

ALPHAFOLD 2

Viola Chen, Xiaxin Shen





OUTLINE



Motivation

Overview

Evoformer

Structure Module

Interpretability



OUTLINE



Motivation

Overview

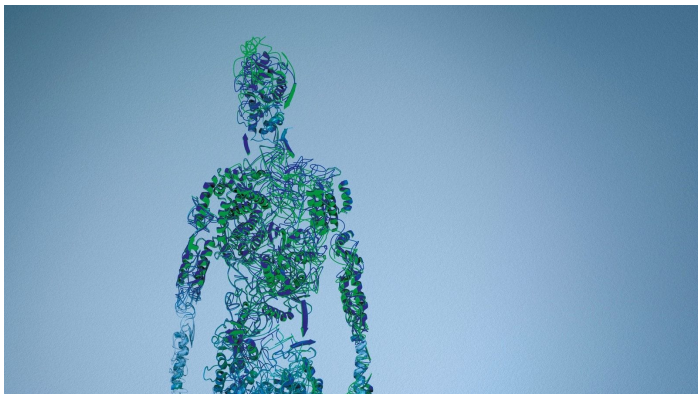
Evoformer

Structure Module

Interpretability

MOTIVATION

- **Physical** interactions
 - Computational intractability
 - Context dependence
 - Not sufficiently accurate
- **Evolutionary** history
 - Bioinformatics analysis
 - Far short of experimental accuracy



- Without AI, it takes **hundreds of thousands of dollars** to determine protein 3D structures
 - Costs **\$100,000~1M** per structure
 - Determination of 3D structure by X-ray crystallography takes as much as **three to five** years
- AlphaFold2
 - **Near experimental accuracy** in a majority of cases
 - Even in cases in which **no similar structure is known**.



OUTLINE



Motivation

Overview

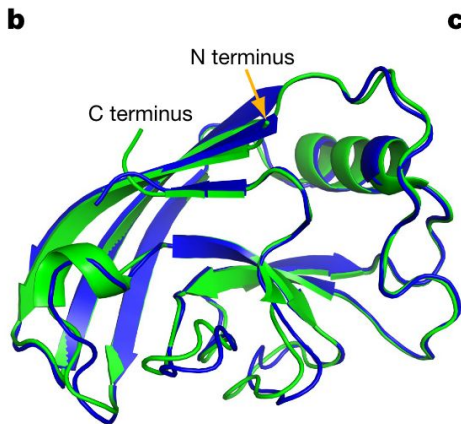
Evoformer

Structure Module

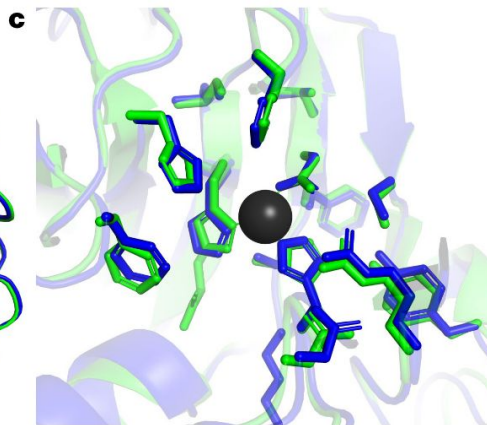
Interpretability

THE PERFORMANCE OF ALPHAFOLD ON THE CASP14 DATASET

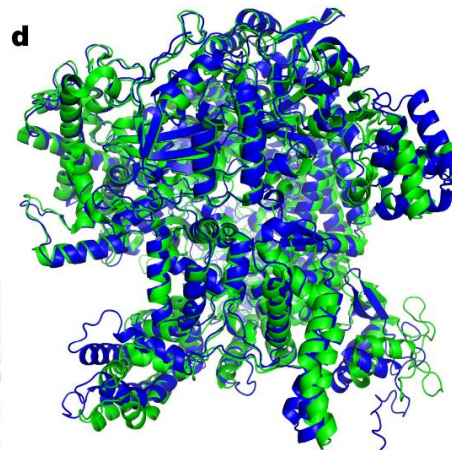
- AlphaFold2 can produce
 - very accurate **domain structures**
 - highly accurate **side chains**
 - **scalable** to very **long** proteins



AlphaFold Experiment
r.m.s.d.₉₅ = 0.8 Å; TM-score = 0.93

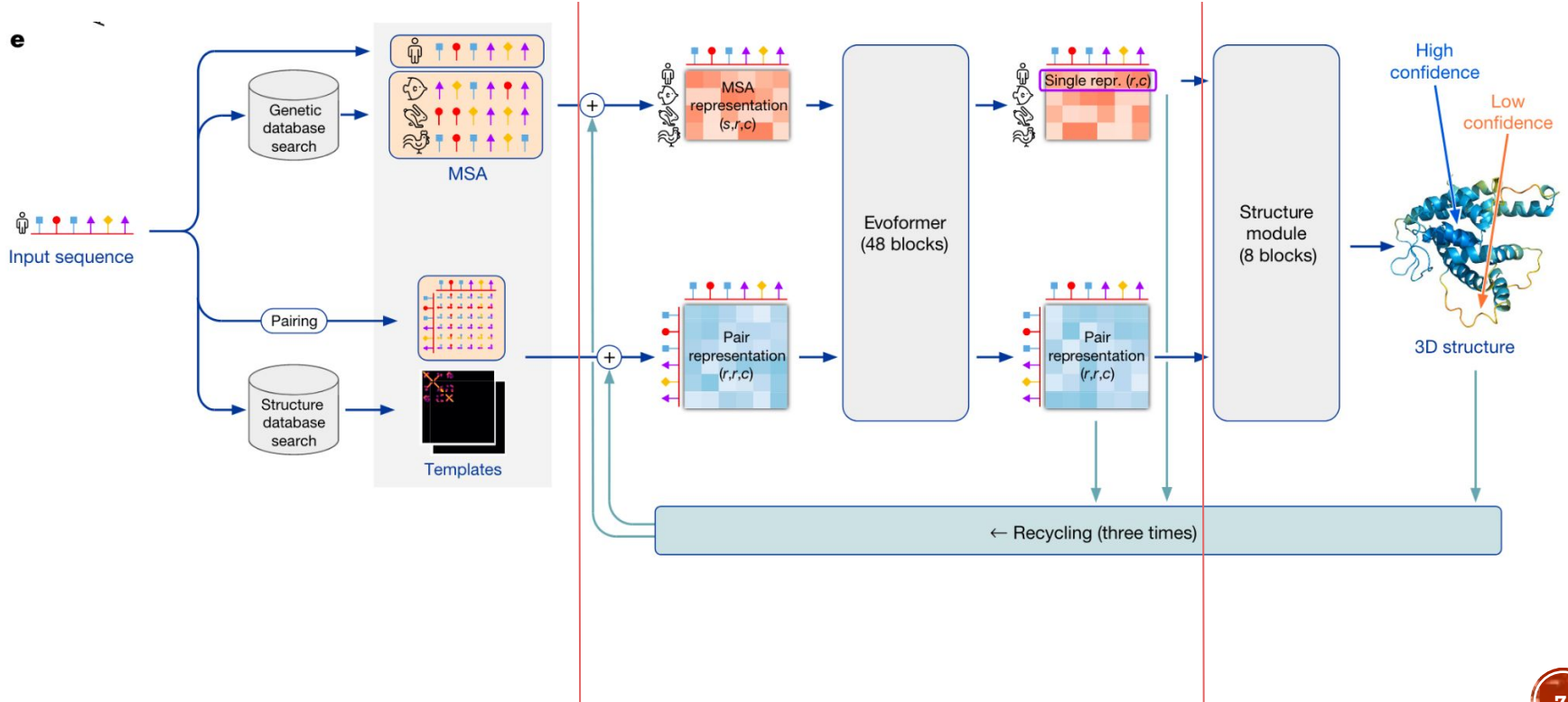


AlphaFold Experiment
r.m.s.d. = 0.59 Å within 8 Å of Zn



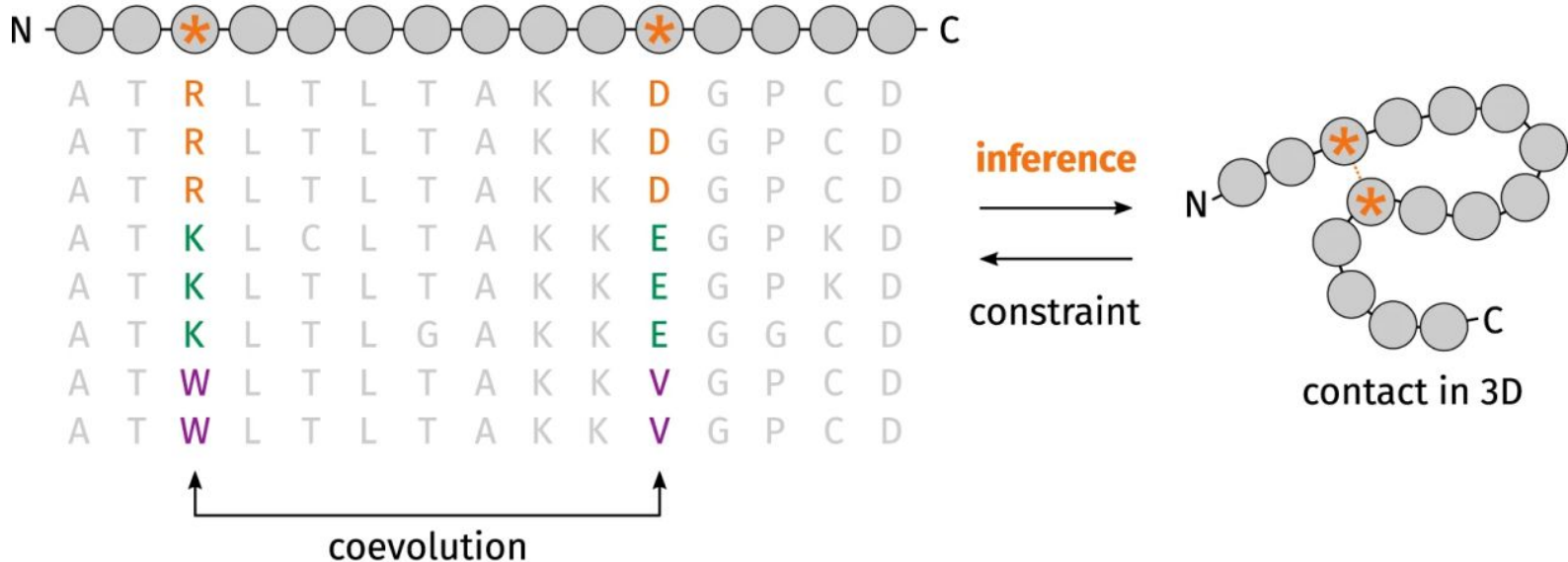
AlphaFold Experiment
r.m.s.d.₉₅ = 2.2 Å; TM-score = 0.96

THE ALPHAFOLD NETWORK



MULTIPLE SEQUENCE ALIGNMENT (MSA)

- If **two amino acids** are in **close contact**, mutations in one of them will be **closely followed** by mutations of the other, in order to preserve the structure.





