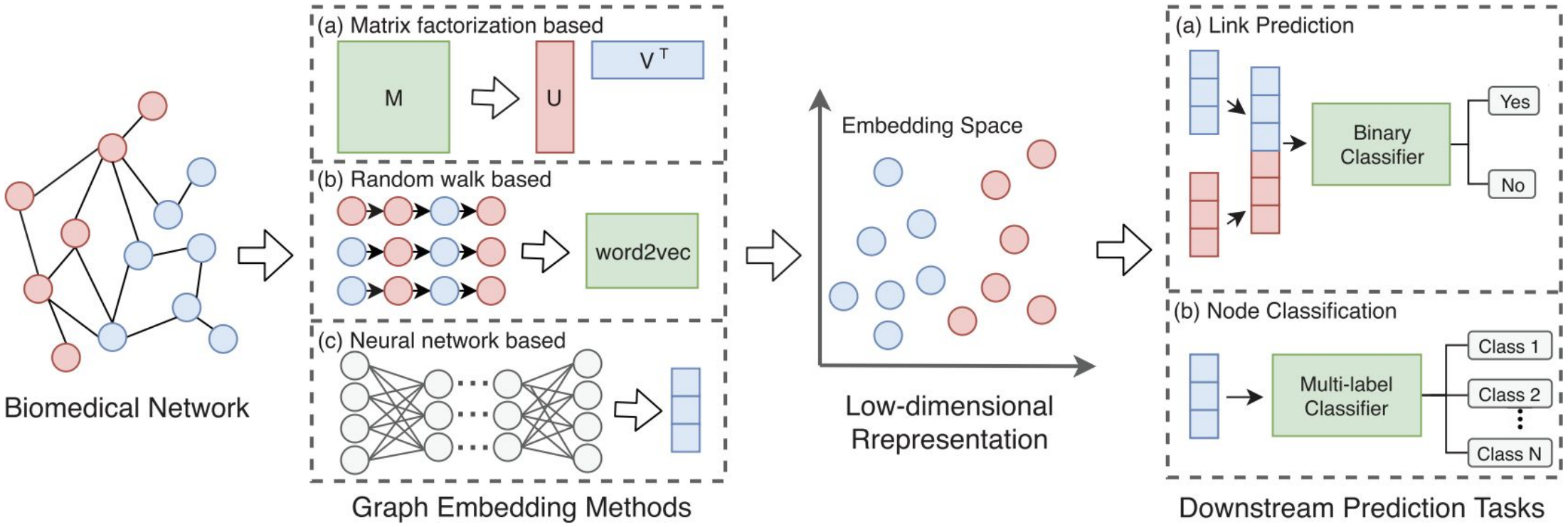


Discussion

Network Science and Graph in Bioinformatics

At the boundary between different fields, new "mountains" rise up.

