Multi-state Protein Design with DynamicMPNN

Alex Abrudan [*1] **Sebastian Pujalte Ojeda** [*1] **Chaitanya K. Joshi** [2] **Matthew Greenig** [1] **Felipe Engelberger** [3]
Alena Khmelinskaia [4] **Jens Meiler** [3] **Michele Vendruscolo** [1] **Tuomas P. J. Knowles** [1]

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 initial guess, DynamicMPNN outperforms ProteinMPNN by up to 13% on structure-normalized RMSD across our challenging multi-state protein benchmark.

1. Introduction

The conformational diversity of proteins underlies biological functions such as enzyme catalysis, protein recognition, allostery, protein evolution modulation or human disease (Monzon et al., 2016). Of particular importance is the design of bioswitches, with huge implications in biotechnology, *e.g.* in creating artificial 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 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 are known to take two main functional states (Dishman & Volkman, 2018) and a finite number of conformations (see Discussion), thus this study is not aimed at the design of protein sequences adopting continuous conformational landscapes (e.g. intrinsically disordered proteins (Tompa & Fuxreiter, 2008)).

Multi-state protein design was first achieved through rational design and physics-based RosettaDesign (Liu & Kuhlman, 2006) of: 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 equilibrium between states (Zhang et al., 2022; Quijano-Rubio et al., 2021; Praetorius et al., 2023). Remarkably, Praetorius et al. (2023) additionally employed ProteinMPNN Multi-state Design (ProteinMPNN-MSD) (Dauparas et al., 2022)—an inference strategy for extending ProteinMPNN to multiple states by averaging the logits of two independent single-state ProteinMPNN embeddings during the decoding step—to design proof-of-concept multi-helix hinge proteins.

Praetorius et al. (2023) however report an overall very low experimental success rate: 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 with the corresponding target peptide. This and the absence of any reported *de novo* designs for full natural multi-state backbones is proof that multi-state protein design lags behind single-state design (Appendix E) due to limited conformational datasets, weak benchmarks, and the poor record of folding models in predicting alternative states (Chakravarty et al., 2024) - which made for poor self-consistency filters.

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 pro-

[*]Equal contribution [1]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK [2]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK [3]Institute for Drug Discovery, Leipzig University, Leipzig, Germany [4]Department of Chemistry, Ludwig-Maximilans-University, Munich, Germany. Correspondence to: Alex Abrudan, Sebastian Pujalte Ojeda <aca41@cam.ac.uk, sp2120@cam.ac.uk>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