OUTLINE



Motivation

Overview

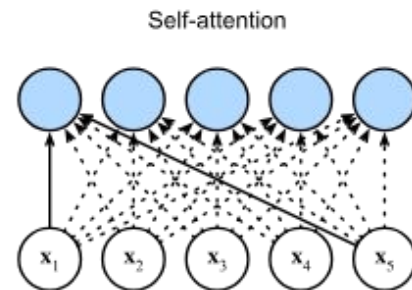
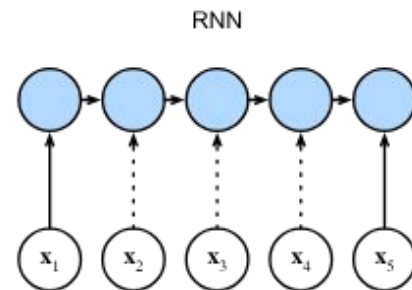
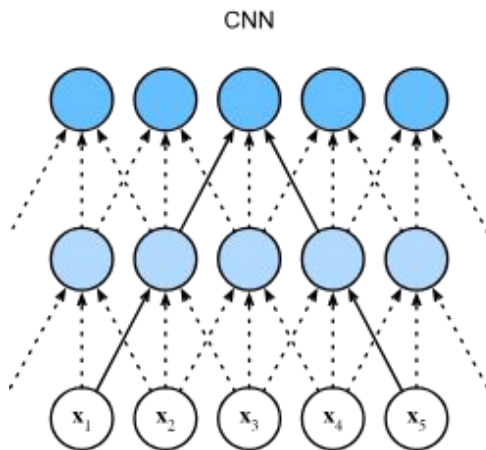
Evoformer

Structure Module

Interpretability

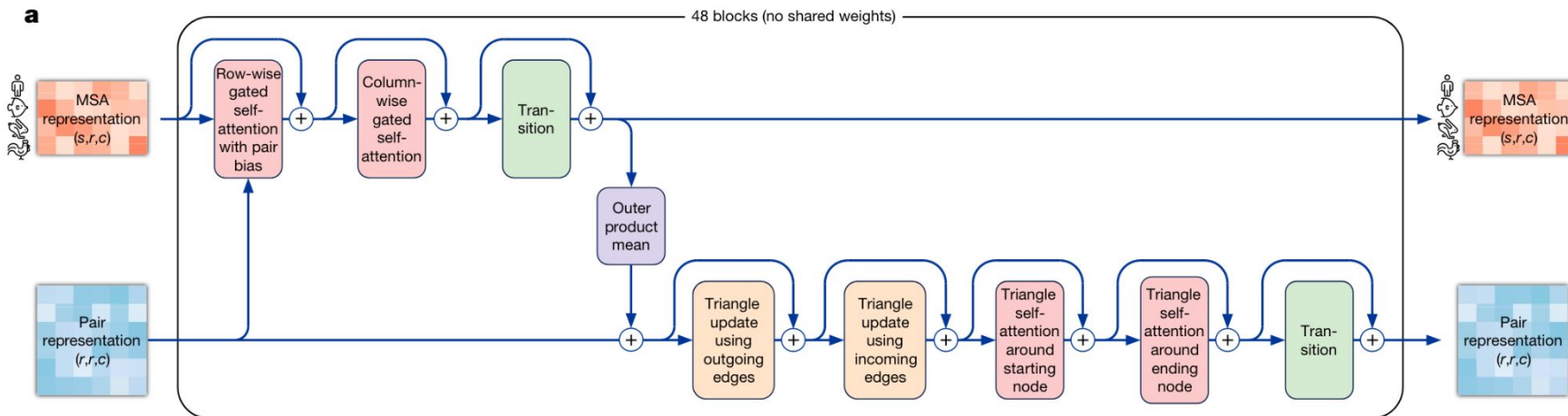
TRANSFORMER

- An encoder-decoder model that can manipulate pairwise connections within and between sequences.
- Transformers are great with sequences that have pairwise connections.



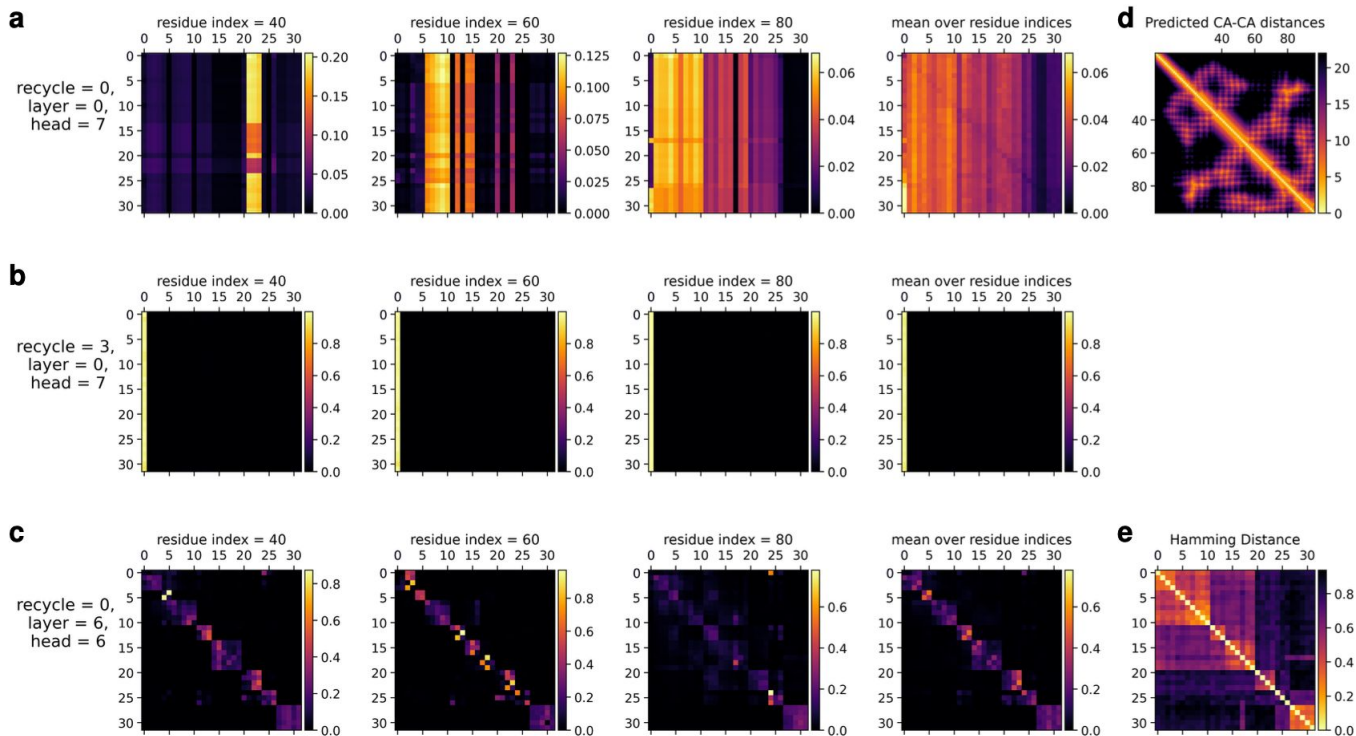
EVOFORMER BLOCK

- Updates the MSA with axial attention, using the info from pair representation
- Updates the pair representation from updated MSA, using outer product mean block
- Applies triangle inequalities to the updated pair representation to enforce consistency

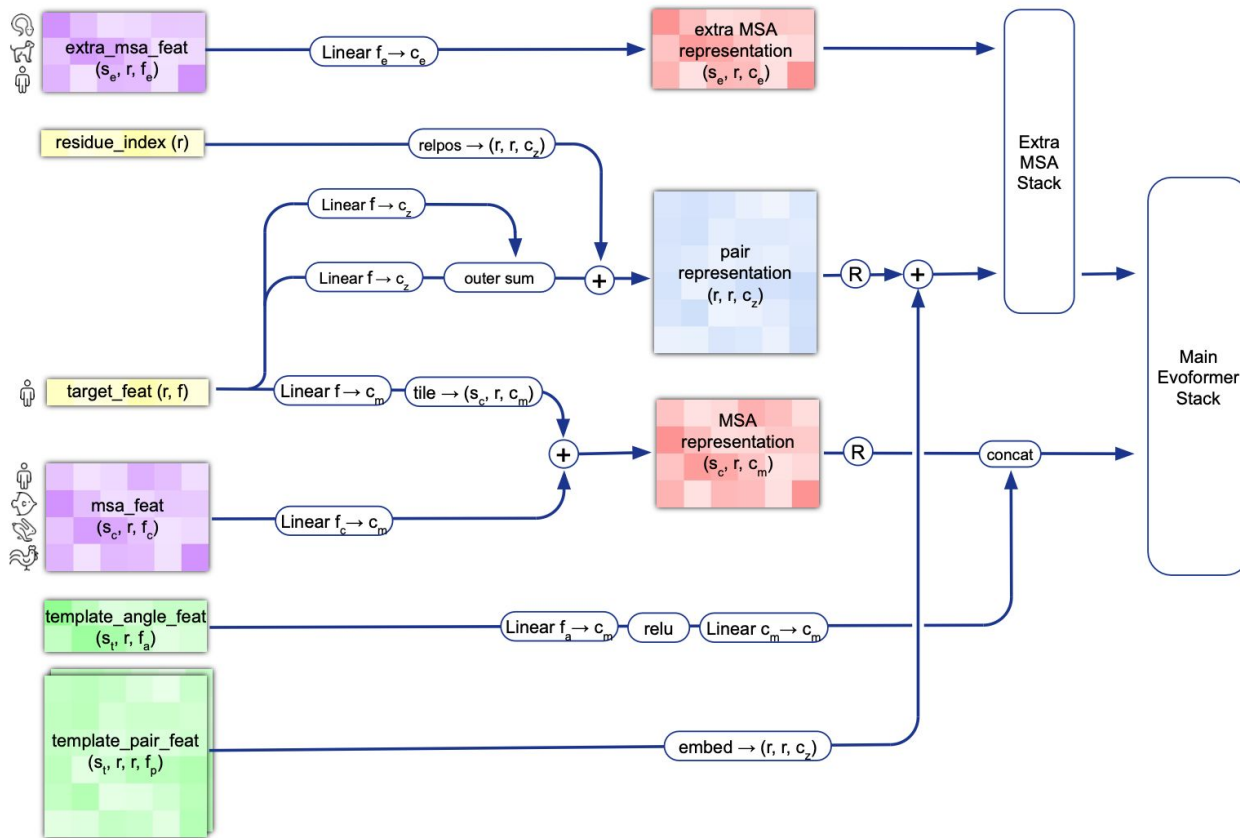


VISUALIZATIONS OF THE ATTENTION MAPS

MSA Columnwise Attention: CASP14 target T1082

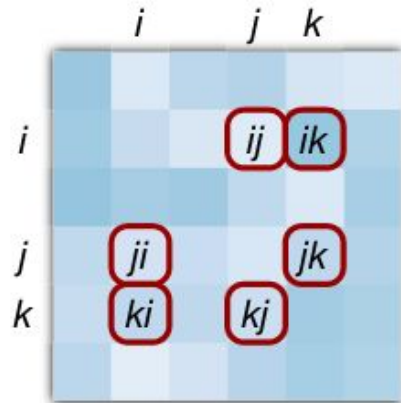


INPUT FEATURE EMBEDDINGS

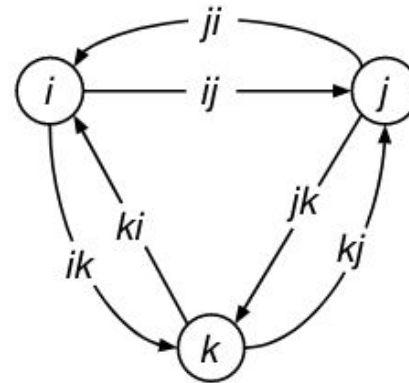


PAIR REPRESENTATION

b Pair representation
(r, r, c)



