MULTI-STATE PROTEIN DESIGN WITH DYNAMICMPNN

Anonymous authorsPaper under double-blind review

ABSTRACT

Structural biology has long been dominated by the *one sequence, one structure, one function* paradigm, yet many critical biological processes—from enzyme catalysis to membrane transport—depend on proteins that adopt multiple conformational states. Existing multi-state design approaches rely on post-hoc aggregation of single-state predictions, achieving poor experimental success rates compared to single-state design. We introduce DynamicMPNN, an inverse folding model explicitly trained to generate sequences compatible with multiple conformations through joint learning across conformational ensembles. Trained on 46,033 conformational pairs covering 75% of CATH superfamilies and evaluated using Alphafold 3, DynamicMPNN outperforms ProteinMPNN by up to 25% on decoy-normalized RMSD and by 12% on sequence recovery across our challenging multi-state protein benchmark.

1 Introduction

A commonly derived assumption from Anfinsen's experiment is that proteins adopt only one native 3D structure, leading to the "one sequence, one structure, one function" canon. This view has been indirectly reinforced by the predominant use of X-Ray crystallography in experimental protein structure determination, which requires that proteins form a homogenous, diffractible crystal to be characterised (Dishman & Volkman, 2018). The large collection of static protein structures in the PDB has enabled the development of high-accuracy machine learning models for tasks such as structure prediction (Jumper et al., 2021; Kryshtafovych et al., 2019) and inverse folding (Dauparas et al., 2022; Hsu et al., 2022a). Amongst contemporary inverse folding models, ProteinMPNN has been particularly widely adopted in applied protein design projects due to its low inference costs and robust experimental success rates (Dauparas et al., 2022; Watson et al., 2023; Goverde et al., 2024), outperforming traditional physics-based design methods on both fronts (Liu & Kuhlman, 2006).

Although direct experimental characterisation of protein dynamics remains a challenge, the conformational diversity of proteins underlies crucial biological functions such as enzyme catalysis, protein-protein interactions, allostery, and human disease (Monzon et al., 2016). In applied protein design, bio-switches - proteins that switch between two structural states - are of particular importance, with key applications in engineering artificial bio-motors, signalling pathways, biosensors, or drug delivery systems (Stein & Alexandrov, 2015; Praetorius et al., 2023). While most known switches undergo rearrangements in the context of a single fold (Ambroggio & Kuhlman, 2006a), the class of metamorphic proteins undergo changes in both their secondary structure and fold (Fig. 1a) and have been predicted to represent up to 4% of the PDB chains (Porter & Looger, 2018). These proteins typically adopt two main functional states (Dishman & Volkman, 2018) and a finite number of conformations (see Discussion). Beyond the world of switches, other dynamic proteins are characterised by continuous conformational landscapes (e.g. intrinsically disordered proteins (Tompa & Fuxreiter, 2008)).

Multi-state protein design was first achieved through rational design and physics-based methods such as RosettaDesign (Liu & Kuhlman, 2006). Previous campaigns leveraging these methods have attempted to design metamorphic metal-binding peptides (Ambroggio & Kuhlman, 2006b; Cerasoli et al., 2005), closely related sequences that adopt diverging folds (Wei et al., 2020), and hinge proteins with binder-regulated thermodynamic equilibria, allowing the relative populations of different structural states to be modulated by exogenous proteins (Zhang et al., 2022; Quijano-Rubio et al.,

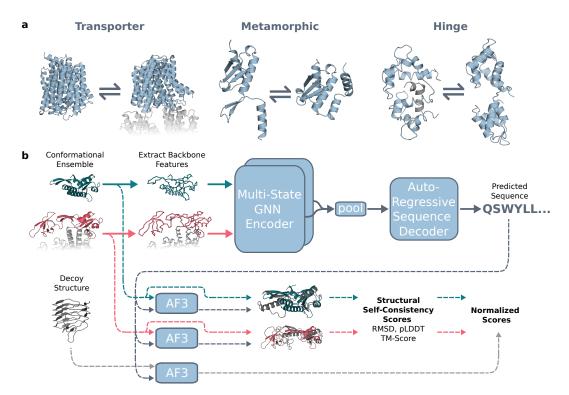


Figure 1: DynamicMPNN for multi-state protein design. (a) Examples of proteins with multiple conformational states: transporters in closed and open states (PDB: 6NC7, 6NC9), metamorphic protein with alternative folds (PDB: 4QHH, 4QHF) and hinges showing domain movement (PDB: 5D0W, 1CFC). (b) Schematic of DynamicMPNN, an inverse folding model trained to generate protein sequences with multiple conformational states. Conformations are encoded with their respective chemical environments (i.e. interaction partners shown in gray). Solid lines show the flow of information in the model, while dashed lines show the evaluation pipeline using AlphaFold 3 (AF3); employing target structures as templates during inference and measuring the deviations between predicted and target structures, with decoy structures serving as negative controls.

2021; Praetorius et al., 2023). More recently, Praetorius et al. (2023) (2023) developed Protein-MPNN Multi-state Design (ProteinMPNN-MSD) (Dauparas et al., 2022): an inference strategy for extending ProteinMPNN inverse folding to multiple structural states by averaging the logits of two independent single-state ProteinMPNN embeddings during the decoding step. In a similar vein, ProteinGenerator (Lisanza et al., 2024) applies a sequence diffusion process conditioned on structural predictions by RosettaFold (Baek et al., 2021), and attempts to design sequences that adopt multiple states by averaging the sequence logits predicted by the model across distinct target structures.

Still, significant limitations in current multi-state design pipelines remain. Designs generated from ProteinMPNN-MSD report low overall experimental success rates: only 46 2-state hinge sequences, corresponding to roughly 0.002% of all designs, were successfully expressed in solution, out of which only nine showed binding to the corresponding target peptide. Likewise, the authors of ProteinGenerator reported a significantly lower *in silico* design success rate of 0.05% for their multi-state design task compared to rates of 2-10% observed for various single-state sequence design objectives using their framework. These observations - combined with the relative scarcity of published data on ML-driven multi-state *de novo* design campaigns - suggest that current ML methods for multi-state protein design have been significantly less successful than their single-state counterparts (Appendix ??). We propose that this gap can be attributed to limited multi-conformational datasets, weak benchmarks, and the poor performance of folding models in predicting alternative states (Chakravarty et al., 2024) (which adversely affects their efficacy as self-consistency filters in protein design workflows).

