CryoSAMU: Enhancing 3D Cryo-EM Density Maps of Protein Structures at Intermediate Resolution with Structure-Aware Multimodal U-Nets

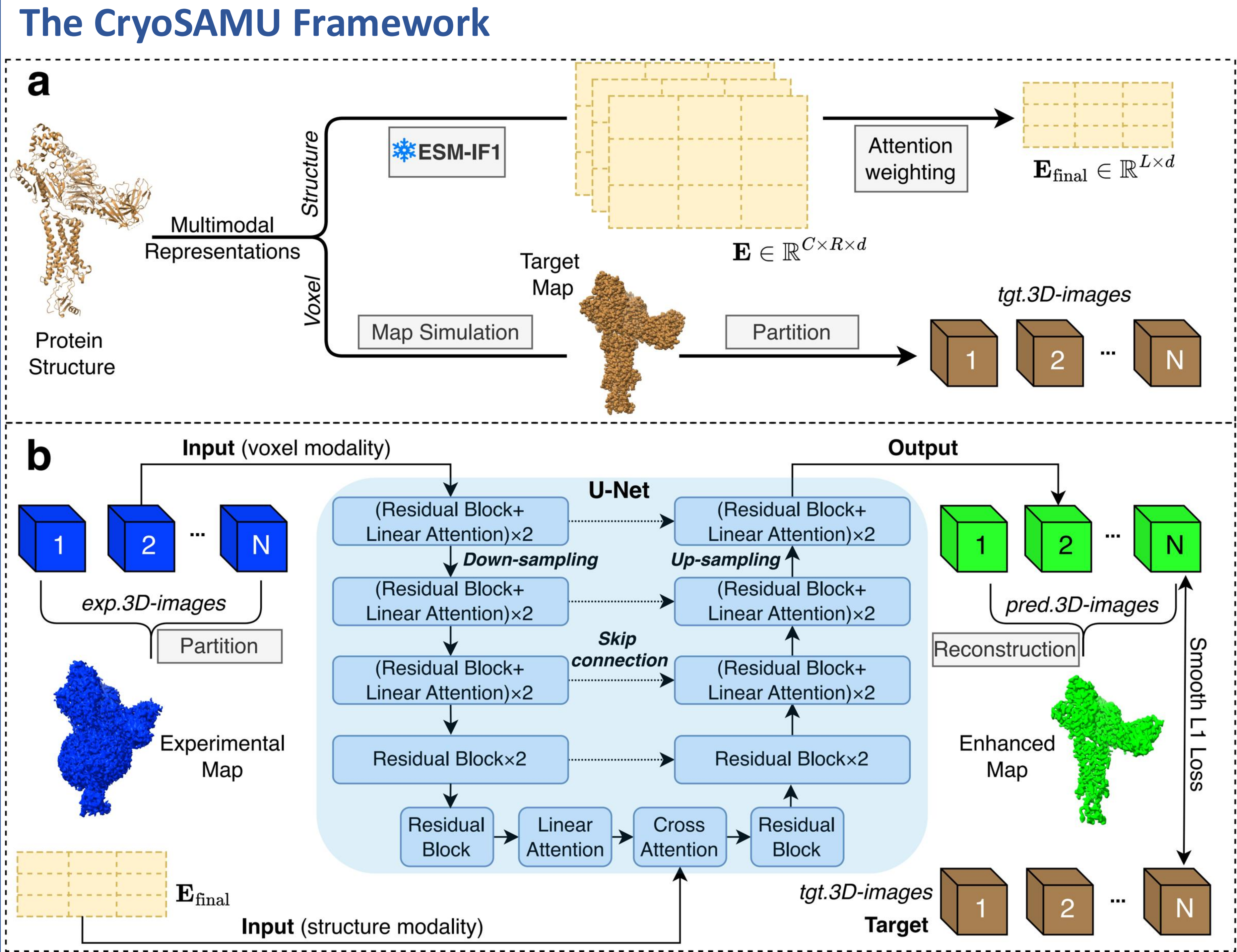
Chenwei Zhang[1] (cwzhang@cs.ubc.ca), Khanh Dao Duc[2] (kdd@math.ubc.ca)
[1]Department of Computer Science, [2]Department of Mathematics, University of British Columbia, Vancouver, Canada

