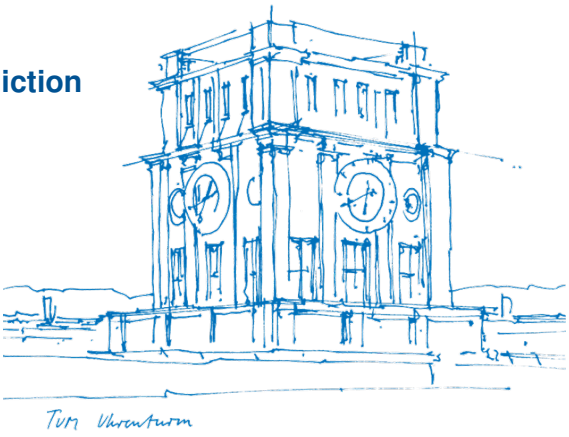
Computer Vision Group
TUM School of Computation, Information and Technology
Technical University of Munich

AlphaFold2

Highly Accurate Protein Structure Prediction



Zeyu Li

Computer Vision Group
TUM School of Computation, Information and Technology
Technical University of Munich

October 20th, 2024

Outline

Proteins
Definition

- **Protein** Complex biomolecules that perform a vast array of functions in living organisms. Comprised of long chains of amino acids, proteins fold into unique 3D shapes essential for their specific functions.
- **Amino Acid Sequence** represents the linear order of amino acids in a protein
- **Backbone** The chain of repeating atoms (N, C) that forms the main structure of the protein
- **Residues** Individual amino acids, including their unique side chains, attach to the backbone and determine the protein's properties and interactions.
- **Multiple Sequence Alignment (MSA)** A method to align sequences of homologous proteins to identify conserved regions, which often indicate structurally or functionally important areas.

The Protein Folding Problem

Importance of protein structure prediction

- Structures of around 100,000 unique proteins determined by experiments
- Only represents a small fraction of the billions of known protein sequences
- Months to years of painstaking effort required
- **Accurate computational approaches are needed**
- Protein Folding Problem

The Protein Folding Problem

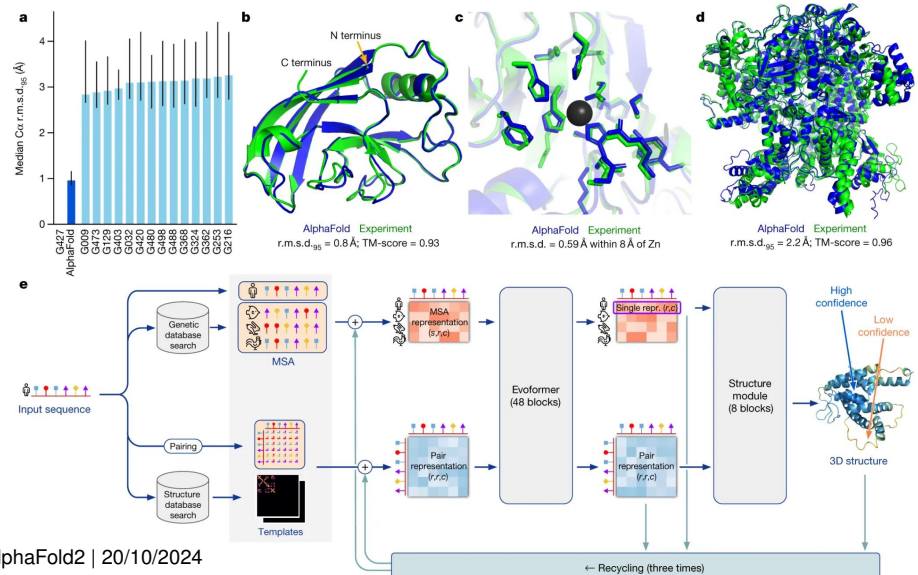
Existing methods

- Two complementary paths of predicting 3D protein structures
  - **Physical interactions**
  - proved highly challenging for even moderate-sized proteins
  - **Evolutionary history**
  - the constraints on protein structure are derived from the evolutionary history of proteins
- Despite these advances, both approaches face the same problem
- When no homologous structure available accuracy falls short