Our contributions. This paper introduces DynamicMPNN (Fig. 1b), a novel geometric deep learning-based pipeline for multi-state protein sequence design.

- DynamicMPNN is the first explicit multi-state inverse folding model for protein design. To train DynamicMPNN, we create a new ML-ready dataset of proteins with multiple conformations using the PDB and CoDNaS (Monzon et al., 2016) databases, and evaluate the method on 96 biologically relevant metamorphic, hinge, and transporter proteins.
- We introduce a novel data processing pipeline and architecture able to handle sequencealigned structure pairs with heterogeneous sequences, enabling us to leverage the conformational diversity across non-identical protein states.
- We propose a multi-state self-consistency metric and benchmark based on *Alphafold3* (AF3) (Abramson et al., 2024) using target structures as templates.
- DynamicMPNN improves performance over ProteinMPNN (Dauparas et al., 2022) on AF3 by up to 25% on RMSD and 8% on pLDDT decoy normalized self-consistency values.

2 THE DYNAMICMPNN PIPELINE

2.1 PROTEIN MULTI-CONFORMATIONAL DATASET

While over 900,000 individual protein chains (sequence-structure pairs) are available in the PDB, multi-conformational data is at a first glance far more scarce with only roughly 12,000 NMR-derived protein ensembles covering just 21% of CATH superfamilies. To overcome this limitation, we exploit the sequence redundancy in the PDB: following Monzon et al. (2016) and Hrabe et al. (2015), we cluster all protein chain sequences with a very high sequence similarity threshold (≥95%) to build a multi-conformational dataset of 46,033 conformational clusters with at least 2 members expanding coverage to 75% of CATH superfamilies (Fig. 2a). The 95% similarity threshold accounts for soluble tags or point mutations/alterations performed in different experiments which are unlikely to affect the overall fold. Clusters contain highly unbalanced conformational populations (Fig. 2c): a relatively small number of clusters contain thousands of sequences, while the majority contain only a few. This contributes to the relatively small intra-cluster RMSD in the majority of the clusters (Fig. 2b), since many alternative protein states live in low populations.

To maximise the conformational signal and minimize alignment artifacts, we built our dataset using only pairs of chains from one or two PDB entries that have the largest RMSD from each cluster (Fig. 2c). Critically, this selection emphasizes our focus on proteins exploring a finite number of functional states, with most switching between just two conformations (Dishman & Volkman, 2018; Leaver-Fay et al., 2011; Alberstein et al., 2022). While we hope to explore design of more continuous conformational spaces in future work, we focus presently on training DynamicMPNN to design two-state proteins. Furthermore, molecular dynamics datasets were opted against due to the poor training signal given by limited sequence variability and scarcity of trajectories showing large conformational changes (Vander Meersche et al., 2023; Mirarchi et al., 2024), which normally happen across long biological timescales. At the other end of the spectrum, single-chain disordered protein simulations (Tesei et al., 2024) expose continuous conformational landscapes which are hard to discretize and have a high degree of stochasticity with a low signal-to-noise ratio rather than the ordered topologies of globular multi-state proteins (Tompa & Fuxreiter, 2008).

For dataset splitting, we first curate a benchmark composed of five previous studies of proteins with large 2-state conformational changes: (1) 92 metamorphic proteins (Porter & Looger, 2018), (2) 91 apo-holo proteins (Saldaño et al., 2022), (3) the OC23 and OC85 open-closed datasets (Kalakoti & Wallner, 2025), and (4) 20 transporter proteins (Kalakoti & Wallner, 2025). The proteins with the highest inter-state RMSD were assigned to the test set (96 samples), while the rest were assigned to the validation set (100 samples). Training clusters were filtered to exclude any with TM-score > 0.4 (Zhang & Skolnick, 2004) and > 30% sequence similarity to test/validation structures, preventing structural similarity leakage and yielding a final training set of 44,243 conformer pairs.

We additionally curate a set of single-state sequence-structure pairs from the 30% sequence similarity clusters not represented in our multi-conformational training dataset (n=27,394). This augmentation strategy maximizes coverage of protein fold space while preserving the multi-conformational learning signal.

See Appendix C for further details on dataset composition.

2.2 DynamicMPNN for multi-state inverse folding

Single-state inverse folding methods seek to model the conditional distribution p(Y|X) where $X \in \mathbb{R}^{n \times 3 \times 3}$ represents a protein backbone with n residues, and $Y = (y_1, \dots, y_n)$ is the amino acid sequence. Extensions of these methods to multi-state design have thus far been limited to post-hoc aggregation of independent single-state predictions. We believe such methods favour logits which are highly biased towards one conformation, whose average over the 2 states is higher than the one of logits which are moderately valued across both states (Joshi et al., 2025). Instead, DynamicMPNN learns the joint conditional distribution of $p(Y|X_1,\dots,X_m)$ directly through autoregressive sequence generation, where $\{X_1,\dots,X_m\}$ represent distinct protein conformations encoded into a shared latent space; thus learning a sequence distribution that simultaneously satisfies multiple structural constraints. We decompose this joint conditional probability using the autoregressive factorization:

$$p(Y|X_1, ..., X_m) = \prod_{i=1}^n p(y_i|y_{i-1}, ..., y_1; X_1, ..., X_m)$$
(1)

where each factor represents the probability of selecting residue y_i given the sequence prefix and the complete structural ensemble.

Overall architecture. DynamicMPNN independently encodes each of the functional states of a protein, together with their binding partners, into a shared latent feature space (Fig. 1b). Embeddings of the chains-to-be-designed chains are then pooled across conformations to obtain a single embedding from which a sequence is auto-regressively generated.