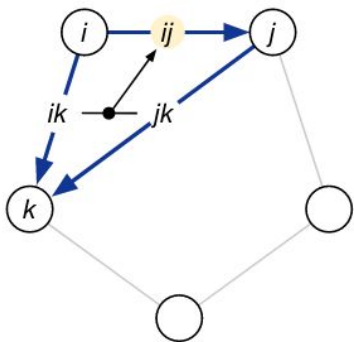
Corresponding edges
in a graph



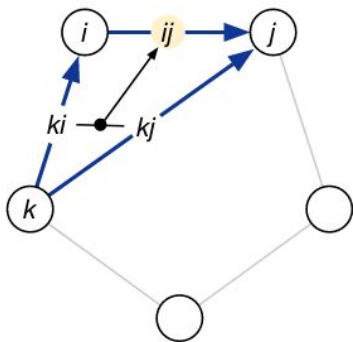
TRIANGLE MULTIPLICATIVE UPDATE AND TRIANGLE SELF-ATTENTION.

c

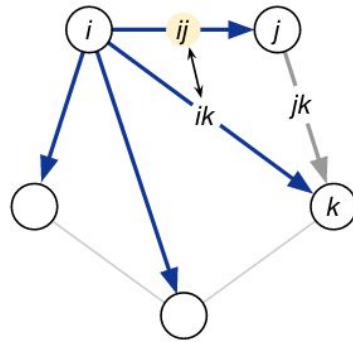
Triangle multiplicative update using 'outgoing' edges



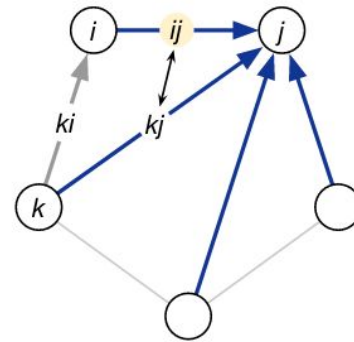
Triangle multiplicative update using 'incoming' edges



Triangle self-attention around starting node



Triangle self-attention around ending node





OUTLINE



Motivation

Overview

Evoformer

Structure Module

Interpretability

STRUCTURE MODULE

How does AlphaFold2 “translates” output of Evoformer to 3D coordinates of atoms?



Input:

Evoformer's single representation

$$\{\mathbf{s}_i^{\text{initial}}\} \text{ with } \mathbf{s}_i^{\text{initial}} \in \mathbb{R}^{c_s}$$

Evoformer's pair representation

$$\{\mathbf{z}_{ij}\} \text{ with } \mathbf{z}_{ij} \in \mathbb{R}^{c_z} \text{ and } i, j \in \{1, \dots, N_{\text{res}}\}$$

Backbone frames

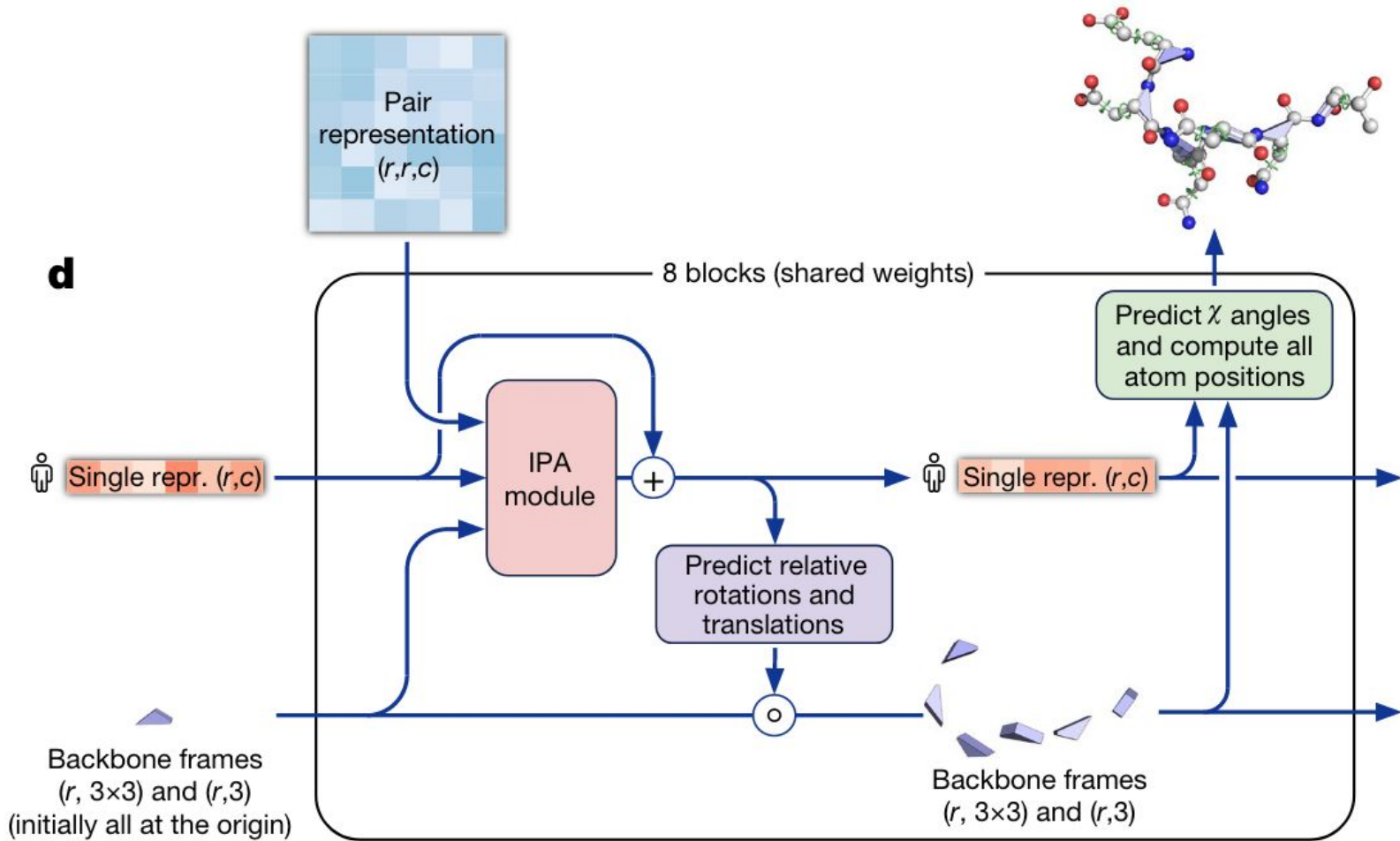
Rotations (r , 3×3) and translation vector (r , 3)

Output:

A global 3D coordinate for each residue

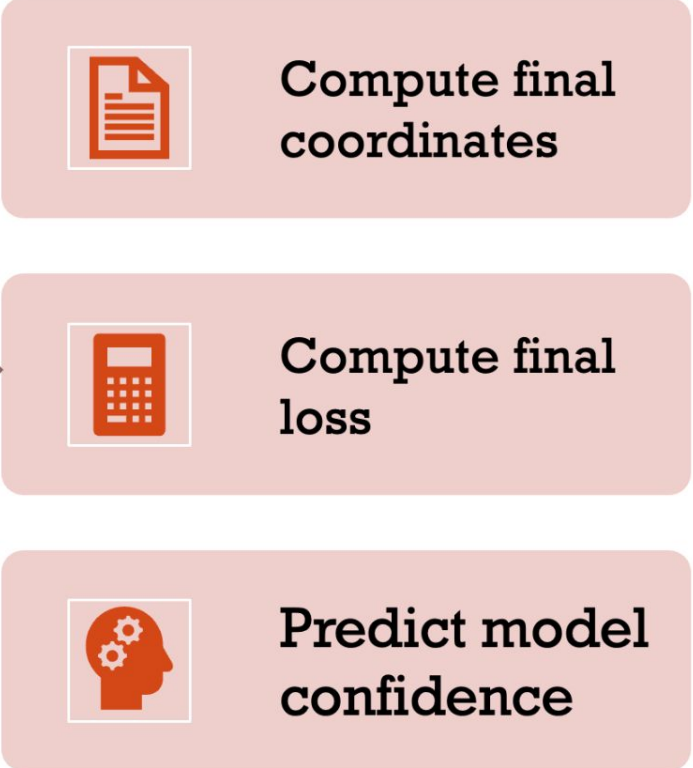
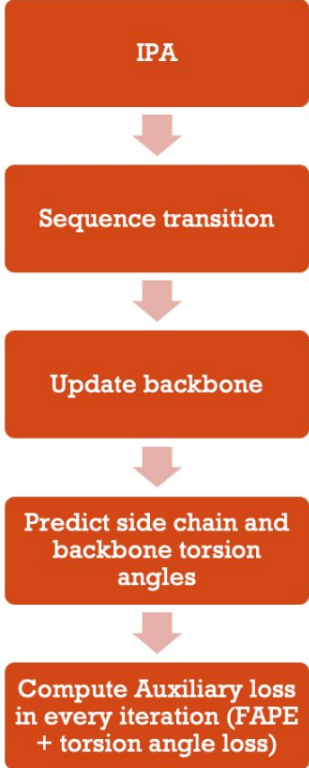
$$\vec{\mathbf{x}}_{\text{global}} \in \mathbb{R}^3$$

**WHAT ARE THE *INPUT* AND
OUTPUT OF THE *STRUCTURE*
*MODULE***



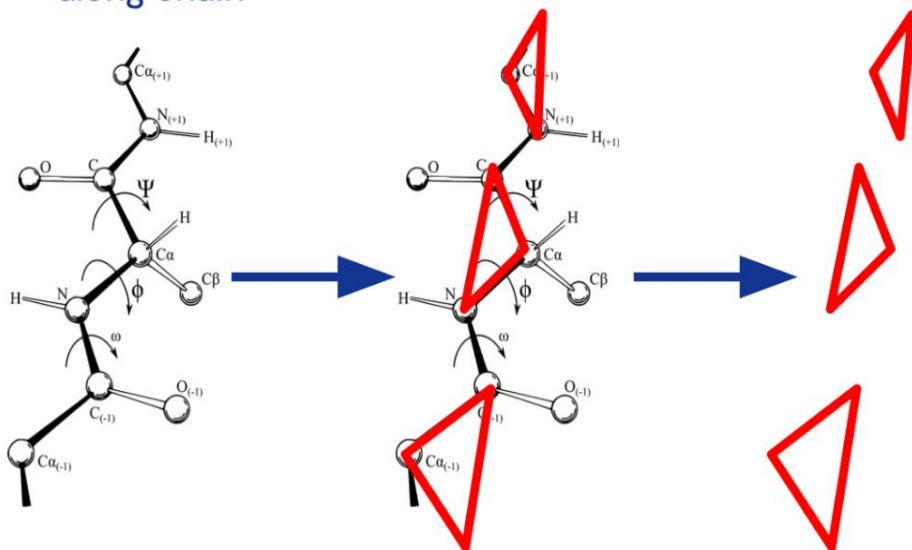
Pseudo-algorithm

- For each layer



Local Frames

- Protein backbone = gas of 3-D rigid bodies (chain is learned!)
- Rigid body orientation defines local coordinate frame along chain