AlphaFold2's Breakthrough

- Near experimental accuracy in a majority of cases
- Median backbone accuracy of 0.96 Å r.m.s.d
- root-mean-square deviation at 95% residue coverage (95% confidence interval = 0.85–1.16 Å)
- As comparison, the width of a carbon atom is approximately 1.4 Å.
- All-atom accuracy was 1.5 Å r.m.s.d.95
- Scalable to very long proteins with accurate domains and domain-packing
- Finally, able to provide precise, per-residue estimates of its reliability
- Innovative design of architecture

AlphaFold2's Breakthrough



AlphaFold2's Breakthrough

Innovation

- Jointly processes multiple sequence alignments (MSAs) and pairwise feature
- Iteratively refine structural hypotheses
- Inspect key components that matters

AlphaFold2's Breakthrough

Key Components

- Multiple sequence alignments (MSA) and Pairwise features
- Evoformer
- Structure modules
- Recycling
- Self-distillation

Outline

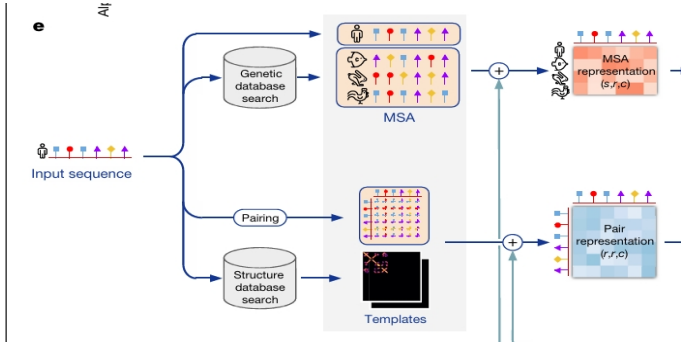
Step by Step Inspection

In general, the architecture can be divided in three parts:

- Feature Extraction
- Encoder(Evoformer)
- Decoder(Structure module)

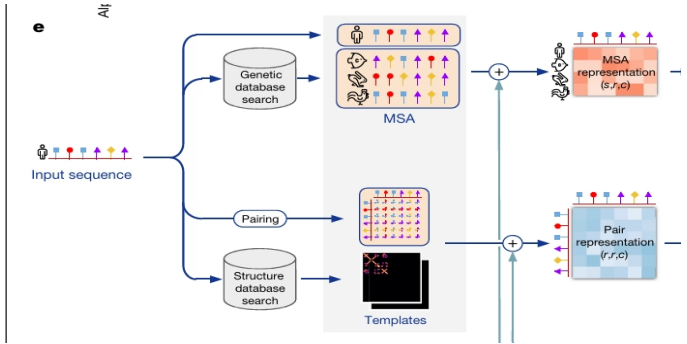
Each of which contains rich details. We will inspect step by step.

Feature Extraction



- Input sequence of human DNA is delivered simultaneously into two representations: MSA representation and Pairing representations
- Jointly embed multiple sequence alignments (MSAs) and pairwise features
- Pair residues features

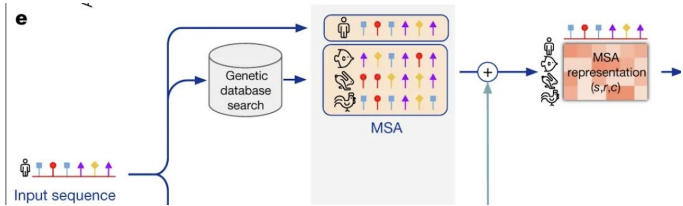
Feature Extraction
Motivation



- **Goal:** Directly predicting the 3D coordinates of all heavy atoms for a given protein
- Not only serial residues information of a given sequence
- But also relationships between different residues