Our architecture is built upon gRNAde (Joshi et al., 2025), a multi-state GNN for RNA inverse folding. For both encoder and decoder, we employ SE(3)-equivariant Geometric Vector Perceptron (Jing et al., 2021) layers which maintain computational efficiency through edge sparsity (k-NN edges with k=32). Within the GVP, the scalar features $s_i \in \mathbb{R}^{k \times f}$ and vector features $\vec{v}_i \in \mathbb{R}^{k \times f' \times 3}$ for each node i (Duval et al., 2024):

$$m_i, \vec{m}_i \coloneqq \sum_{j \in \mathcal{N}_i} \mathrm{MsG}((s_i, \vec{v}_i), (s_j, \vec{v}_j), e_{ij}),$$
 (2)

$$s_i', \vec{v}_i' \coloneqq \text{UPD}((s_i, \vec{v}_i), (m_i, \vec{m}_i)),$$
 (3)

where Msg, UPD are Geometric Vector Perceptrons, a generalization of MLPs to take tuples of scalar and vector features as input and apply O(3)-equivariant non-linear updates. The overall GNN encoder is SO(3)-equivariant due to the use of reflection-sensitive input features (dihedral angles) combined with O(3)-equivariant GVP-GNN layers. Both encoder and decoder are assigned 8 GVP layers, following findings in Hsu et al. (2022b) (2022). See Appendix D for further details.

We present two different encoder architectures:

- **DynamicMPNN**: Independent encoder channels for each conformation, followed by Deep Set pooling (Zaheer et al., 2017) it is invariant to conformation order and does not add extra parameters to our model. We note that while some more expressive pooling strategies have been shown to provide marginal performance improvements, they usually come at a great cost in efficiency (Joshi et al., 2025). Only node features are updated.
- DynamicMPNN + DSS: We implement cross-attention between the encoder channels after
 each layer using a Deep Symmetric Set (DSS) (Maron et al., 2020) module, which allows
 for richer inter-conformation interactions at the cost of increased computational complexity. In the scatter/gather DSS strategy, node embeddings of all design chains are averaged,
 passed through GVP layers, and added back to the residual features of each channel. Both
 node and edge features are updated.

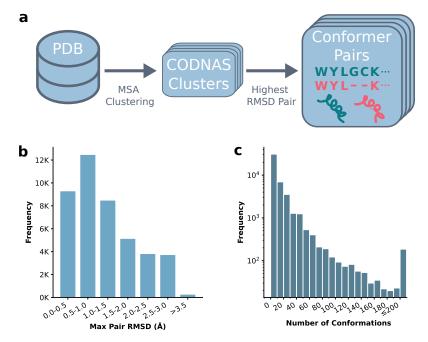


Figure 2: Multi-state protein dataset. (a) Data processing pipeline used to construct sequencealigned structure pairs. (b) Distribution of the number of conformations per CoDNaS cluster. (c) Distribution of the maximum $C\alpha$ -RMSD between pairs of structures in each CoDNaS cluster.

A major distinction from gRNAde is that DynamicMPNN models multi-chain complexes and can thus condition conformational changes on protein binders and oligomeric states. It may be possible to use this principle to engineer controllable protein conformational switches by tuning the free energy difference between folds and their binding interactions (Alberstein et al., 2022). However, this also introduces challenges in dealing with nonidentical sequences and missing residues across conformations, which we address as follows.

Alignment and pooling. To deal with nonidentical sequences and missing residues during training, structure pairs are featurized and encoded independently, and subsequently aggregated based on pairwise sequence alignments (i.e. only nongap residues are taken into account when pooling).

Multi-chain encoding and masking. Given that conformational shifts often depend on interactions, DynamicMPNN encodes the chemical environment for each conformational state (currently limited to proteins). During training, we expose the sequence information of binding partners in the encoding, only masking chains with >70% sequence similarity to the ground truth sequence of the chains of interest.

2.3 MULTI-STATE DESIGN EVALUATION

Following previous work Wang et al. (2023), we evaluate the refoldability of generated sequences rather than sequence recovery (Appendix B). Existing refoldability methods compare target structures to single conformations predicted by folding models (e.g., AlphaFold2 Jumper et al. (2021)). We argue that this approach is unsuitable for multi-state design since folding models typically predict one dominant state or interpolate between conformations rather than sampling the full conformational ensemble Lane (2023); Chakravarty et al. (2024); Saldaño et al. (2022). Furthermore, as mentioned in Section 2.1, recent DL-based protein folding models have been shown not to learn protein folding physics, but rather to compress the protein space information while preserving lower complexity motifs such as sequence segment pairing (Outeiral et al., 2022; Zhang et al., 2024; Chakravarty et al., 2024).

We propose using a template-based Alphafold 3 (AF3) framework Abramson et al. (2024): for each de novo sequence, we perform 2 AF3 runs, one with each conformational state given as template to moderately bias the model towards the target conformation. The similarity - $C\alpha$ -RMSD or TM-

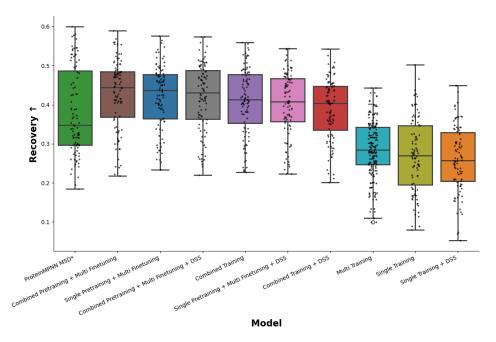


Figure 3: Sequence recovery performance across DynamicMPNN model variants and ProteinMPNN baseline on multi-state protein benchmark (n=96). Combined training approaches achieve highest performance, with models that only incorporate multi or single state training data performing poorly.

score (Zhang & Skolnick, 2004) - between predicted structures and the templates, along with AF3 confidence scores, serve as a proxy for the likelihood that the designed sequence will fold into the target structure. This framework naturally enables evaluation of refoldability across multiple conformations.