Motivations

- Intermediate-resolution cryo-EM maps (4–8 Å) are common but challenging for accurate protein structure modeling due to low contrast and structural ambiguity.
- Existing enhancement methods are not optimized for this resolution range and use only density information, ignoring valuable structural context.
- Protein Language Models (pLLMs) like ESM-IF1[1] offer rich structural embeddings but remain underutilized in map enhancement.
- There's a need for fast, accurate, and structure-aware approaches to facilitate map interpretability for downstream applications like protein structure modeling.
- ✅ To address these challenges, we developed **CryoSAMU** — the first multimodal network that integrates protein structural embeddings into a 3D voxel-based U-Net using cross-attention[2], enabling enhanced cryo-EM maps optimized for intermediate resolution.

Methods

The CryoSAMU Framework



Self-Attention Weighting for Structural Embeddings

- Handles proteins with variable chains and residues
- Use soft attention to preserve informative structures
- Produce fixed-size structural embeddings for multimodal learning

Procedure:

1. Output from ESM-IF1 for all chains and residues:

$$E \in \mathbb{R}^{C\times R\times d} \quad (C\text{: no. chains}, R\text{: no. residues}, d\text{: emb size})$$

2. Aggregate residue embeddings into chain-level embeddings:

$$E_i^{chain} = \frac{1}{R}\sum_{j=1}^{R} E_{i,j} \implies E^{chain} \in \mathbb{R}^{C\times d}$$

3. Compute pairwise chain similarity and Softmax attention weights:

$$S = E^{chain}\cdot (E^{chain})^{\mathrm{T}}, \quad W_{ij} = \frac{\exp(S_{ij})}{\sum_k \exp(S_{ik})}$$

4. Get chain importance and perform weighted aggregation:

$$w_i = \frac{1}{C}\sum_j W_{ij}, \quad E^{pooled} = \sum_{i=1}^{C} w_i \cdot E_i \implies E^{pooled} \in \mathbb{R}^{R\times d}$$

5. Repeat the above steps to E^{pooled} for residue-level attention to obtain a scalar weight α_j for each residue j = 1,2, ..., R.

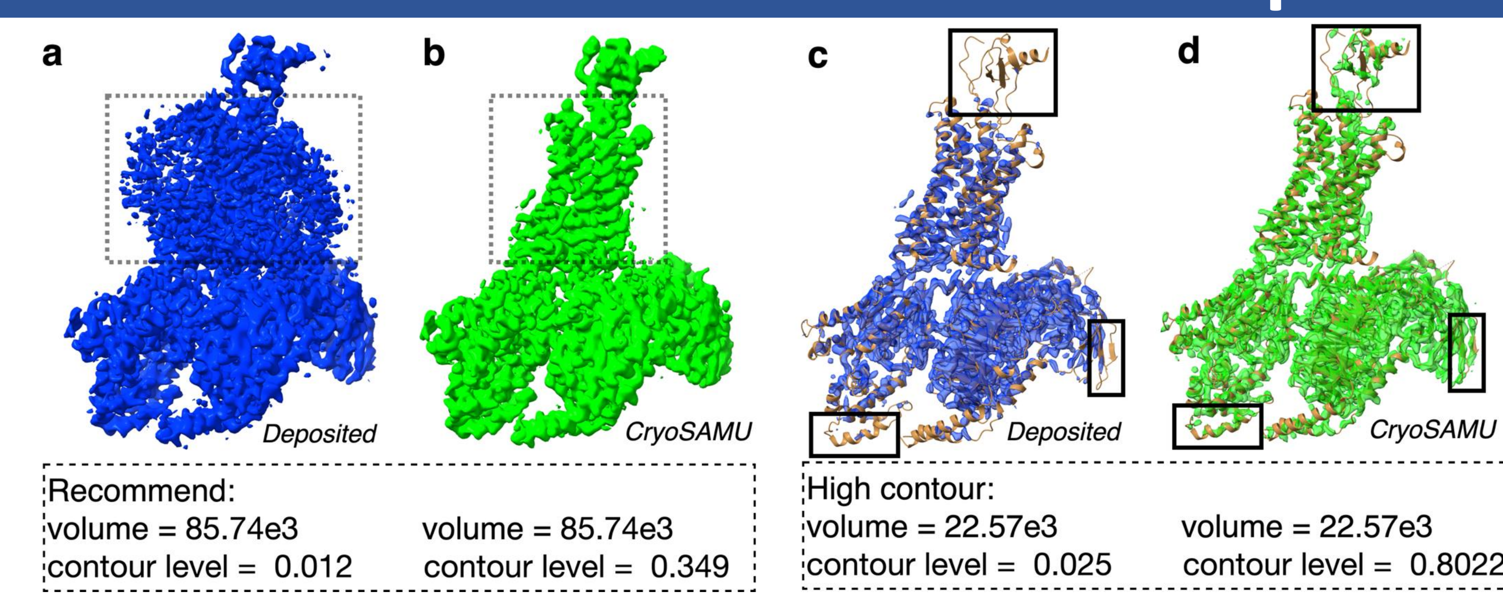
6. Normalize and resample E^{pooled} w.r.t α_j to select top-L informative residues:

$$E^{final} \in \mathbb{R}^{L\times d}$$

Dataset

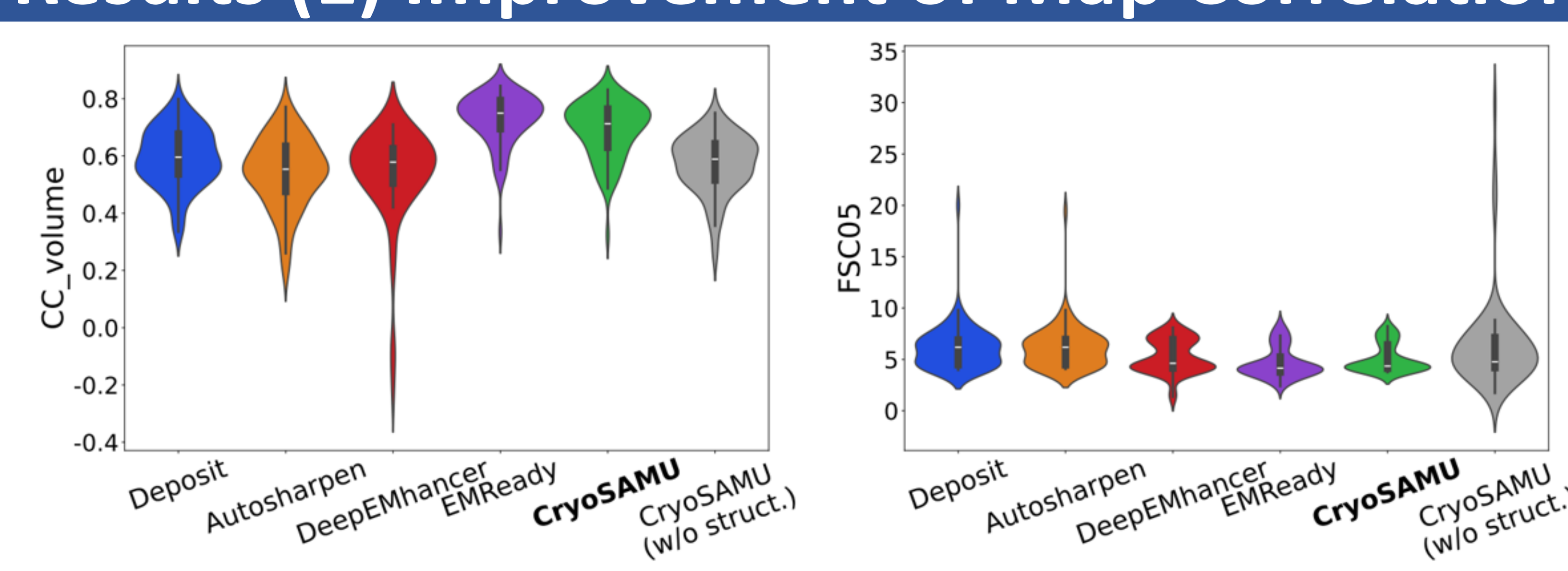
- Dataset includes 384 pairs of cryo-EM density maps and protein structures from EMDB[3] and PDB[4].
- Excluded maps with misaligned PDB structures or non-protein macromolecules.
- Filtered pairs with correlation score below 0.65 to ensure complete mappings.
- Retained unique PDB structures with sequence identity over 30%.