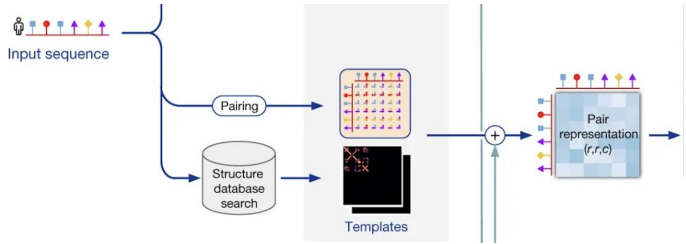
Feature Extraction

Multiple Sequence Alignments



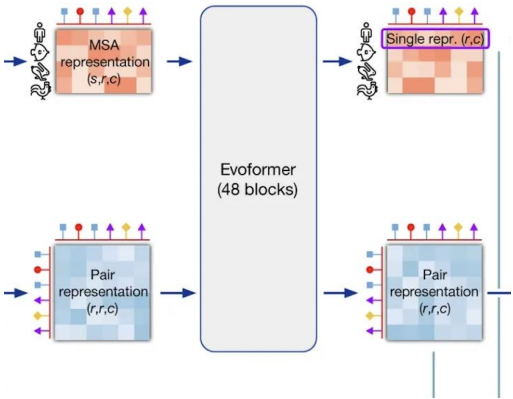
- The MSA representation: $N_{seq} \times N_{res}$ array
- N_{seq}: number of sequences
- N_{res}: number of residues
- Randomly selecting representative sequences from a full set(Genetic Database)
- Associating each sequence with its nearest representative
- “Clustering” with random cluster centres ensures similarity of sequences to some extend

Feature Extraction
Pair Residues Features



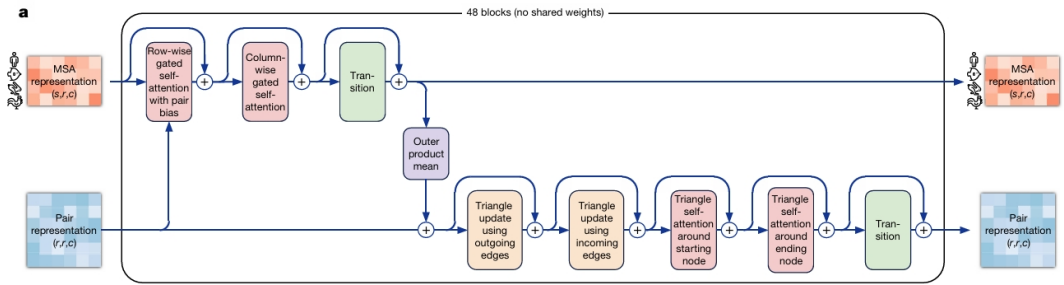
- The residue pairs representation: $N_{res} \times N_{res}$ array
- Templates store structural information about residues pairing existing

Evoformer



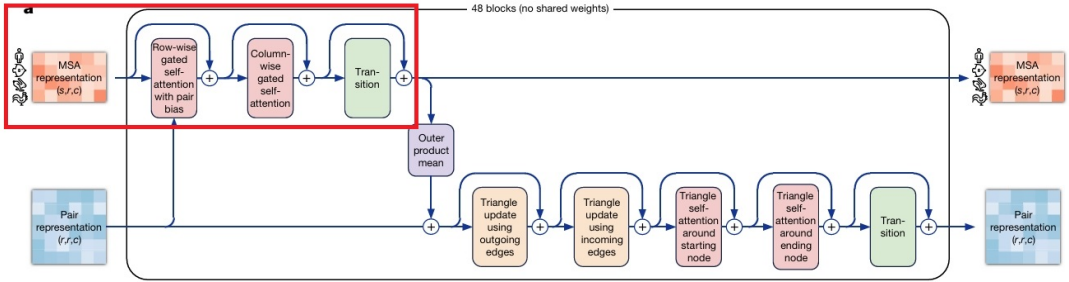
- Input and output of the same size
- Exchange information within MSA and pairs
- (s, r, c) refers to sequence, residues and channels/features, recalling **Multi-head Attention** in Transformer