Formally, for a protein with conformational states $X = \{X_1, X_2, \dots, X_m\}$ and designed sequence Y, we define the AF3 RMSD for each target conformation X_k as:

$$AF3_{template}(Y, X_k) = RMSD(AF3(Y, X_k), X_k)$$
(4)

where $AF3_{template}(Y, X_k)$ is the structure predicted by AlphaFold3 for sequence Y when a template of X_k is provided. To account for the structural bias induced by the template, we define a normalization strategy to contextualize observed deviations:

Decoy normalization (Decoy Norm): We provide AF3 structurally dissimilar decoy structures as templates (TM-score < 0.4) using the same sequences designed and measure the resulting deviations. This control assesses whether sequences fold specifically into their targets or may fold equally well into arbitrary structures:

$$RMSD_{decoy}(Y, X_k; D) = \frac{AF3_{template}(Y, X_k)}{AF3_{template}(Y, D)}$$
(5)

where D is a decoy that is structurally dissimilar to X_k . Additionally, we measure pLDDT confidence scores to evaluate AF3 fold uncertainty. High RMSD with low pLDDT indicates poor template matching, while low RMSD with high pLDDT suggests a successful design.

3 RESULTS AND DISCUSSION

Setup. We evaluate several training strategies and model architectures to assess the impact on multistate learning. Single Training models follow the standard single state inverse folding approach, learning exclusively from single-state sequence structure pairs, similarly to ProteinMPNN Dauparas

Table 1: Refoldability performance comparison of DynamicMPNN model variants on multi-state protein design benchmark (n=96). Raw metrics show absolute performance values, while normalized metrics show performance relative to random decoy structures.

	Raw Metrics			Decoy-Normalized Metrics		
Model Variant	pLDDT ↑	$RMSD(\mathring{A})\downarrow$	TM-score ↑	pLDDT ↑	$\mathbf{RMSD}\downarrow$	TM-score ↑
Combined Training	79.28 (9.28)	4.37 (4.76)	0.774 (0.246)	1.513 (0.503)	0.200 (0.183)	7.396 (2.192)
Combined Training + DSS	78.87 (8.82)	4.38 (4.47)	0.769 (0.239)	1.532 (0.524)	0.200 (0.164)	7.384 (2.304)
Single Training	62.59 (16.84)	13.48 (13.02)	0.503 (0.350)	1.466 (0.537)	0.479 (0.339)	5.014 (3.068)

Table 2: Sequence recovery performance comparison across DynamicMPNN model variants and ProteinMPNN baseline on multi-state protein design benchmark.

Model Variant	Sequence Recovery (%) ↑		
Combined Pretraining + Multi Finetuning	42.7 (8.8)		
Single Pretraining + Multi Finetuning	42.1 (8.3)		
Combined Pretraining + Multi Finetuning + DSS	41.9 (8.6)		
Combined Training	41.0 (8.5)		
Single Pretraining + Multi Finetuning + DSS	40.3 (8.2)		
Combined Training + DSS	38.8 (7.8)		
ProteinMPNN MSD*	38.0 (11.0)		
Multi Training	28.7 (7.3)		
Single Training	27.1 (9.4)		
Single Training + DSS	26.2 (8.5)		

Table 3: Performance comparison of DynamicMPNN variants and ProteinMPNN baseline on subset (n=61) of multi-state protein design benchmark. Standard deviations shown in parentheses. Note that ProteinMPNN MSD's handling of gap tokens and missing residues (i.e., X tokens) limited the number of designs that could be refolded using AF3, necessitating separate comparison.

	Raw Metrics			Decoy-Normalized Metrics		
Model	pLDDT ↑	RMSD (Å) \downarrow	TM-score ↑	pLDDT ↑	$\mathbf{RMSD}\downarrow$	TM-score ↑
Combined Training	79.23 (7.79)	4.30 (3.50)	0.748 (0.233)	1.401 (0.411)	0.206 (0.170)	7.093 (2.034)
Combined Training + DSS	78.90 (7.68)	4.56 (3.52)	0.737 (0.225)	1.404 (0.425)	0.217 (0.170)	7.090 (2.384)
Single Training	63.10 (15.65)	13.21 (12.18)	0.477 (0.331)	1.364 (0.458)	0.508 (0.337)	4.527 (2.777)
ProteinMPNN MSD*	74.29 (11.37)		0.689 (0.244)			

et al. (2022). In Combined Training, models are presented with both single and multi-state training samples. We also explore training over multiple stages, where single-state data is provided during pretraining to teach general principles of protein geometry from abundant single-state structures before specializing on multi-state conformational pairs (Multi Finetuning). For each model, the training checkpoint with the highest sequence recovery on the multi-state validation set was selected for evaluation.

All DynamicMPNN versions were trained on either 8 A100-80GB or 8 H100-80GB GPUs using a batch size of 32 and Adam (Kingma, 2014) optimizer with learning rate 10^{-3} . Training for each stage was run until the respective validation sequence recovery (on single-state validation dataset for single training runs and multi-state for multi-state runs) plateaued or began to decline, typically after 20-50 epochs. Then, DynamicMPNN and ProteinMPNN (using Multi-state Design inference strategy) were used to sample 16 sequences for the benchmark test set of 96 paired conformations, which were each run through the AF3 pipeline separately against both target states.

Refoldability was evaluated for designed sequences of the model checkpoints with the highest sequence recovery scores by predicting the structure using AF3 with target and decoy structures as templates. Aggregated evaluation metrics were computed by averaging across all 16 sequences and both states.

DynamicMPNN outperforms existing benchmarks across multiple evaluation metrics. Our best-performing model achieves substantial improvements over baseline methods: a 25% reduction in decoy-normalized RMSD (Tab. 3) and a 12% improvement in sequence recovery (Fig.3; Tab. 2) compared to ProteinMPNN Multi-State Design (MSD). This performance gain is particularly noteworthy considering that ProteinMPNN's training dataset contains proteins within 30% sequence similarity clusters of 91 out of 96 benchmark proteins, making our improvement on this established baseline significant. We leave the retraining of ProteinMPNN on a rigorous train-test split for future work.