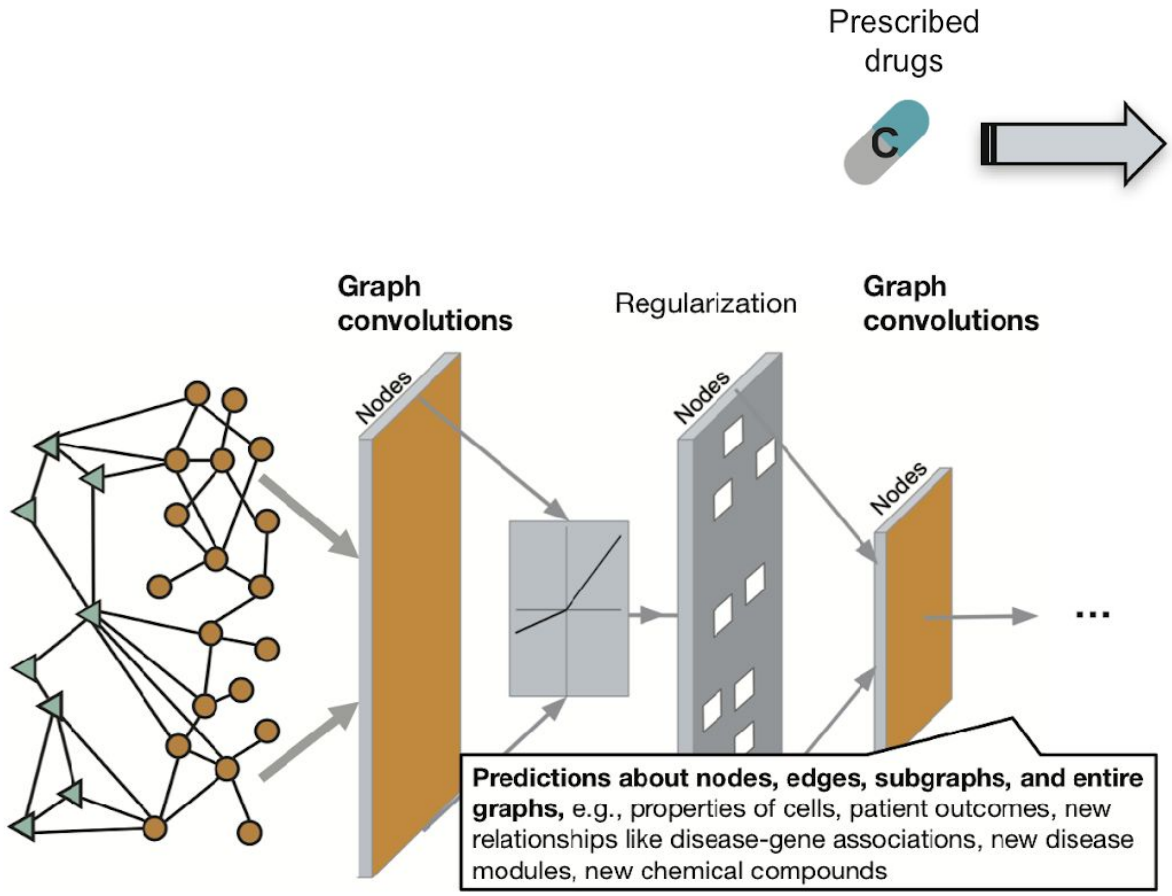
Learn Embeddings on Biological Networks

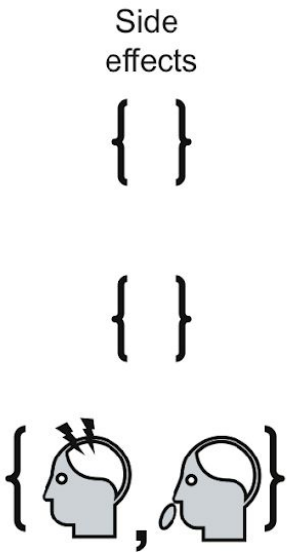


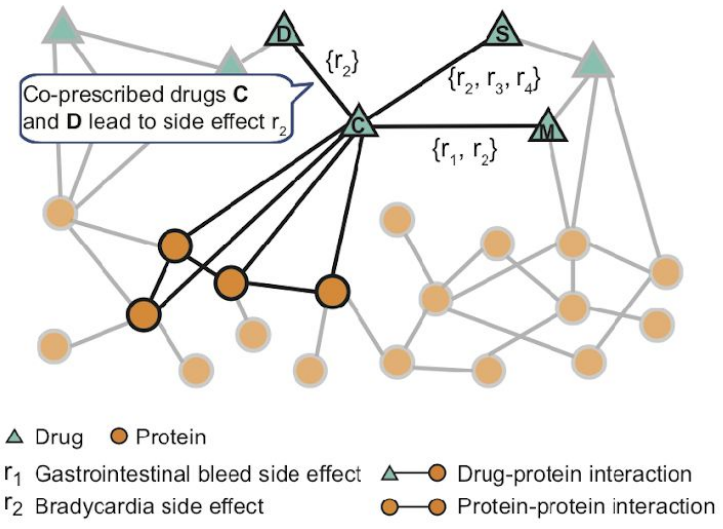
Reference: Yue, Xiang, et al. "Graph embedding on biomedical networks: methods, applications and evaluations." *Bioinformatics* 36.4 (2020): 1241-1251.

Example: Drug-Protein Network, Side Effects

Credit: https://zitniklab.hms.harvard.edu/research/







More Applications

- Molecular biology, compound structures, pathways
- Pandemic prediction, disease spreading
- Healthcare knowledge graphs, biomedical ontologies
- Clinical report analysis and personal health record

DeepMind's AlphaFold Team & Posts

- https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery (AlphaFold v1, Jan 2020)
- https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology (AlphaFold v2, Dec 2020)
- https://deepmind.com/blog/article/putting-the-power-of-alphafold-into-the-worlds-hands (AlphaFold v2 release, Jul 2021)

Resource List: AlphaFold and New Frontier of Protein Folding

Tutorials, Blogs, and Related resources of AlphaFold and AlphaFold2 (collected by Junheng)

UCLA Samueli Computer Science

Thank you!

Contact: jhao@cs.ucla.edu

Website: http://www.haojunheng.com/

Appendix

Related Topics and Tutorials

More about MSA, Protein structure and spatial representation, etc.