Evoformer
Architecture perspective



- Upper-left sub-net for MSA sequence representation
- Bottom-right sub-net for pair representation
- Both make use of MSA and pair-wise information

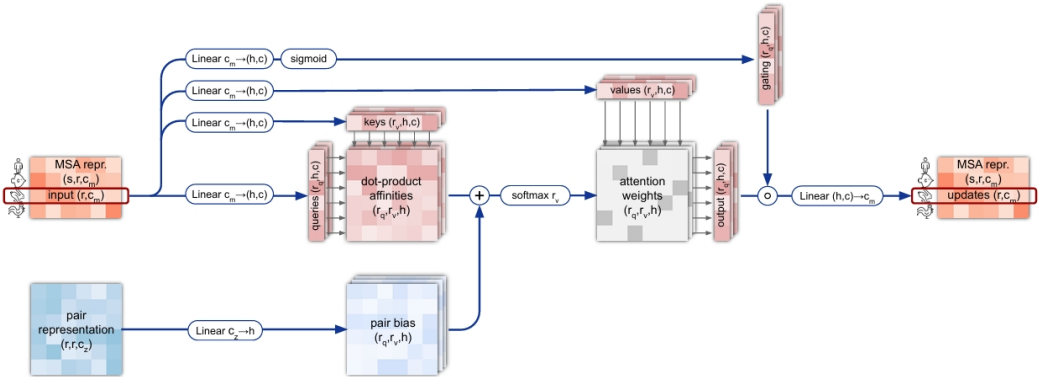
Evoformer
MSA



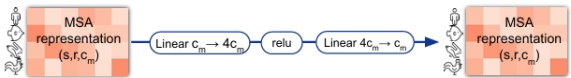
- Forward pass design
- Split into row-wise and column-wise training
- MLP transition

Evoformer

MSA row-wise gated self-attention

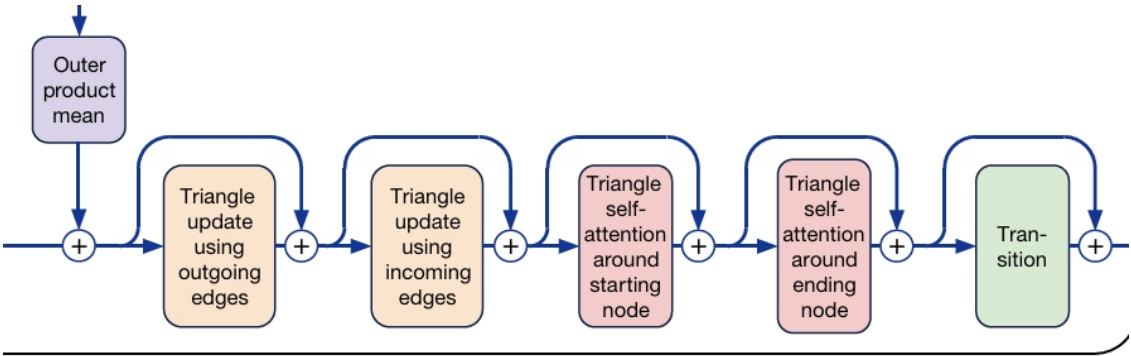


Evoformer
MSA Transition



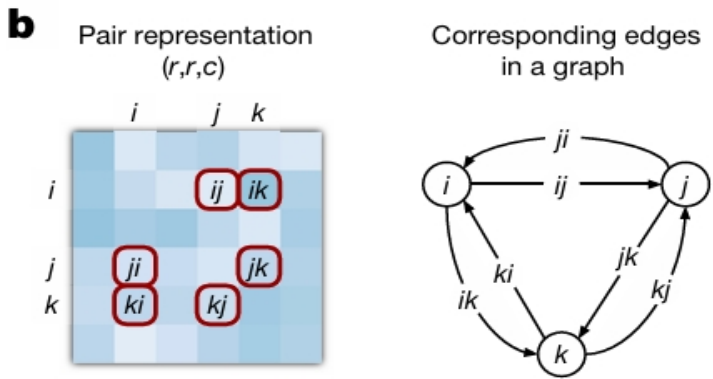
- 2-layer MLP as the transition layer
- Intermediate number of channels expands the channels by a factor of 4