However, a more rigorous comparison involves our single-state trained model, which adheres to the same training-test split and thus eliminates potential data leakage concerns. Against this baseline, DynamicMPNN demonstrates substantial improvements across all evaluation metrics, providing clear evidence that explicit multi-state training is superior to single-state approaches for multi-conformational protein design tasks (Tab. 2). Our single-state trained model performs considerably worse than both DynamicMPNN variants and ProteinMPNN MSD, achieving only 63.10 pLDDT compared to 79.23 for our best model, and showing nearly three-fold higher RMSD values (13.21 Å vs 4.30 Å; Tab. 1). Some likely reasons to explain the under performance of our model checkpoint trained exclusively on single-state data is the aforementioned unfair train splitting of ProteinMPNN. Furthermore, the aggregation of different protein embeddings in the latent space prior to decoding likely creates incoherent latent representations that cannot be properly interpreted by the decoder which has only seen single-state proteins in training.

While DynamicMPNN's absolute AF3 RMSD values fall outside typical thresholds used for design screening, these results should be contextualized within our deliberately challenging benchmark data set. Our test set specifically comprises proteins exhibiting the largest documented conformational changes in the known proteome, representing an extrapolation task that extends beyond the typically modest conformational variations observed in most PDB structures and consequently in our training data. The high RMSD values reflect the inherent difficulty of designing sequences that must satisfy multiple different structural constraints simultaneously.

Combined training strategies prove most effective. Models trained exclusively on multi-state data (Multi Training) show poor performance across all metrics, highlighting the importance of leveraging abundant single-state structural data to cover a larger portion of the protein space during training. Conversely, models trained only on single-state data (Single Training) perform poorly on multi-state design as they are unable to create and decode a meaningful latent representation of the conformational changes. The Combined Training approach, which exposes models to both single-state and multi-state examples during training, achieves the optimal balance and consistently outperforms other training strategies.

The Deep Symmetric Set (DSS, 5M parameters) architecture shows little to no improvement across metrics over the simple pooling strategy (4.2 mil. parameters), suggesting that communication between conformational channels before pooling is not necessary for the model to learn meaningful multi-state embeddings and fit the data. This is in agreement with the findings of Hsu et al. (2022b), who show that even on single state inverse folding when a 1M parameter model outperforms a model 20 times larger when trained on CATH.

4 Conclusion

We present DynamicMPNN, the first explicit multi-state inverse folding model, achieving up to 25% improvement over ProteinMPNN on our multi-state benchmark. By jointly learning across conformational ensembles rather than aggregating single-state predictions, DynamicMPNN captures sequence constraints required for multiple functional conformations. This opens possibilities for engineering synthetic bioswitches, allosteric regulators, and molecular machines. It is unclear if the presented one-size-fits-all approach to multi-state design will be effective experimentally, or if specialized models for different classes of conformational changes depending on their thermodynamic complexity will be beneficial.

REFERENCES

- Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.
- Robert G. Alberstein, Aidan B. Guo, and Tanja Kortemme. Design principles of protein switches. *Current Opinion in Structural Biology*, 72:71–78, Feb 2022. doi: 10.1016/j.sbi.2021.08.004. URL https://doi.org/10.1016/j.sbi.2021.08.004. PMID: 34537489; PMCID: PMC8860883.
- Xavier I Ambroggio and Brian Kuhlman. Design of protein conformational switches. *Current Opinion in Structural Biology*, 16(4):525–530, 2006a. ISSN 0959-440X. doi: https://doi.org/10.1016/j.sbi.2006.05.014. URL https://www.sciencedirect.com/science/article/pii/S0959440X06000984. Membranes / Engineering and design.
- Xavier I. Ambroggio and Brian Kuhlman. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *Journal of the American Chemical Society*, 128(4):1154–1161, 2006b. doi: 10.1021/ja054718w. URL https://doi.org/10.1021/ja054718w. PMID: 16433531.
- Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, August 2021. ISSN 1095-9203. doi: 10.1126/science.abj8754. URL http://dx.doi.org/10.1126/science.abj8754.
- Robert B Best, Kresten Lindorff-Larsen, Mark A DePristo, and Michele Vendruscolo. Relation between native ensembles and experimental structures of proteins. *Proceedings of the National Academy of Sciences*, 103(29):10901–10906, 2006.
- N. Bordin, I. Sillitoe, V. Nallapareddy, C. Rauer, S. D. Lam, V. P. Waman, N. Sen, M. Heinzinger, M. Littmann, S. Kim, S. Velankar, M. Steinegger, B. Rost, and C. Orengo. Alphafold2 reveals commonalities and novelties in protein structure space for 21 model organisms. *Communications Biology*, 6(1):160, 2023. ISSN 2399-3642. doi: 10.1038/s42003-023-04488-9. URL https://doi.org/10.1038/s42003-023-04488-9.
- Eleonora Cerasoli, Belinda K. Sharpe, and Derek N. Woolfson. Zico: A peptide designed to switch folded state upon binding zinc. *Journal of the American Chemical Society*, 127(43):15008–15009, 2005. doi: 10.1021/ja0543604. URL https://doi.org/10.1021/ja0543604. PMID: 16248623.
- Debayan Chakravarty, Jonah W. Schafer, Elizabeth A. Chen, Daniel A. Keedy, Maximillian D. Smith, and Lauren L. Porter. Alphafold predictions of fold-switched conformations are driven by structure memorization. *Nature Communications*, 15:7296, 2024. doi: 10.1038/s41467-024-51801-z. URL https://doi.org/10.1038/s41467-024-51801-z.
- J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning-based protein sequence design using proteinmpnn. *Science*, 378(6615):49-56, 2022. doi: 10.1126/science.add2187. URL https://www.science.org/doi/abs/10.1126/science.add2187.
 - Acacia F. Dishman and Brian F. Volkman. Unfolding the mysteries of protein metamorphosis. *ACS Chemical Biology*, 13(6):1438–1446, 2018. doi: 10.1021/acschembio.8b00276. URL https://doi.org/10.1021/acschembio.8b00276. PMID: 29787234.