- Coordinates measured in local Frame are **invariant to choice of global frame**

$$\begin{aligned}\vec{x}_{\text{global}} &= T_i \circ \vec{x}_{\text{local}} \\ &= R_i \vec{x}_{\text{local}} + \vec{t}_i.\end{aligned}$$

```

def StructureModule( $\{\mathbf{s}_i^{\text{initial}}\}, \{\mathbf{z}_{ij}\}, N_{\text{layer}} = 8, c = 128,$ 
 $\{T_i^{\text{true},f}\}, \{T_i^{\text{alt},f}\}, \{\bar{\alpha}_i^{\text{true},f}\}, \{\bar{\alpha}_i^{\text{alt},f}\}, \{\bar{\mathbf{x}}_i^{\text{true},a}\}, \{\bar{\mathbf{x}}_i^{\text{alt},a}\}$ ) :
 $\mathbf{s}_i^{\text{initial}} \in \mathbb{R}^c$ 
 $\mathbf{I} \in \mathbb{R}^{3 \times 3}, \bar{\mathbf{O}} \in \mathbb{R}^3$ 
1:  $\mathbf{s}_i^{\text{initial}} \leftarrow \text{LayerNorm}(\mathbf{s}_i^{\text{initial}})$ 
2:  $\mathbf{z}_{ij} \leftarrow \text{LayerNorm}(\mathbf{z}_{ij})$ 
3:  $\mathbf{s}_i = \text{Linear}(\mathbf{s}_i^{\text{initial}})$ 
 $\mathbf{s}_i \in \mathbb{R}^{c_a}$ 
4:  $T_i = (\mathbf{I}, \bar{\mathbf{O}})$ 
5: for all  $l \in [1, \dots, N_{\text{layer}}]$  do # shared weights
6:  $\{\mathbf{s}_i\} += \text{InvariantPointAttention}(\{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{T_i\})$ 
7:  $\mathbf{s}_i \leftarrow \text{LayerNorm}(\text{Dropout}_{0.1}(\mathbf{s}_i))$ 
# Transition
8:  $\mathbf{s}_i \leftarrow \mathbf{s}_i + \text{Linear}(\text{relu}(\text{Linear}(\text{relu}(\text{Linear}(\mathbf{s}_i))))))$  all intermediate activations  $\in \mathbb{R}^{c_a}$ 
9:  $\mathbf{s}_i \leftarrow \text{LayerNorm}(\text{Dropout}_{0.1}(\mathbf{s}_i))$ 
# Update backbone.
10:  $T_i \leftarrow T_i \circ \text{BackboneUpdate}(\mathbf{s}_i)$ 
# Predict side chain and backbone torsion angles  $\omega, \phi, \psi, \chi_1, \chi_2, \chi_3, \chi_4$ 
11:  $\mathbf{a}_i = \text{Linear}(\mathbf{s}_i) + \text{Linear}(\mathbf{s}_i^{\text{initial}})$   $\mathbf{a}_i \in \mathbb{R}^c$ 
12:  $\mathbf{a}_i \leftarrow \mathbf{a}_i + \text{Linear}(\text{relu}(\text{Linear}(\text{relu}(\mathbf{a}_i))))$  all intermediate activations  $\in \mathbb{R}^c$ 
13:  $\mathbf{a}_i \leftarrow \mathbf{a}_i + \text{Linear}(\text{relu}(\text{Linear}(\text{relu}(\mathbf{a}_i))))$  all intermediate activations  $\in \mathbb{R}^c$ 
14:  $\bar{\alpha}_i^f = \text{Linear}(\text{relu}(\mathbf{a}_i))$   $\bar{\alpha}_i^f \in \mathbb{R}^2, f \in \mathcal{S}_{\text{torsion names}}$ 
# Auxiliary losses in every iteration.
15:  $(\vec{R}_i, \vec{t}_i) = T_i$ 
16:  $\vec{\mathbf{x}}_i^{C\alpha} = \vec{t}_i$ 
17:  $\mathcal{L}_{\text{aux}}^l = \left( \text{computeFAPE}(\{T_i\}, \{\bar{\mathbf{x}}_i^{C\alpha}\}, \{T_i^{\text{true}}\}, \{\bar{\mathbf{x}}_i^{\text{true},C\alpha}\}, \epsilon = 10^{-12}) \right.$ 
18:  $\quad \left. + \text{torsionAngleLoss}(\{\bar{\alpha}_i^f\}, \{\bar{\alpha}_i^{\text{true},f}\}, \{\bar{\alpha}_i^{\text{alt},f}\}) \right)$ 
# No rotation gradients between iterations to stabilize training.
19: if  $l < N_{\text{layer}}$  then
20:  $T_i \leftarrow (\text{stopgrad}(\vec{R}_i), \vec{t}_i)$ 
21: end if
22: end for
23:  $\mathcal{L}_{\text{aux}} = \text{mean}_l(\{\mathcal{L}_{\text{aux}}^l\})$ 
24:  $T_i^f, \vec{\mathbf{x}}_i^a = \text{computeAllAtomCoordinates}(T_i, \bar{\alpha}_i^f)$   $a \in \mathcal{S}_{\text{atom names}}$ 
25:  $T_i^f \leftarrow \text{concat}(T_i, T_i^f)$ 
# Final loss on all atom coordinates.
26:  $\{T_i^{\text{true},f}\}, \{\bar{\mathbf{x}}_i^{\text{true},a}\} \leftarrow \text{renameSymmetricGroundTruthAtoms}(\{T_i^f\}, \{\bar{\mathbf{x}}_i^a\}, \{T_i^{\text{true},f}\}, \{T_i^{\text{alt},f}\}, \{\bar{\mathbf{x}}_i^{\text{true},a}\}, \{\bar{\mathbf{x}}_i^{\text{alt},a}\})$ 
28:  $\mathcal{L}_{\text{FAPE}} = \text{computeFAPE}(\{T_i^f\}, \{\bar{\mathbf{x}}_i^a\}, \{T_i^{\text{true},f}\}, \{\bar{\mathbf{x}}_i^{\text{true},a}\}, \epsilon = 10^{-4})$ 
# Predict model confidence.
29:  $\{r_i^{\text{true LDDT}}\} = \text{perResidueLDDT\_Ca}(\{\bar{\mathbf{x}}_i^a\}, \{\bar{\mathbf{x}}_i^{\text{true},a}\})$ 
30:  $\{r_i^{\text{pLDDT}}\}, \mathcal{L}_{\text{conf}} = \text{predictPerResidueLDDT\_Ca}(\{\mathbf{s}_i\}, \{r_i^{\text{true LDDT}}\})$ 
31: return  $\{\bar{\mathbf{x}}_i^a\}, \{r_i^{\text{pLDDT}}\}, \mathcal{L}_{\text{FAPE}}, \mathcal{L}_{\text{conf}}, \mathcal{L}_{\text{aux}}$ 

```

Algorithm 22 Invariant point attention (IPA)

def InvariantPointAttention($\{\mathbf{s}_i\}, \{\mathbf{z}_{ij}\}, \{T_i\}, N_{\text{head}} = 12, c = 16, N_{\text{query points}} = 4, N_{\text{point values}} = 8$) :

- 1: $\mathbf{q}_i^h, \mathbf{k}_i^h, \mathbf{v}_i^h = \text{LinearNoBias}(\mathbf{s}_i)$ $\mathbf{q}_i^h, \mathbf{k}_i^h, \mathbf{v}_i^h \in \mathbb{R}^c, h \in \{1, \dots, N_{\text{head}}\}$
 - 2: $\vec{\mathbf{q}}_i^{hp}, \vec{\mathbf{k}}_i^{hp} = \text{LinearNoBias}(\mathbf{s}_i)$ $\vec{\mathbf{q}}_i^{hp}, \vec{\mathbf{k}}_i^{hp} \in \mathbb{R}^3, p \in \{1, \dots, N_{\text{query points}}\}$, units: nanometres
 - 3: $\vec{\mathbf{v}}_i^{hp} = \text{LinearNoBias}(\mathbf{z}_{ij})$ $\vec{\mathbf{v}}_i^{hp} \in \mathbb{R}^3, p \in \{1, \dots, N_{\text{point values}}\}$, units: nanometres
 - 4: $b_{ij}^h = \text{LinearNoBias}(\mathbf{z}_{ij})$
 - 5: $w_C = \sqrt{\frac{2}{9N_{\text{query points}}}}$,
 - 6: $w_L = \sqrt{\frac{1}{3}}$
 - 7: $a_{ij}^h = \text{softmax}_j \left(w_L \left(\frac{1}{\sqrt{c}} \mathbf{q}_i^{h\top} \mathbf{k}_j^h + b_{ij}^h - \frac{\gamma^h w_C}{2} \sum_p \left\| T_i \circ \vec{\mathbf{q}}_i^{hp} - T_j \circ \vec{\mathbf{k}}_j^{hp} \right\|^2 \right) \right)$
 - 8: $\tilde{\mathbf{o}}_i^h = \sum_j a_{ij}^h \mathbf{z}_{ij}$
 - 9: $\mathbf{o}_i^h = \sum_j a_{ij}^h \mathbf{v}_j^h$
 - 10: $\vec{\mathbf{o}}_i^{hp} = T_i^{-1} \circ \sum_j a_{ij}^h (T_j \circ \vec{\mathbf{v}}_j^{hp})$
 - 11: $\tilde{\mathbf{s}}_i = \text{Linear} \left(\text{concat}_{h,p}(\tilde{\mathbf{o}}_i^h, \mathbf{o}_i^h, \vec{\mathbf{o}}_i^{hp}, \|\vec{\mathbf{o}}_i^{hp}\|) \right)$
 - 12: **return** $\{\tilde{\mathbf{s}}_i\}$
-