Pairwise Representation



- Goal: Amino acids be pairwise representable in as a 3D structure
- Many constraints must be satisfied: e.g. the triangle inequality on distances
- Pairwise Representation must incorporate these constraints

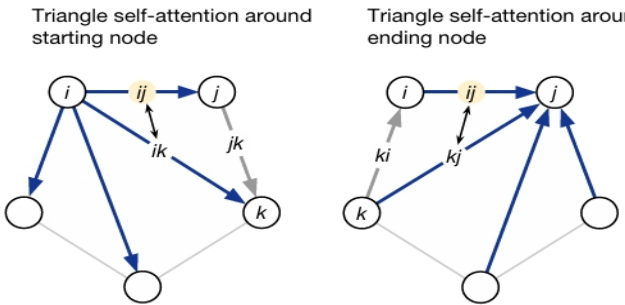
Pairwise Representation
Illustration of Pair in Graph



- Recall GNN, represented as weighted vectors

Pairwise Representation

Triangle self-attention around nodes



- Recall GNN's message passing
- In particular, given i, j, k three nodes, update the information of node i and node j to node k(starting node), or vice versa(ending node)

Pairwise Representation

Triangle self-attention in algorithm

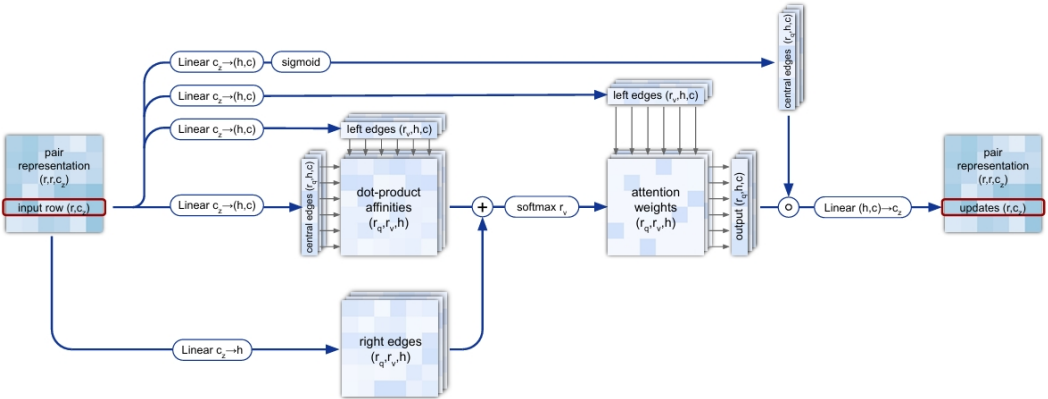
$$a_{ijk}^h = \text{softmax}_k \left(\frac{1}{\sqrt{c}} \mathbf{q}_{ij}^{h\top} \mathbf{k}_{ik}^h + b_{jk}^h \right)$$

$$o_{ij}^h = g_{ij}^h \circ \sum_k a_{ijk}^h \mathbf{v}_{ik}^h$$

- Inspect self-attention around starting node
- Element-wise Multiplication with gate
- A Combination of attention and message passing
- g_{ij}^h represents the output gating.
- Overall, very similar to MSA attention we've seen earlier.