Alexandre Duval, Simon V. Mathis, Chaitanya K. Joshi, Victor Schmidt, Santiago Miret, Fragkiskos D. Malliaros, Taco Cohen, Pietro Liò, Yoshua Bengio, and Michael Bronstein. A hitchhiker's guide to geometric gnns for 3d atomic systems, 2024. URL https://arxiv.org/abs/2312.07511.

Casper A. Goverde, Martin Pacesa, Nicolas Goldbach, Lars J. Dornfeld, Petra E. M. Balbi, Sandrine Georgeon, Stéphane Rosset, Srajan Kapoor, Jagrity Choudhury, Justas Dauparas, Christian Schellhaas, Simon Kozlov, David Baker, Sergey Ovchinnikov, Alex J. Vecchio, and Bruno E. Correia. Computational design of soluble and functional membrane protein analogues. *Nature*, 631(8020):449–458, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07601-y. URL http://dx.doi.org/10.1038/s41586-024-07601-y.

- Thomas Hrabe, Zhanwen Li, Mayya Sedova, Piotr Rotkiewicz, Lukasz Jaroszewski, and Adam Godzik. Pdbflex: exploring flexibility in protein structures. *Nucleic Acids Research*, 44(D1): D423–D428, 11 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv1316. URL https://doi.org/10.1093/nar/gkv1316.
- Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. April 2022a. doi: 10. 1101/2022.04.10.487779. URL http://dx.doi.org/10.1101/2022.04.10.487779.
- Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), Proceedings of the 39th International Conference on Machine Learning, volume 162 of Proceedings of Machine Learning Research, pp. 8946–8970. PMLR, 17–23 Jul 2022b. URL https://proceedings.mlr.press/v162/hsu22a.html.
- Arian R. Jamasb, Alex Morehead, Chaitanya K. Joshi, Zuobai Zhang, Kieran Didi, Simon V. Mathis, Charles Harris, Jian Tang, Jianlin Cheng, Pietro Lio, and Tom L. Blundell. Evaluating representation learning on the protein structure universe. In *The Twelfth International Conference on Learning Representations*, 2024.
- Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael J. L. Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons, 2021. URL https://arxiv.org/abs/2009.01411.
- Chaitanya K. Joshi, Arian R. Jamasb, Ramon Viñas, Charles Harris, Simon V. Mathis, Alex Morehead, Rishabh Anand, and Pietro Liò. grnade: Geometric deep learning for 3d rna inverse design, 2025. URL https://arxiv.org/abs/2305.14749.
- John Jumper, Richard Evans, Alexander Pritzel, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589, 2021. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.
- Y. Kalakoti and B. Wallner. Afsample2 predicts multiple conformations and ensembles with alphafold2. *Communications Biology*, 8(1):373, 3 2025. doi: 10.1038/s42003-025-07791-9. URL https://doi.org/10.1038/s42003-025-07791-9. PMID: 40045015; PM-CID: PMC11882827.
- Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Andriy Kryshtafovych, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moult. Critical assessment of methods of protein structure prediction (casp)-round xiii. *Proteins*, 87(12): 1011–1020, 12 2019. doi: 10.1002/prot.25823. URL https://doi.org/10.1002/prot.25823. PMID: 31589781; PMCID: PMC6927249.
- Thomas J Lane. Protein structure prediction has reached the single-structure frontier. *Nature Methods*, 20(2):170–173, 2023.

Andrew Leaver-Fay, Robert Jacak, P Brian Stranges, and Brian Kuhlman. A generic program for multistate protein design. *PLoS One*, 6(7):e20937, Jul 2011. doi: 10.1371/journal.pone.0020937. URL https://doi.org/10.1371/journal.pone.0020937. PMID: 21754981; PM-CID: PMC3130737.

Sidney Lyayuga Lisanza, Jacob Merle Gershon, Samuel W. K. Tipps, Jeremiah Nelson Sims, Lucas Arnoldt, Samuel J. Hendel, Miriam K. Simma, Ge Liu, Muna Yase, Hongwei Wu, Claire D. Tharp, Xinting Li, Alex Kang, Evans Brackenbrough, Asim K. Bera, Stacey Gerben, Bruce J. Wittmann, Andrew C. McShan, and David Baker. Multistate and functional protein design using rosettafold sequence space diffusion. *Nature Biotechnology*, 43(8):1288–1298, September 2024. ISSN 1546-1696. doi: 10.1038/s41587-024-02395-w. URL http://dx.doi.org/10.1038/s41587-024-02395-w.