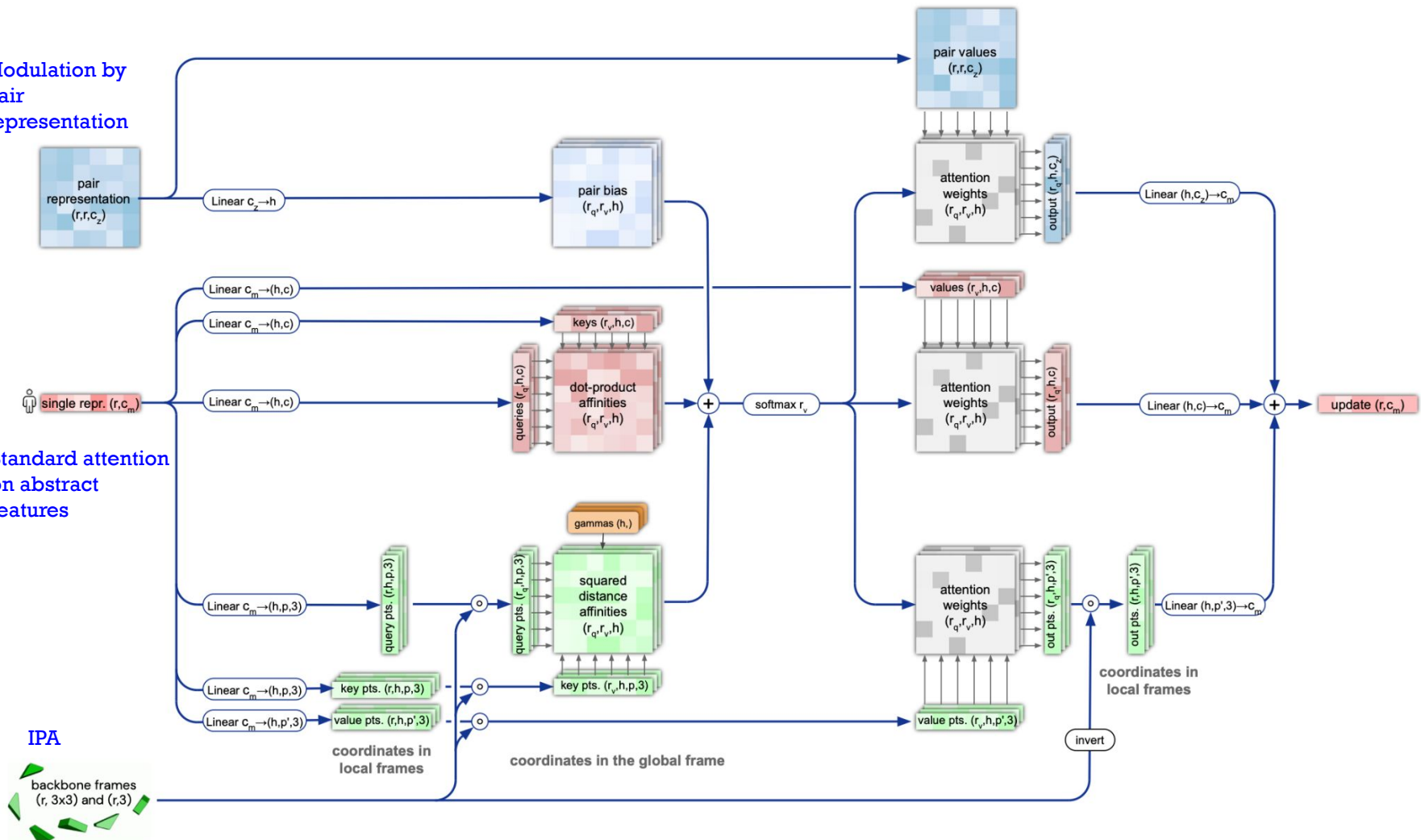


IPA module

Modulation by pair representation

Standard attention on abstract features

IPA



LOSS

Attaching individual loss to each
model component

- Frame Aligned Point Error (FAPE)
- Auxiliary loss
- Distogram loss
- MSA loss
- Confidence loss
- Violation loss

Algorithm 28 Compute the Frame aligned point error

def computeFAPE($\{T_i\}, \{\vec{x}_j\}, \{T_i^{\text{true}}\}, \{\vec{x}_j^{\text{true}}\}, Z = 10\text{\AA}, d_{\text{clamp}} = 10\text{\AA}, \epsilon = 10^{-4}\text{\AA}^2$) :

$$T_i, T_i^{\text{true}} \in (\mathbb{R}^{3 \times 3}, \mathbb{R}^3)$$

$$\vec{x}_j, \vec{x}_j^{\text{true}} \in \mathbb{R}^3,$$

$$i \in \{1, \dots, N_{\text{frames}}\}, j \in \{1, \dots, N_{\text{atoms}}\}$$

1: $\vec{x}_{ij} = T_i^{-1} \circ \vec{x}_j$

$$\vec{x}_{ij} \in \mathbb{R}^3$$

2: $\vec{x}_{ij}^{\text{true}} = T_i^{\text{true}^{-1}} \circ \vec{x}_j^{\text{true}}$

$$\vec{x}_{ij}^{\text{true}} \in \mathbb{R}^3$$

3: $d_{ij} = \sqrt{\|\vec{x}_{ij} - \vec{x}_{ij}^{\text{true}}\|^2 + \epsilon}$

$$d_{ij} \in \mathbb{R}$$

4: $\mathcal{L}_{\text{FAPE}} = \frac{1}{Z} \text{mean}_{i,j}(\text{minimum}(d_{\text{clamp}}, d_{ij}))$

5: **return** $\mathcal{L}_{\text{FAPE}}$

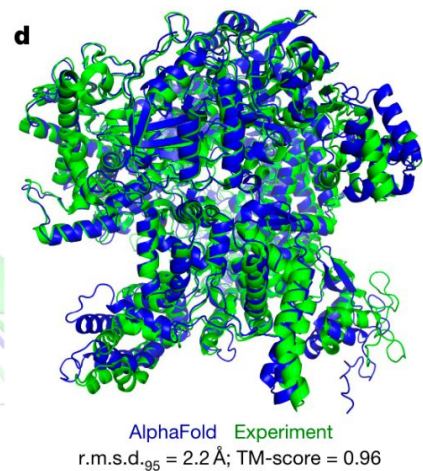
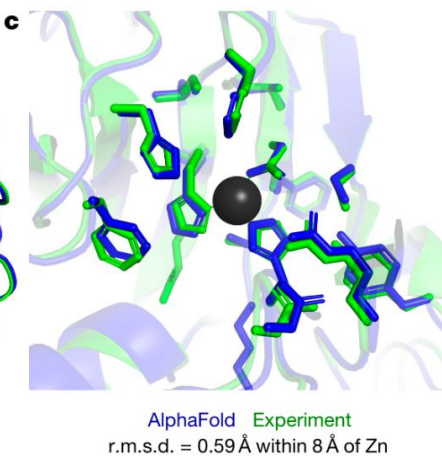
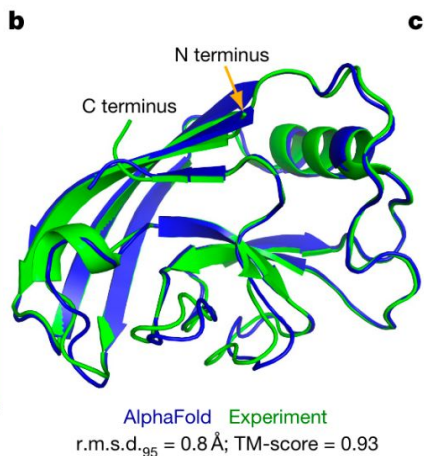
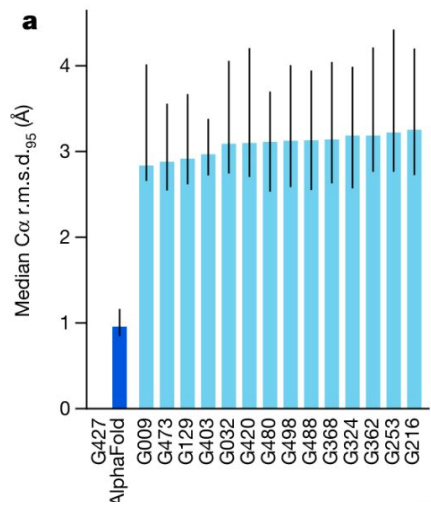
- Final FAPE loss: score all atoms in backbone and sidechain
- Auxiliary: score only Ca atoms in backbone frames