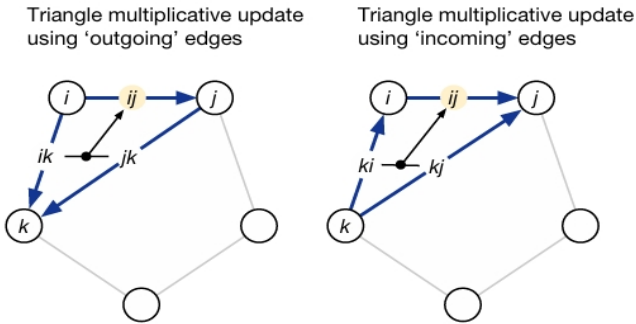
Pairwise Representation

Triangle self-attention around starting node



Pairwise Representation

Triangle self-attention using edges



- Similar as before, update the missing third edge using two edges adjacent
- Given edges ik, jk, update edge ij using attention

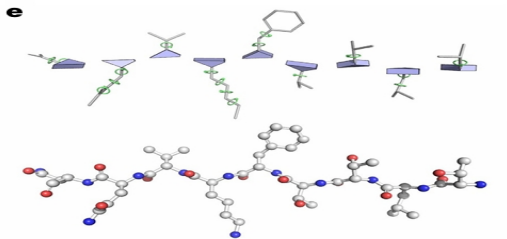
Pairwise Representation

Triangle self-attention using edges

$$g_{ij} = \sigma(\text{Linear}(\mathbf{z}_{ij}))$$
$$\tilde{\mathbf{z}}_{ij} = g_{ij} \circ \text{Linear}(\text{LayerNorm}(\sum_k a_{ik} \circ \mathbf{b}_{jk}))$$

- g_{ij}^h represents the output gating.
- Element-wise Multiplication with each other and gate
- Can be regarded as message passing across edges
- Initially designed to simplified the computation
- In the end proved to be indispensable.
- Make the prediction more accurate

Decoder
Design Principle

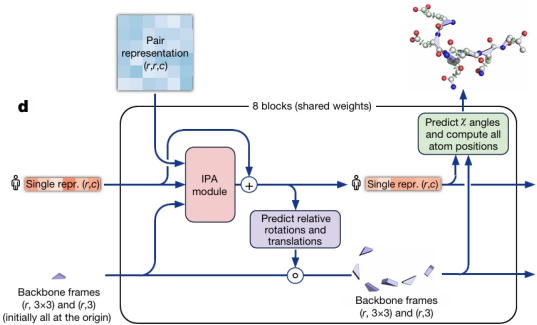


- To represent a protein structure in 3D space, what problem do we need to solve?
- Using the pair representation and the original sequence row (single representation)
- First we need 3D backbone structure of the MSA representation
- N_{res} independent rotations and translations, each with respect to the global frame

Decoder
Design Principle

- Aside from pure geometry, exact biological constraints are enforced in the post-prediction relaxation of the structure
- How?
- Designing context-specified violation loss term
- e.g. Satisfaction of the peptide bond geometry is encouraged during fine-tuning by a violation loss term
- To simplify the representations, coordinates are separated int o two frame: Global Frame and Local Frame
- Local Frame: Translations and Rotations suffice. Or in the words of paper, “rigid motions")
- Why?

Decoder
Predicting the structure



- Recall the structure of LSTM
- Explicit building of backbone pipe (analog to cell states in LSTM)
- Key Component: IPA modules

Decoder
Invariant point attention (IPA)

$$a_{ij}^h = \text{softmax}_j \left(w_L \left(\frac{1}{\sqrt{c}} \mathbf{q}_i^{h\top} \mathbf{k}_j^h + b_{ij}^h - \frac{\gamma^h w_c}{2} \sum_p \left\| T_i \circ \vec{q}_i^{hp} - T_j \circ \vec{k}_j^{hp} \right\|^2 \right) \right)$$

- i, j represents two different residues
- T_i, T_j "rigid motions" of corresponding residues
- w_c empirical constants w.r.t. query points, and γ^h another constant w.r.t heads