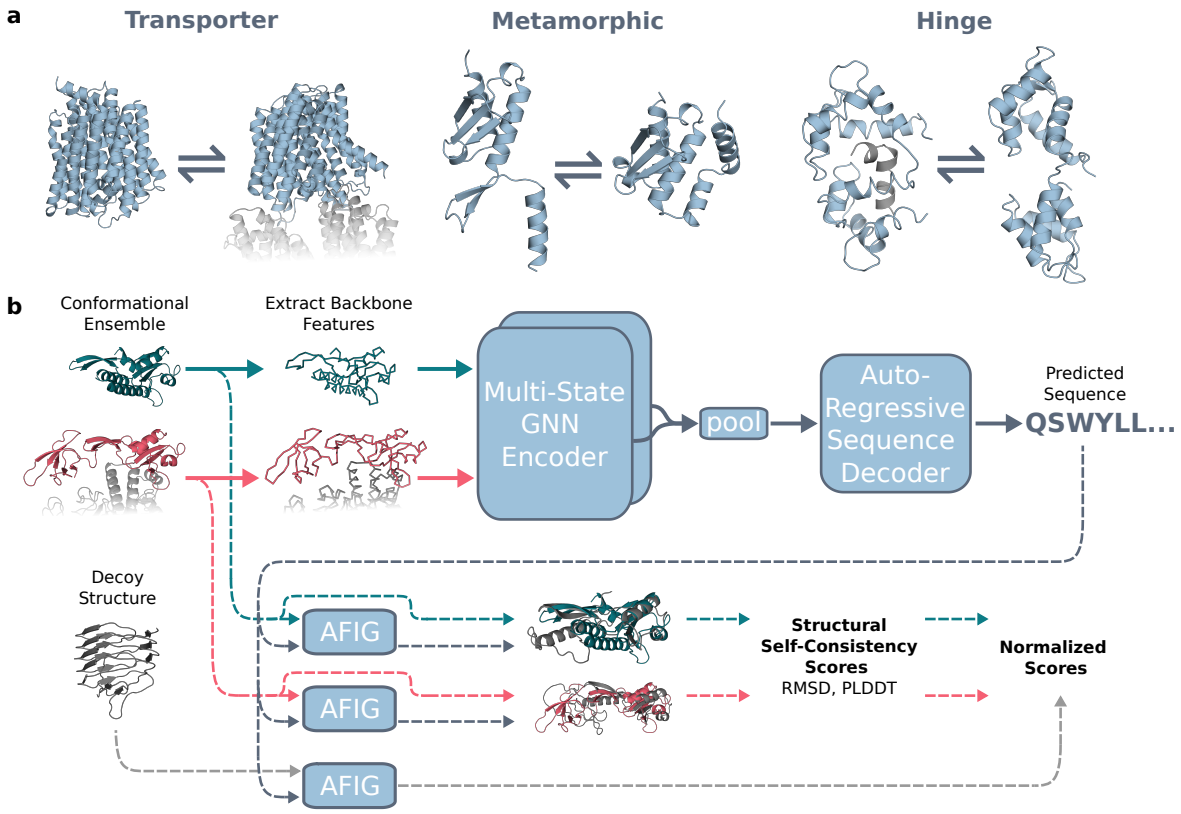


Figure 1: **DynamicMPNN for multi-state protein design.** (a) Examples of proteins with multiple conformational states: transporters in closed and open states (PDB: 6NC7, 6NC9), metamorphic protein with alternative folds (PDB: 4QHH, 4QHF) and hinges showing domain movement (PDB: 5D0W, 1CFC). (b) Schematic 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 Initial Guess (AFIG), which evaluates structural self-consistency by initializing AlphaFold2 backbone frames on target structure coordinates and measuring the deviations between predicted and target structures, with decoy structures serving as negative controls.

teins with multiple conformations using the PDB and CoDNaS (Monzon et al., 2016) databases, and evaluate the method on 94 biologically relevant metamorphic, hinge, and transporter proteins.

- We propose a multi-state self-consistency metric and benchmark based on *AlphaFold initial guess* (AFIG) (Roney & Ovchinnikov, 2022; Bennett et al., 2023) which we argue to be superior to sequence recovery.
- DynamicMPNN improves performance over ProteinMPNN (Dauparas et al., 2022) on AFIG by up to 13% on RMSD and 3% on pLDDT self-consistency values.

2. The DynamicMPNN pipeline

2.1. Protein multi-conformational dataset

While over 750,000 individual protein chains (sequence-structure pairs) are available in the PDB, multi-conformational data is far more scarce with only 11,833 NMR-derived protein ensembles covering just 21% of CATH superfamilies. To overcome this limitation, we exploit the redundancy of chains with very high sequence similarity (≥95%) in the PDB to build a multi-conformational dataset of 46,033 conformer pairs that expands coverage to 75% of CATH superfamilies (Fig. 2a).