Frame aligned point error (FAPE) loss

Scoring predicted atom coordinates under predicted *local* frames

against

corresponding ground truth coordinates and *local* frames



RESULTS - CASP14





OUTLINE



Motivation

Overview

Evoformer

Structure Module

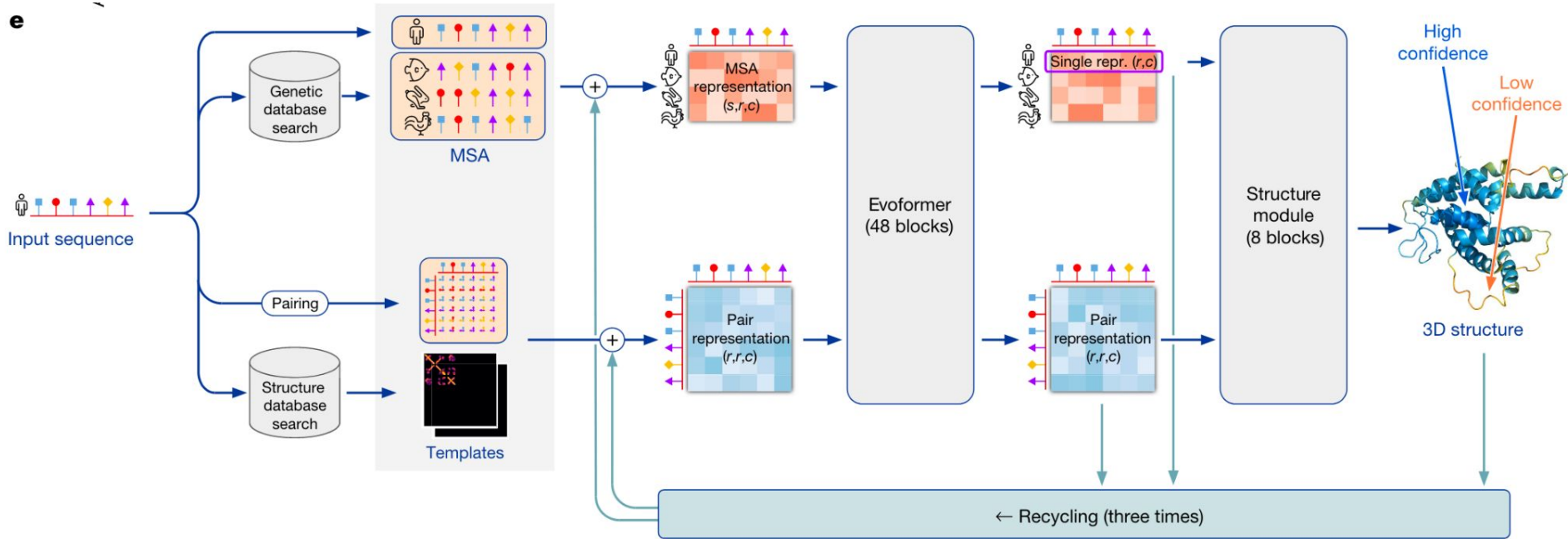
Interpretability

MODEL INTERPRETABILITY

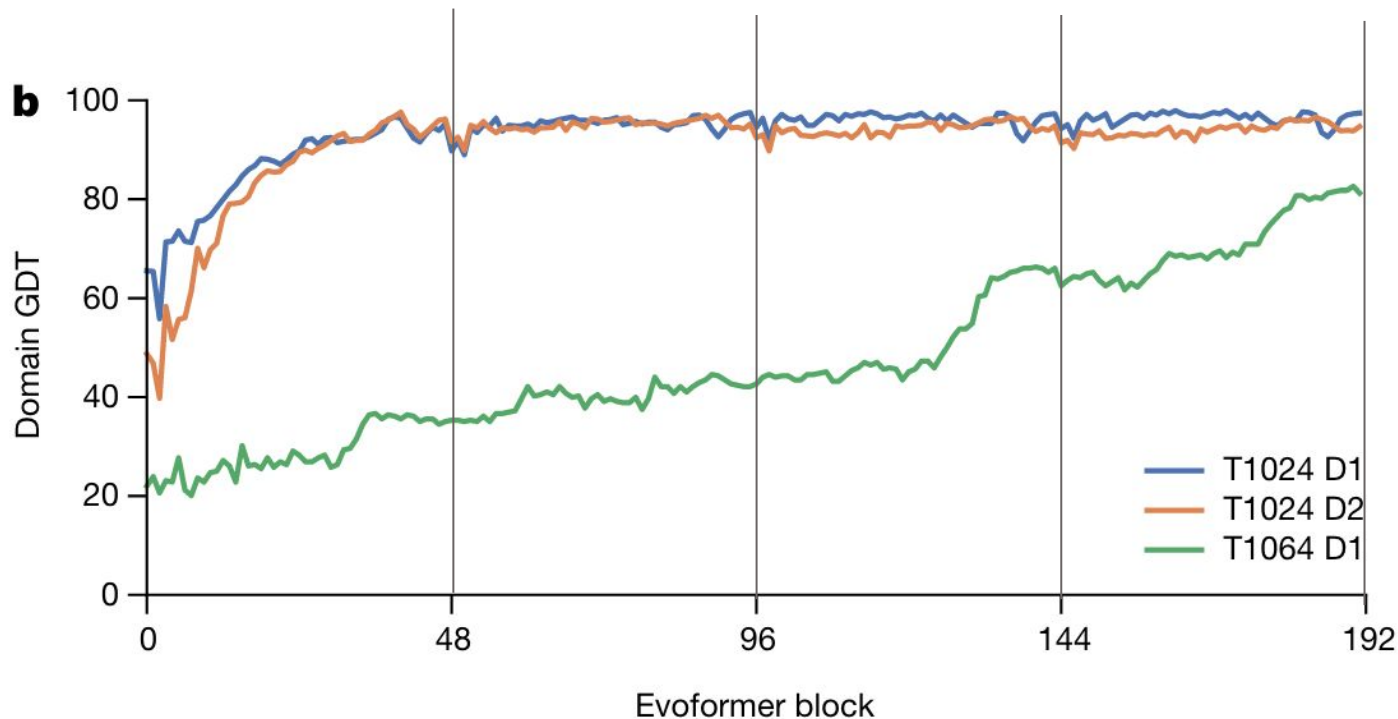
Do we understand why it works so well?



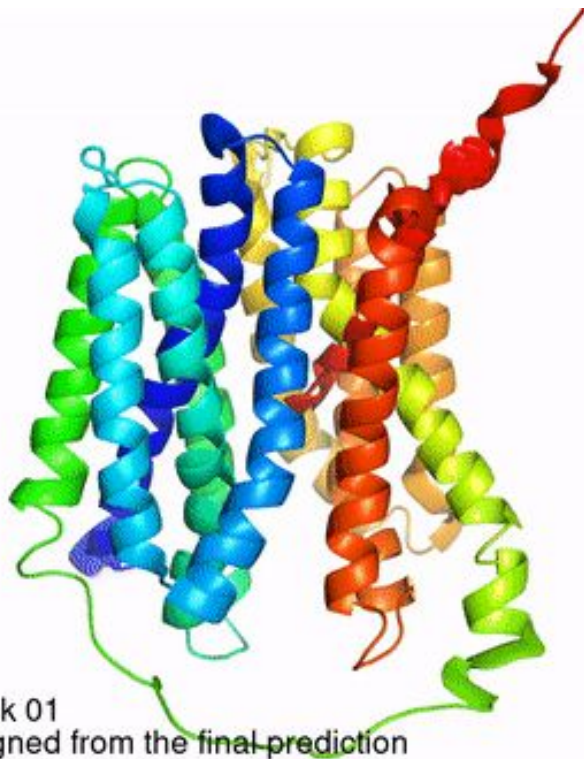
RECALL OVERALL ARCHITECTURE



PERFORMANCE TRAJECTORY

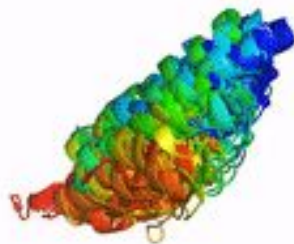


T1024 - easy folding



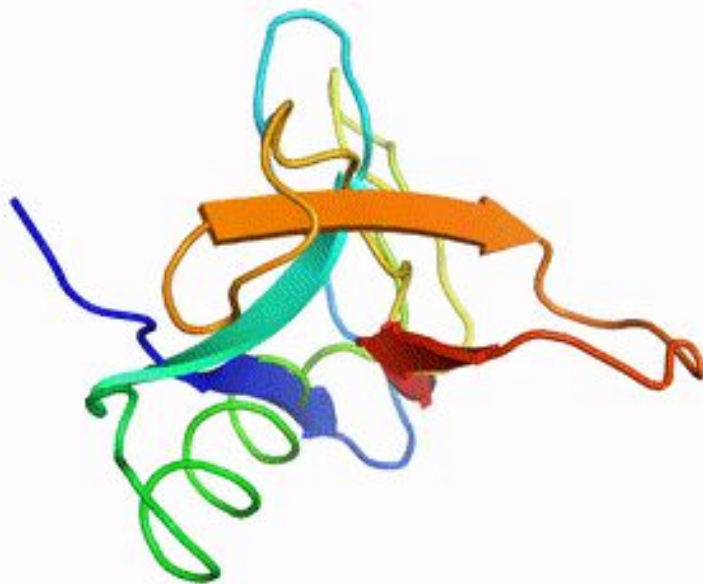
Recycling iteration 0, block 01
Secondary structure assigned from the final prediction

T1044 - some domain folds quickly, others take longer



Recycling iteration 0, block 01
Secondary structure assigned from the final prediction

T1064 - hard one that takes entire training depth to fold



Recycling iteration 0, block 01
Secondary structure assigned from the final prediction

T1091 - very hard, explored “unphysical configurations”



Recycling iteration 0, block 01
Secondary structure assigned from the final prediction

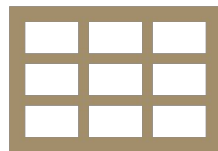
TRAINING DATASET



Supervised:

Protein Database (PDB) data with known structure

High performance

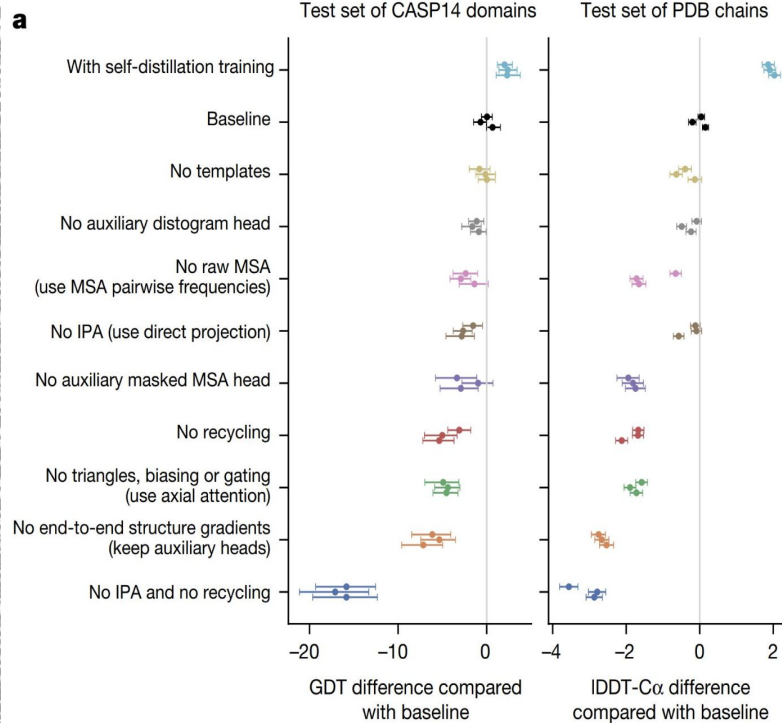


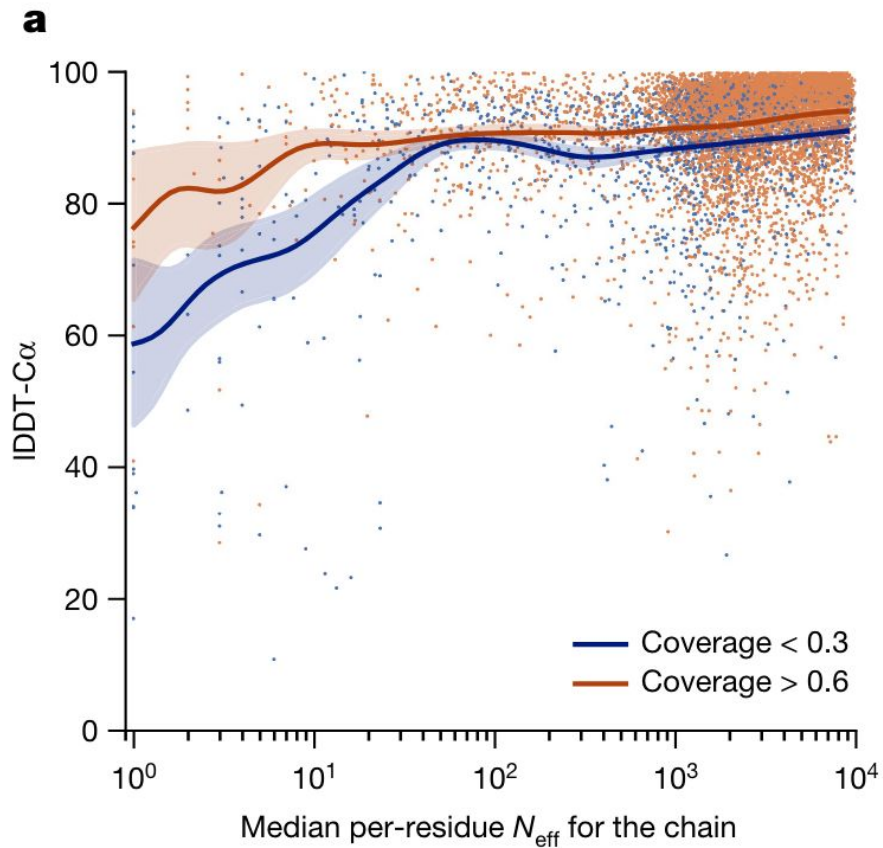
Unsupervised

1. Masked MSA
2. Self-distillation
Uniclust30 sequence only

High “confidence” subset are taken as a new dataset for training

ALL COMPONENTS CONTRIBUTE TO PERFORMANCE





LIMITATIONS

- Doesn't work very well when MSA depth is low
- Doesn't work very well when subunits are heterogeneous (non-homomers)