- Yang Liu and Brian Kuhlman. Rosettadesign server for protein design. *Nucleic Acids Research*, 34(Web Server issue): W235-W238, 7 2006. doi: 10.1093/nar/gkl163. URL https://doi.org/10.1093/nar/gkl163. PMID: 16845000; PMCID: PMC1538902.
- Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. On learning sets of symmetric elements, 2020. URL https://arxiv.org/abs/2002.08599.
- Alessandro Mirarchi, Toni Giorgino, and Gianni De Fabritiis. mdcath: A large-scale md dataset for data-driven computational biophysics. *Scientific Data*, 11:1299, 2024. doi: 10.1038/s41597-024-04140-z. URL https://doi.org/10.1038/s41597-024-04140-z.
- Alexander M. Monzon, Christian O. Rohr, Maria Silvina Fornasari, and Gustavo Parisi. Codnas 2.0: a comprehensive database of protein conformational diversity in the native state. *Database*, 2016:baw038, 2016. doi: 10.1093/database/baw038. URL https://doi.org/10.1093/database/baw038. PMID: 27022160; PMCID: PMC4809262.
- Carlos Outeiral, Daniel A Nissley, and Charlotte M Deane. Current structure predictors are not learning the physics of protein folding. *Bioinformatics*, 38(7):1881–1887, 01 2022. ISSN 1367-4803. doi: 10.1093/bioinformatics/btab881. URL https://doi.org/10.1093/bioinformatics/btab881.
- Lauren L. Porter and Loren L. Looger. Extant fold-switching proteins are widespread. *Proceedings of the National Academy of Sciences*, 115(23):5968–5973, 2018. doi: 10.1073/pnas.1800168115. URL https://www.pnas.org/doi/abs/10.1073/pnas.1800168115.
- Florian Praetorius, Philip J. Y. Leung, Maxx H. Tessmer, Adam Broerman, Cullen Demakis, Acacia F. Dishman, Arvind Pillai, Abbas Idris, David Juergens, Justas Dauparas, Xinting Li, Paul M. Levine, Mila Lamb, Ryanne K. Ballard, Stacey R. Gerben, Hannah Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Brian F. Volkman, Jeff Nivala, Stefan Stoll, and David Baker. Design of stimulus-responsive two-state hinge proteins. *Science*, 381(6659):754–760, 2023. doi: 10.1126/science.adg7731. URL https://www.science.org/doi/abs/10.1126/science.adg7731.
- Alfredo Quijano-Rubio, Hsien-Wei Yeh, Jooyoung Park, et al. De novo design of modular and tunable protein biosensors. *Nature*, 591:482–487, 2021. doi: 10.1038/s41586-021-03258-z. URL https://doi.org/10.1038/s41586-021-03258-z.
- Tadeo Saldaño, Nahuel Escobedo, Julia Marchetti, Diego Javier Zea, Juan Mac Donagh, Ana Julia Velez Rueda, Eduardo Gonik, Agustina García Melani, Julieta Novomisky Nechcoff, Martín N Salas, Tomás Peters, Nicolás Demitroff, Sebastian Fernandez Alberti, Nicolas Palopoli, Maria Silvina Fornasari, and Gustavo Parisi. Impact of protein conformational diversity on alphafold predictions. *Bioinformatics*, 38(10):2742–2748, 04 2022.
- Viktor Stein and Kirill Alexandrov. Synthetic protein switches: design principles and applications. *Trends in Biotechnology*, 33(2):101–110, 2015. ISSN 0167-7799. doi: https://doi.org/10.1016/j. tibtech.2014.11.010. URL https://www.sciencedirect.com/science/article/pii/S016777991400239X. Special Issue: Manifesting Synthetic Biology.

Giulio Tesei, Anders Irbäck Trolle, Niels Jonsson, Johannes Betz, Frederik Elberling Knudsen, Francesco Pesce, Kristoffer E. Johansson, and Kresten Lindorff-Larsen. Conformational ensembles of the human intrinsically disordered proteome. *Nature*, 626(8000):897–904, Feb 2024. doi: 10.1038/s41586-023-07004-5. URL https://doi.org/10.1038/s41586-023-07004-5. PMID: 38297118.

- Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*, 33(1):2-8, 2008. ISSN 0968-0004. doi: https://doi.org/10.1016/j.tibs.2007.10.003. URL https://www.sciencedirect.com/science/article/pii/S096800040700285X.
- Yann Vander Meersche, Gabriel Cretin, Aria Gheeraert, Jean-Christophe Gelly, and Tatiana Galochkina. Atlas: protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research*, 52(D1):D384–D392, 11 2023. ISSN 0305-1048. doi: 10.1093/nar/gkad1084. URL https://doi.org/10.1093/nar/gkad1084.
- Chuanrui Wang, Bozitao Zhong, Zuobai Zhang, Narendra Chaudhary, Sanchit Misra, and Jian Tang. Pdb-struct: A comprehensive benchmark for structure-based protein design. *arXiv* preprint *arXiv*:2312.00080, 2023.
- Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL http://dx.doi.org/10.1038/s41586-023-06415-8.
- Kathy Y. Wei, Dimitar Moschidi, Matthew J. Bick, Santrupti Nerli, Andrew C. McShan, Lauren P. Carter, Po-Ssu Huang, Daniel A. Fletcher, Nikolaos G. Sgourakis, Scott E. Boyken, and David Baker. Computational design of closely related proteins that adopt two well-defined but structurally divergent folds. *Proceedings of the National Academy of Sciences*, 117(13):7208–7215, 2020. doi: 10.1073/pnas.1914808117. URL https://doi.org/10.1073/pnas.1914808117. PMID: 32188784; PMCID: PMC7132107.
- Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.
- J. Z. Zhang, H. W. Yeh, A. C. Walls, et al. Thermodynamically coupled biosensors for detecting neutralizing antibodies against sars-cov-2 variants. *Nature Biotechnology*, 40:1336–1340, 2022. doi: 10.1038/s41587-022-01280-8. URL https://doi.org/10.1038/s41587-022-01280-8.
- Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.
- Zhidian Zhang, Hannah K. Wayment-Steele, Garyk Brixi, Haobo Wang, Dorothee Kern, and Sergey Ovchinnikov. Protein language models learn evolutionary statistics of interacting sequence motifs. *Proceedings of the National Academy of Sciences*, 121(45):e2406285121, 2024. doi: 10.1073/pnas.2406285121. URL https://www.pnas.org/doi/abs/10.1073/pnas.2406285121.

A APPENDIX

A.1 LLM ACKNOWLEDGEMENT

The authors acknowledge that they have used LLMs in the process of paper writing for style suggestions and proofreading.

B SUPPLEMENTARY RESULTS

B.1 DYNAMICMPNN PRELIMINARY DESIGNS

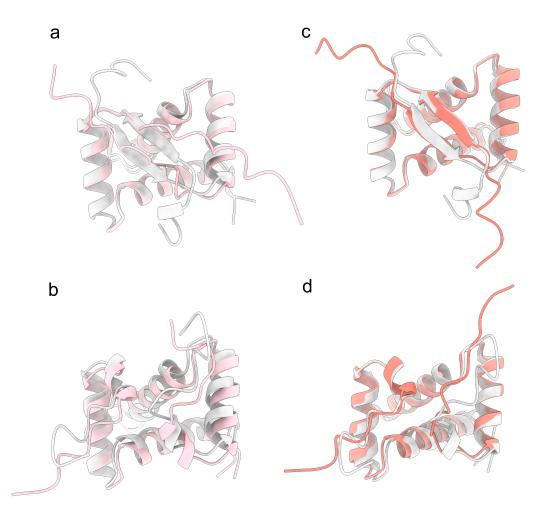


Figure 4: Switch Arc protein case study. (a, b) ProteinMPNN and (c,d) DynamicMPNN best design structure prediction (pink and salmon, respectively) against both Arc states from PDB ID: 1BDT and 1QTG respectively (grey). The DynamicMPNN design recapitulates the beta sheet fold (c), but the ProteinMPNN design does not (a).