Visualization of Enhanced Maps



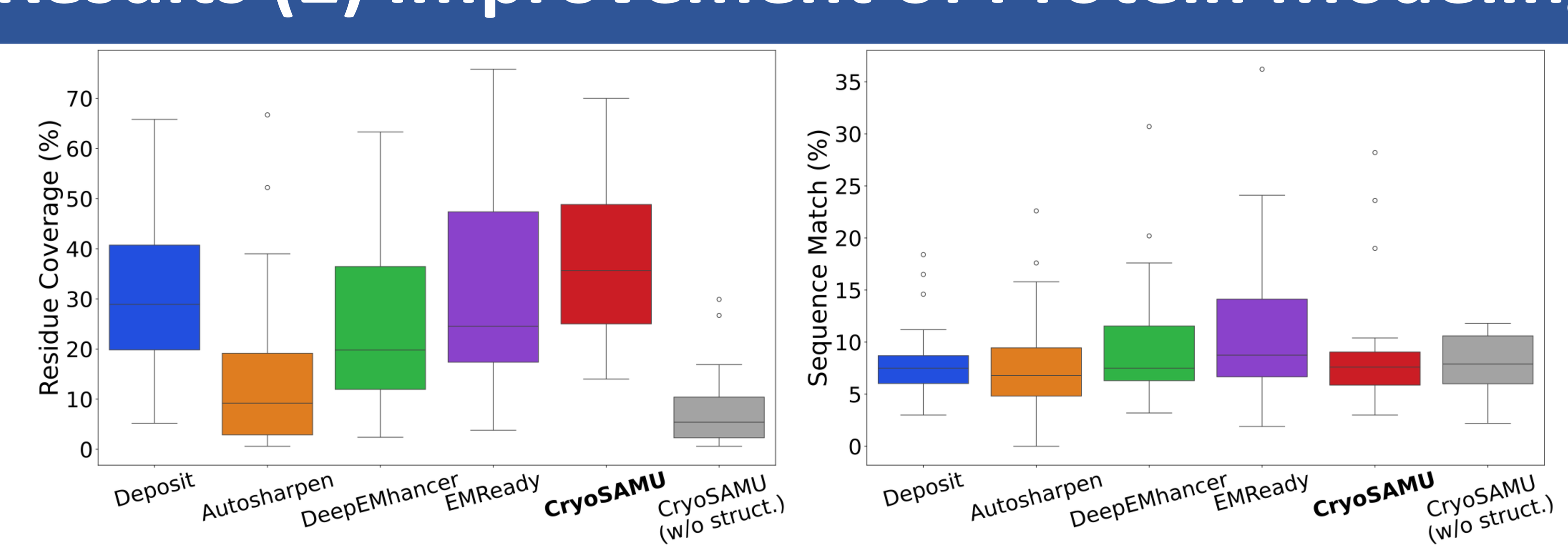
- CryoSAMU-enhanced maps exhibit better alignment with corresponding protein structures, revealing more structural details, outlined by black boxes in Figure (d).