SUMMARY



IMPRESSIVE PERFORMANCE



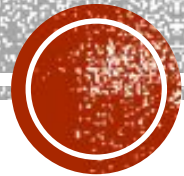
***NOVEL* ARCHITECTURE**



***SOLVED* A LONG-STANDING
PROBLEM, PROVIDE A HUGE
RESERVOIR OF PREDICTED
STRUCTURES (AFDB)**



THANK YOU



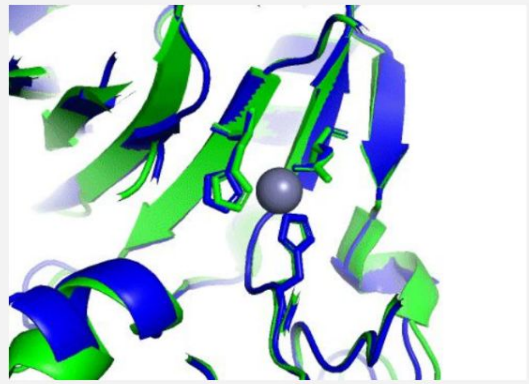
LIST OF FIGURES FROM PAPER

DOES ALPHAFOLD2 UNDERSTAND

TOTAL PROTEIN FOLDS

- Computational structure prediction is typically underspecified
 - Oligomeric state, ligands, DNA-binding, experimental conditions, multiple conformations etc.
- Our network is tolerant to missing context
- AlphaFold is just as good at membrane proteins or novel folds as more typical PDB structures

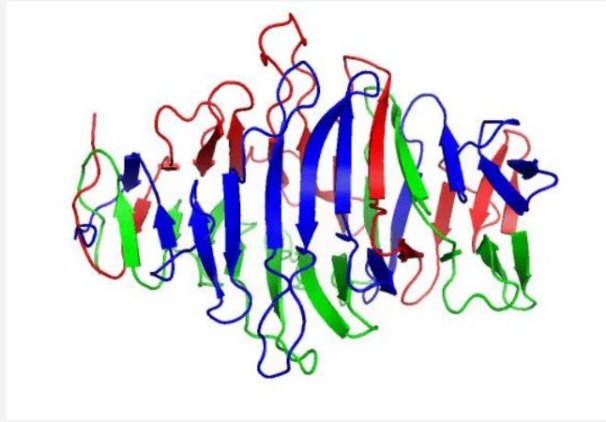
T1056 (zinc binding)



AlphaFold / **Experiment**

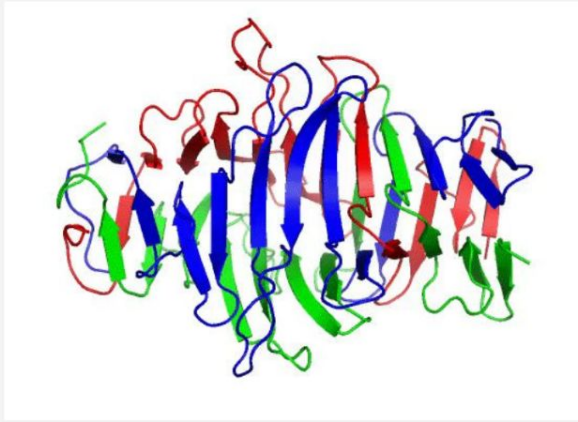
TBM-hard, 98.2 GDT

T1080 (trimer)



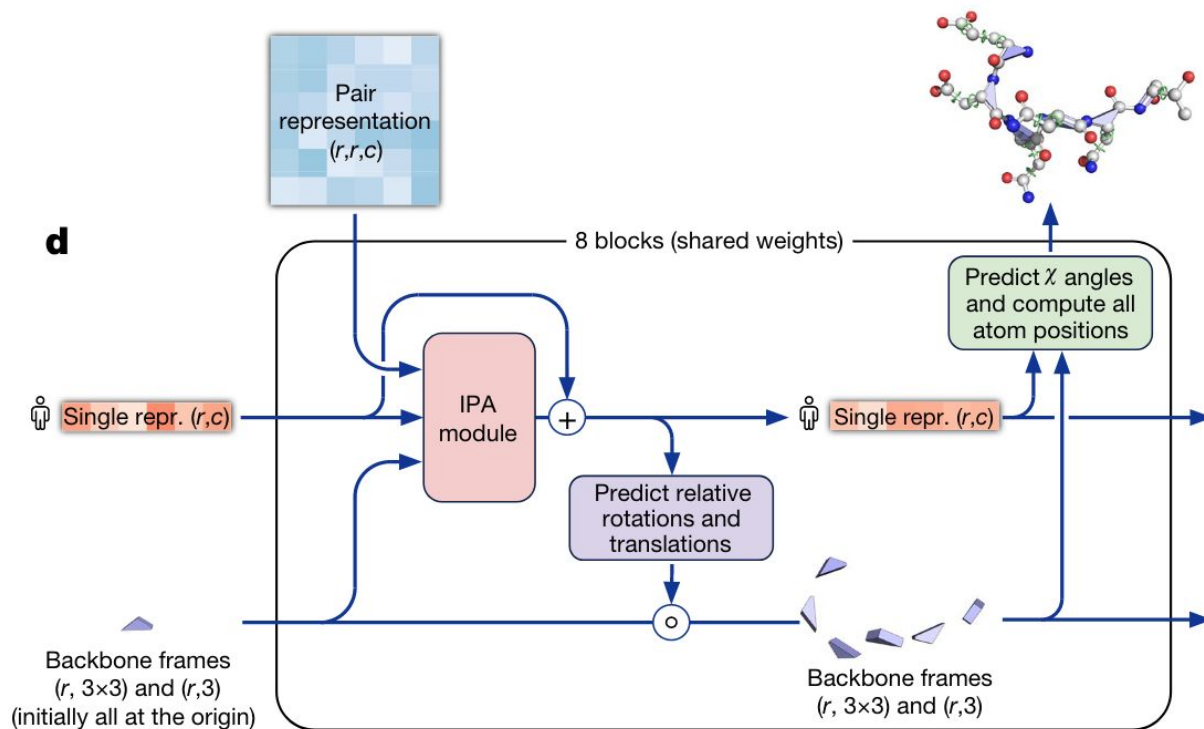
AlphaFold (monomer prediction x3)

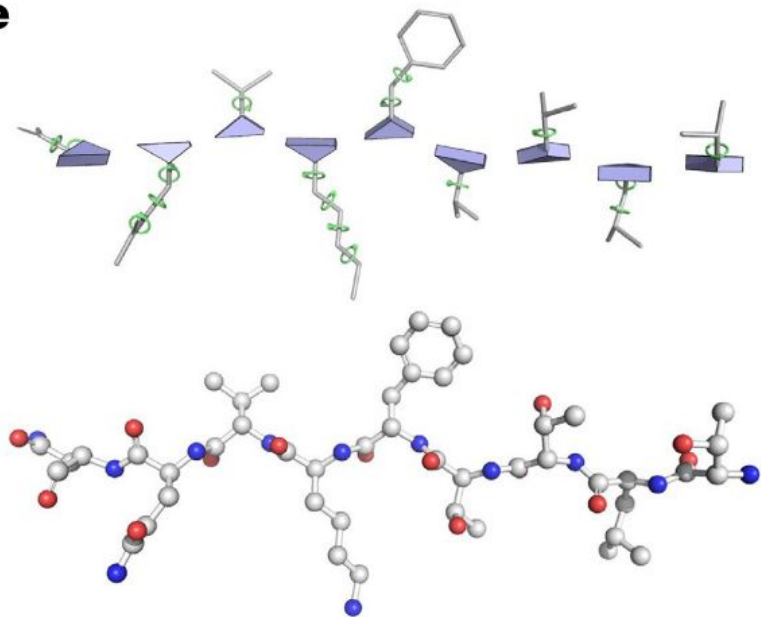
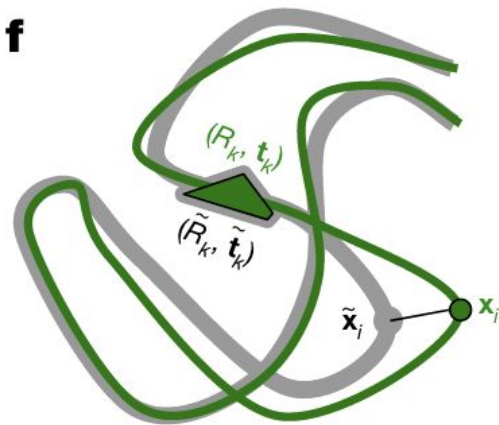
FM/TBM, 85.9 GDT