C DATASET DETAILS

To construct the dataset, we obtained 46,033 Multiple Sequence Alignment (MSA) clusters at $\geq 95\%$ local sequence similarity from the latest version of CoDNaS (v2025) (Monzon et al., 2016), including NMR model structures. Importantly, the CoDNaS dataset prevents potential errors created by mixing different homologues in the same cluster by enforcing the same UniProt ID for all custer members. All available conformations of a protein are included - as different experimental conditions and sequence variations can reveal distinct thermodynamic states of the same protein (Best et al., 2006). While clusters contain varying numbers of conformations (Fig. 2b) we constructed our dataset using only pairs of chains from one or two PDB entries that have the largest RMSD from each cluster (Fig. 2c). Chosen pairs represent the most distinct conformational states. We will include conformational information of the whole cluster in future work.

While other inverse folding models (Hsu et al., 2022b) saw improved performance when trained on orders of magnitude of more protein structures from AFDB, previous studies have found that the

majority of high-confidence structures in AFDB map to known CATH superfamilies (Bordin et al., 2023), and that AlphaFold struggles in predicting alternative states (Chakravarty et al., 2024). We therefore decided against including AFDB structures in our training set.

D MODEL DETAILS

D.1 FEATURISATION SCHEME

We use a similar featurisation scheme as in (Jamasb et al., 2024). Node scalar features are transformer-like positional encoding in a 16-dimensional array; backbone dihedral angles $\phi, \psi, \omega \in \mathbb{R}^6$; the virtual torsion and virtual bond angle $\kappa, \alpha \in \mathbb{R}^4$. Node vector features are position vectors of C_α , $\hat{x_i} \in \mathbb{R}^3$. Edge scalar features are established via k-NN (k=16) and the edge length expressed in 32 Radial Basis Functions, $e_{RBF} \in \mathbb{R}^{32}$, as well as the length of the edge itself. Edge vector features are edge directional unit vectors for both directions $v_{eij}^{\tilde{c}} = \hat{x_i} - \hat{x_j}$. To further prevent overfitting on crystallisation artifacts, random Gaussian noise ($\bar{x} = 0, \sigma = 0.1\text{Å}$) was added to the coordinates (Dauparas et al., 2022).

D.2 MULTI-STATE GNN

DynamicMPNN processes one or multiple protein backbone graphs via a multi-state GNN encoder (Joshi et al., 2025). Overall, DynamicMPNN's encoder is equivariant to 3D roto-translation of coordinates as well as ordering of the states in its input. Encoding is followed by pooling node features across states, which is invariant to the ordering of the states, and autoregressive sequence decoding.

When representing conformational ensembles as a multi-graph, each node feature tensor contains three axes: (#nodes, #conformations, feature channels). Multi-state GNN's encode multi-graphs by performing message passing on the multi-graph adjacency to *independently* process each conformer, while maintaining permutation equivariance of the updated feature tensors along both the first (#nodes) and second (#conformations) axes.

D.3 GEOMETRIC VECTOR PERCEPTRON LAYERS

Geometric Vector Perceptrons (GVPs) (Jing et al., 2021) are a generalization of MLPs to take tuples of scalar and vector features as input and apply O(3)-equivariant non-linear updates. GVP GNN layers process scalar and vector features on separate channels to maintain equivariance. The node scalars $\mathbf{s}_i \in \mathbb{R}^{k \times m}$, node vectors $\tilde{\mathbf{v}}_i \in \mathbb{R}^{k \times m' \times 3}$, and edge scalars \mathbf{e}_{ij} and vectors $\tilde{\mathbf{e}}_{ij}$ communicate through a message passing operation:

$$\mathbf{m}_{i}, \tilde{\mathbf{m}}_{i} := \sum_{j \in N_{i}} \text{GVP}\left((\mathbf{s}_{i}, \tilde{\mathbf{v}}_{i}), (\mathbf{s}_{j}, \tilde{\mathbf{v}}_{j}), \mathbf{e}_{ij}, \tilde{\mathbf{e}}_{ij}\right), \qquad \text{(Message & aggregate steps)}$$
 (6)

$$\mathbf{s}'_{i}, \tilde{\mathbf{v}'}_{i} := \text{GVP}\left((\mathbf{s}_{i}, \tilde{\mathbf{v}}_{i}), (\mathbf{m}_{i}, \tilde{\mathbf{m}}_{i})\right). \tag{Update step}$$

The overall GNN encoder is SO(3)-equivariant due to the use of reflection-sensitive input features (dihedral angles) combined with O(3)-equivariant GVP-GNN layers.

D.4 CONFORMATION ORDER-INVARIANT POOLING

After using message passing layers that are conformation order-equivariant, we add a conformation order-invariant head, which performs average pooling across the conformation channel of the scalar and vector feature tensors, similar to Joshi et al. (2025) (2025): $\mathbf{S} \in \mathbb{R}^{n \times k \times m}$ and $\tilde{\mathbf{V}} \in \mathbb{R}^{n \times k \times m' \times 3}$ to $\mathbf{S} \in \mathbb{R}^{n \times m}$ and $\tilde{\mathbf{V}} \in \mathbb{R}^{n \times m' \times 3}$, where n is the sequence length, k is the number of backbones, k is the number of scalar features, and k is the number of vector features. The only pooling strategy used in this work is the pooling of the maximum RMSD pair of chains - therefore k = 2 - although more pooling strategies for homo-oligomers can be used, such as equal averaging of all chains to be inverse folded in the selected PDB entries.