Results (1) Improvement of Map Correlation



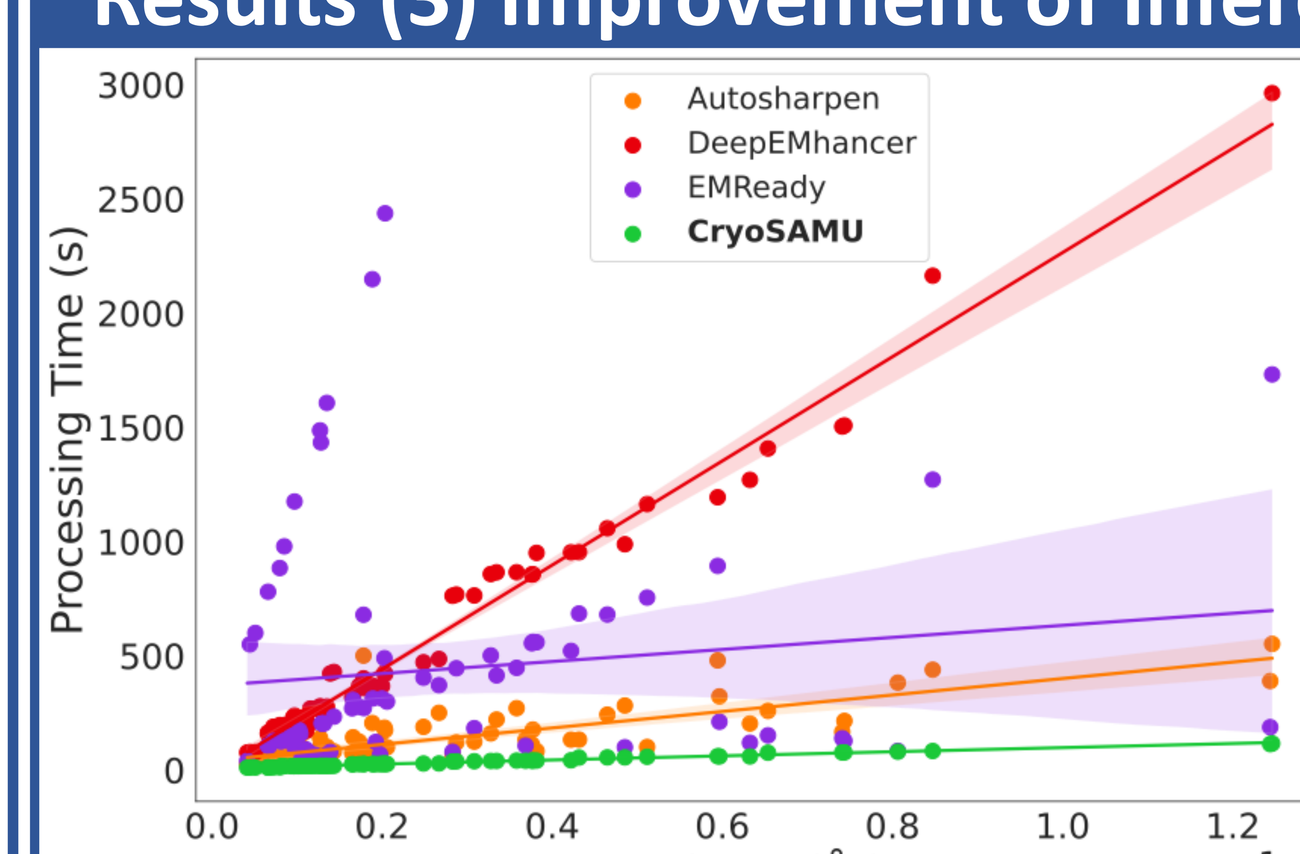
- CryoSAMU outperforms deposited maps in both real-space and Fourier-space correlation
- CryoSAMU shows competitive performance compared to SOTAs, such as EMReady[5].

Results (2) Improvement of Protein Modeling



- CryoSAMU achieves the best residue coverage score, while the sequence match score is slightly lower than EMReady[5] and DeepEMHancer[6].
- Integrating structural embeddings enhances the continuity and interpretability of generated maps, improving protein structure prediction.

Results (3) Improvement of Inference Speed



- CryoSAMU achieves an average processing time of 32.49 seconds, approximately 13.6 times faster than EMReady.
- Promising tool for large-scale and practical applications.

References

1. Hsu et al., ICML 2022, pp.8946-8970.
2. Vaswani et al., NeurIPS 2017, Vol. 30.
3. Lawson et al., Nucleic Acids Res. 2016, pp.D396–D403.
4. Berman et al., Acta Crystallogr. D, 2002, pp. 899–907.
5. He et al., Nat. Commun. 2023, pp. 3217.
6. Sanchez-Garcia et al. Commun. Bio. 2021