Experimental structure

STRUCTURAL MODULE



e**f**

GRAPHICAL OUTLINE

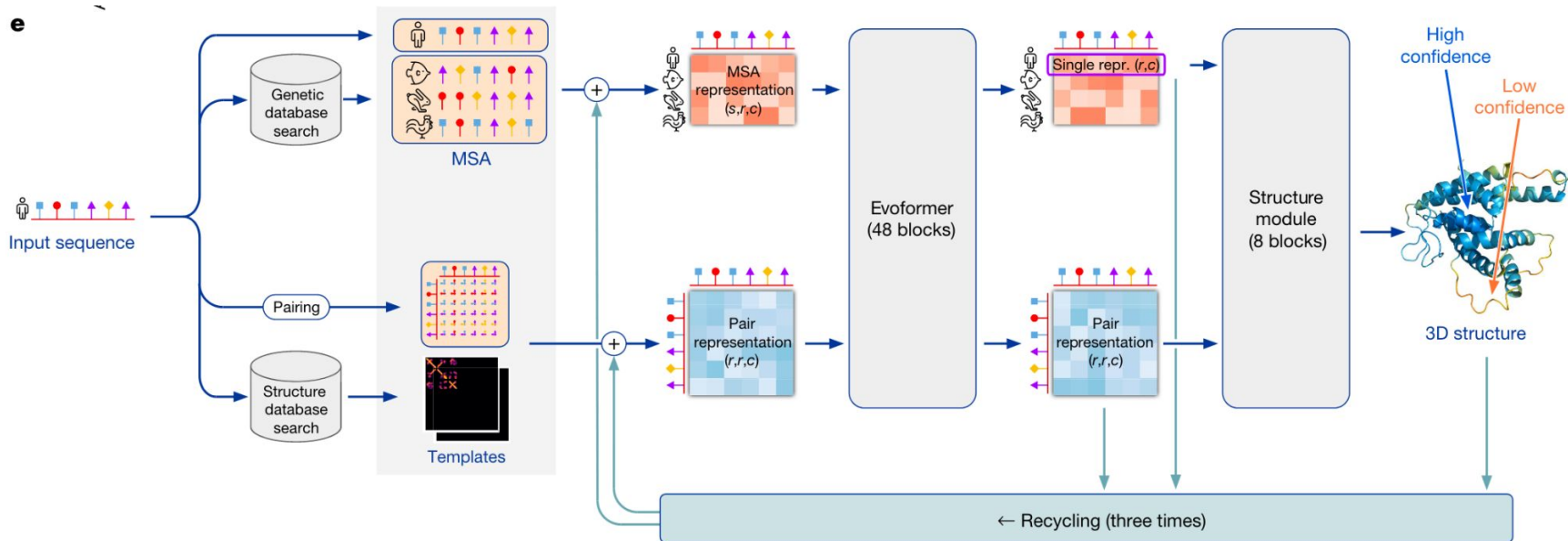
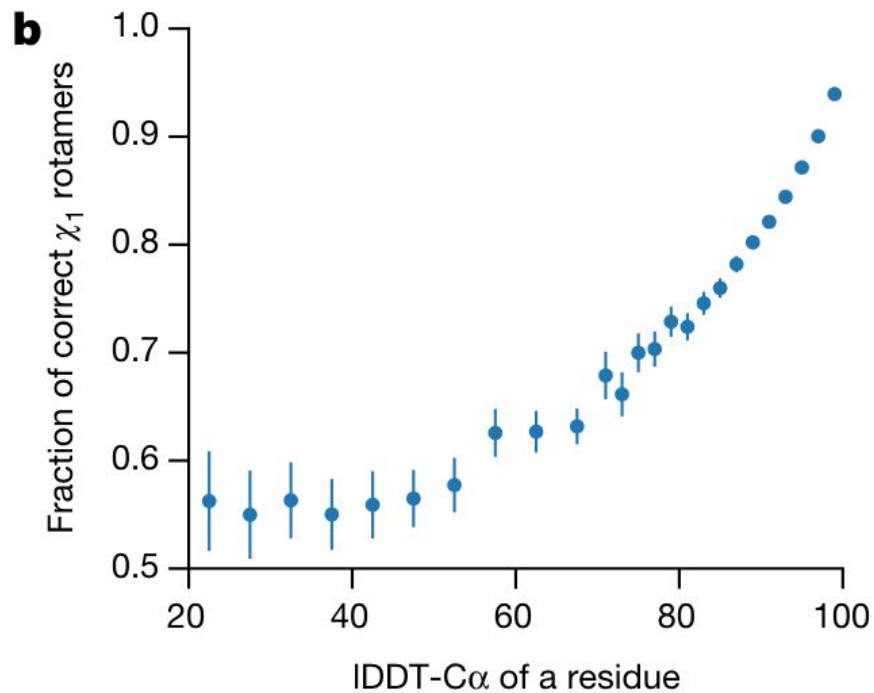
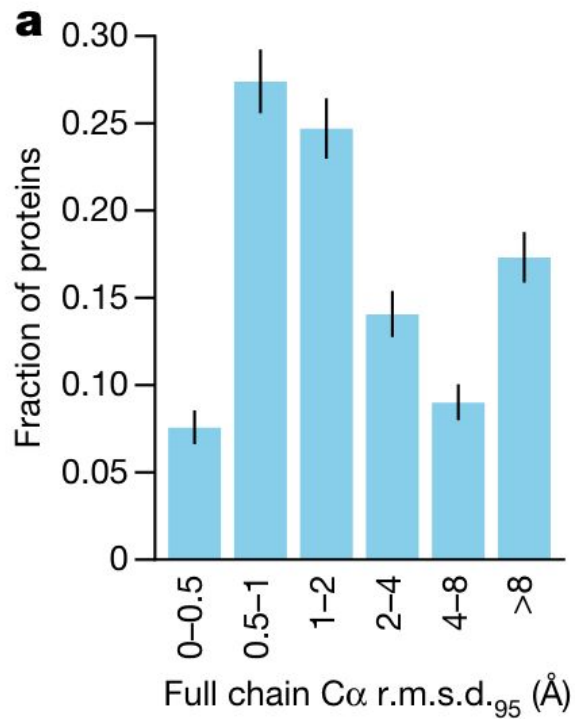
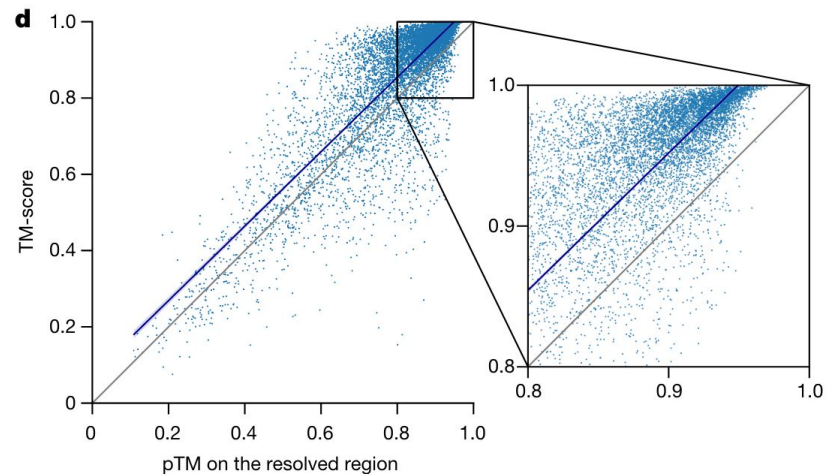
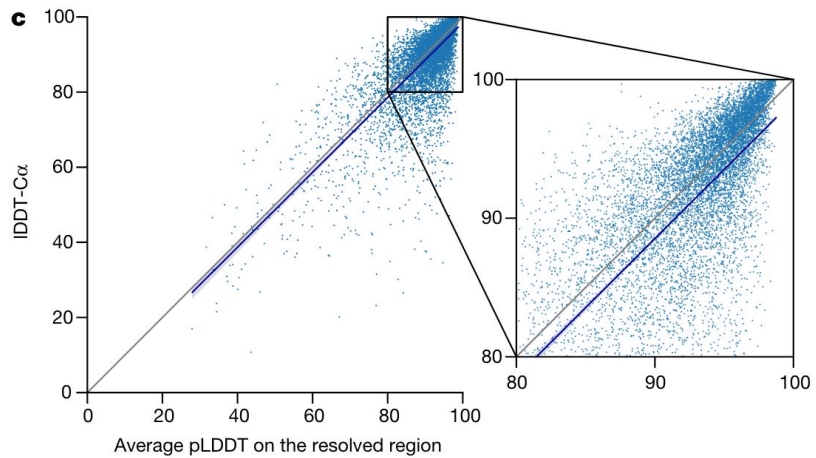
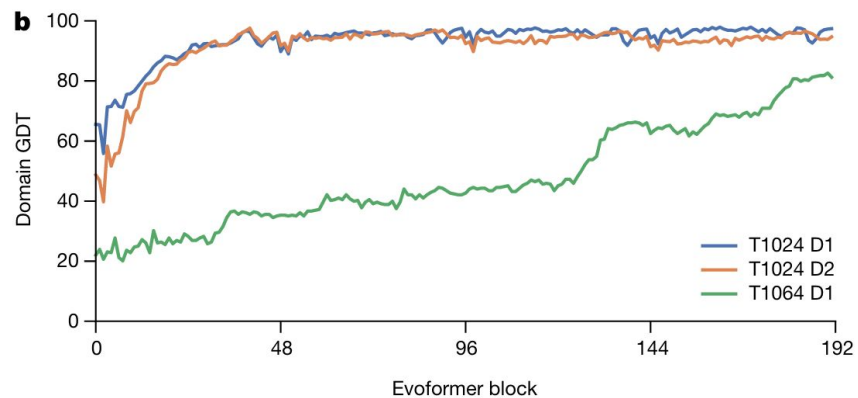
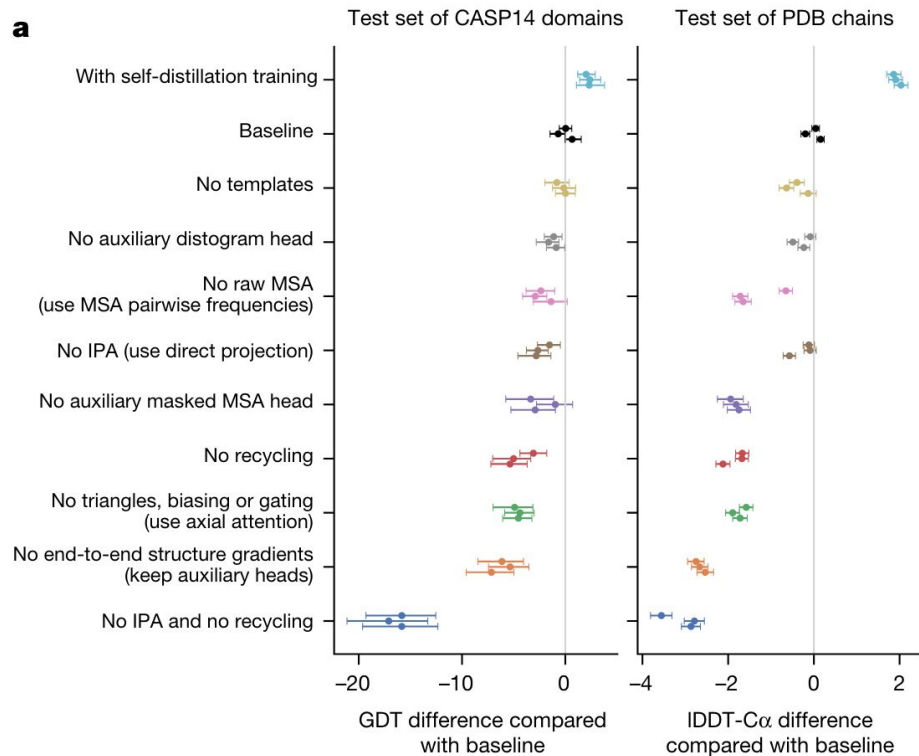
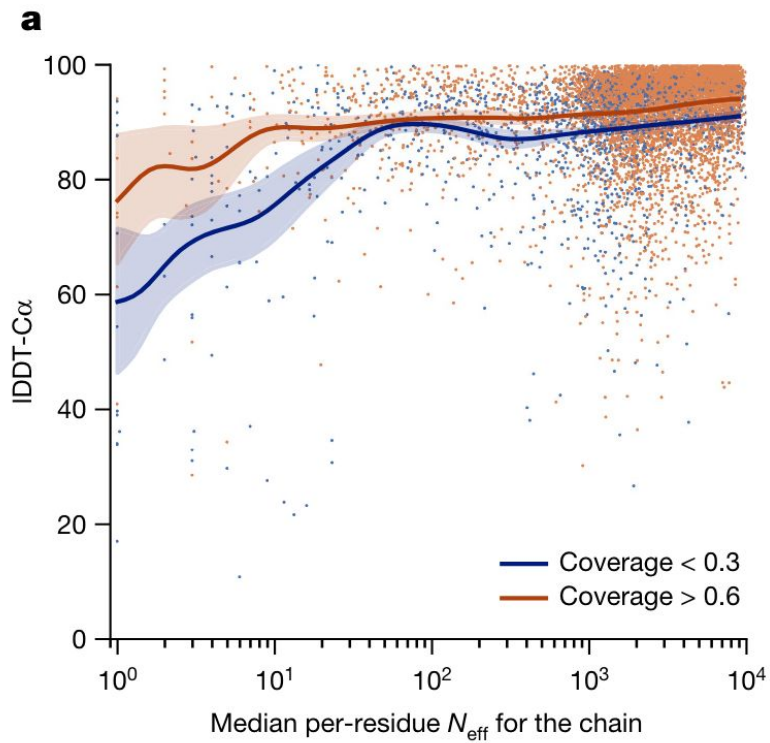


Figure 1(e)









b