IPA v.s. Regularization

- We compare the attention equation with regularization terms to show their similarities
- $R(f) = \Omega(f) + C \sum_{i=1}^{m} \ell(f(x_i), y_i)$
- $\Omega(f)$ structural risk: designed to append the character we want the model to have
- Here, Affinity computation term
- The farther two residues lie after rigid transformations, the less likely they are similar to each other
- To some extend, very similar to the concept of reducing overfitting

IPA
Proof of Invariance

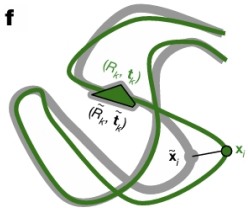
L2-norm of a vector is invariant under rigid transformations

$$
\begin{aligned}
&\left\| (T_{\text{global}} \circ T_i) \circ \vec{q}_i^{hp} - (T_{\text{global}} \circ T_j) \circ \vec{k}_j^{hp} \right\|^2 \\
&= \left\| T_{\text{global}} \circ \left(T_i \circ \vec{q}_i^{hp} - T_j \circ \vec{k}_j^{hp} \right) \right\|^2 \\
&= \left\| T_i \circ \vec{q}_i^{hp} - T_j \circ \vec{k}_j^{hp} \right\|^2 .
\end{aligned}
$$

FAPE Loss

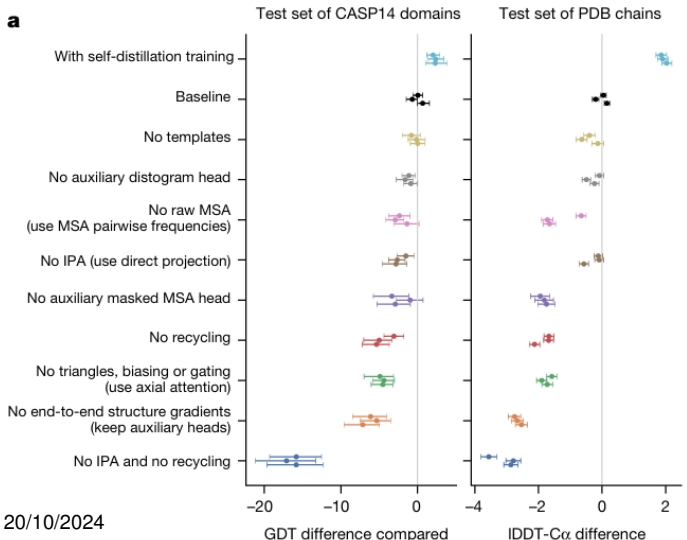
Frame aligned point error

- Compares the predicted atom positions to the true positions under many different alignments
- Compute the distance of all predicted atom positions x_i from the true atom positions
- Penalized with a clamped L1 loss
- A strong bias for atoms to be correct relative to the local frame of each residue and hence correct with respect to its side-chain interactions



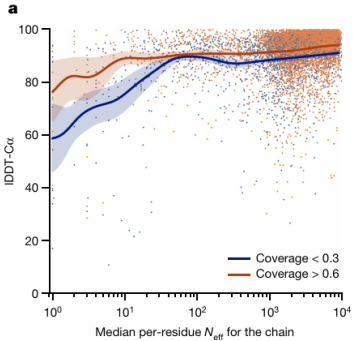
Outline

Interpreting neural network
Ablation results



Interpreting neural network

Effect of MSA information and cross-chain contacts



X-axis represents the median per-residue N_{eff} on logarithmic scale.
Y-axis represents the IDDT-C$_\alpha$ $score$
Shaded areas around each curve represent the 95% confidence intervals

Outline

Discussion

- Rely highly on MSA information
- High-accuracy predictions for homomers, not good enough for hetero-complexes
- Focus on the paper's wise modification of normal transformer
- Still many details not covered in this slides, e.g. data augmentation with noisy student self-distillation
- Seek for *Supplmentary Information* of the paper for implementation details if interested.
- Most of the stuff **based on my own understanding**, do not hesitate to point out!

Q & A

Feel free to ask any questions!

Thanks for your attention!