For dataset splitting, we first curate a benchmark set from five previous studies of proteins with large conformational changes: 92 metamorphic proteins (Porter & Looger, 2018), 91 apo-holo proteins (Saldaño et al., 2022), the OC23/OC85 open-closed datasets, and 20 transporter proteins (Kalakoti

& Wallner, 2025). The 94 highest RMSD pairs were assigned to the test set and the next 100 to the validation set. Training clusters were filtered to exclude any with TM-score > 0.4 (Zhang & Skolnick, 2004) to test/validation structures, preventing structural similarity leakage and yielding a final training set of 44,243 conformer pairs. See Appendix B 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 $X \in \mathbb{R}^{n\times3\times3}$ represents a protein backbone with n residues, and $Y = (y_1, \ldots, y_n)$ 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. Instead, DynamicMPNN learns the joint conditional distribution of $p(Y|X_1, \ldots, X_m)$ directly through autoregressive sequence generation (see Appendix C), where $\{X_1, \ldots, X_m\}$ represent distinct conformations; thus capturing sequence compatibility across all states simultaneously within a single model.

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

Our architecture is based on gRNAde (Joshi et al., 2025), a multi-state GNN for RNA inverse folding. For both the encoder and the decoder, we employ SE(3)-equivariant Geometric Vector Perception (Jing et al., 2021) layers which maintain computational efficiency through edge sparsity. Both encoder and decoder were assigned 8 GVP layers, following Hsu et al. (2022). See Appendix D for further details.

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 chain of interest.

Conformation order-invariant pooling. We employ Deep Set pooling (Zaheer et al., 2017) over the conformations as 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).

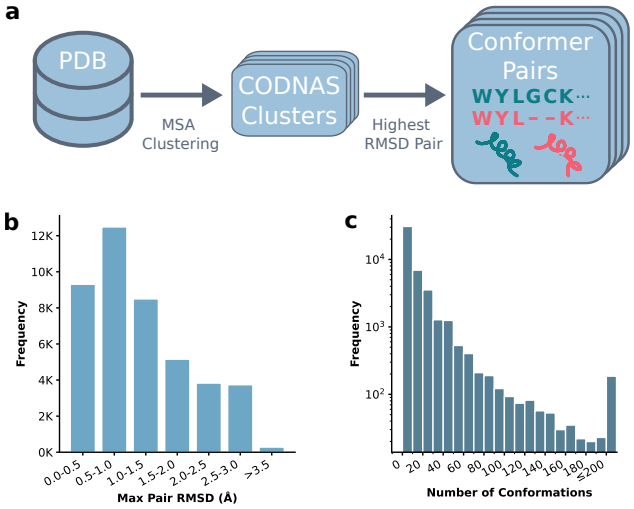


Figure 2: **Multi-state protein dataset.** (a) Data processing pipeline used to construct sequence-aligned structure pairs. (b) Distribution of the number of conformations per CoDNaS cluster. (c) Distribution of the maximum Cα-RMSD between pairs of structures in each CoDNaS cluster.

2.3. Multi-state design evaluation

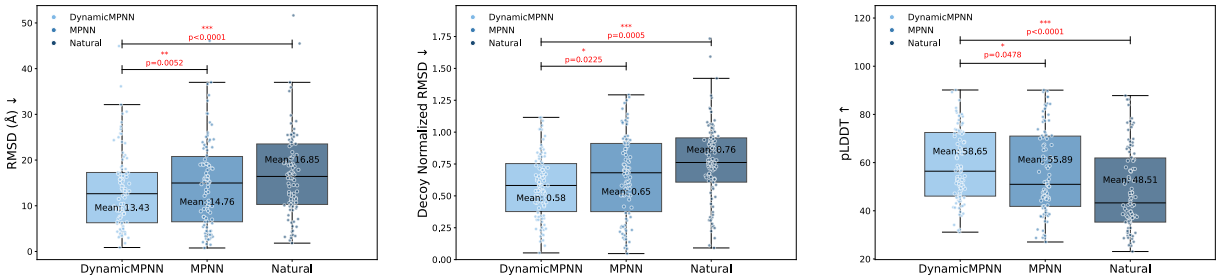
Following previous work (Wang et al., 2023), we evaluate the refoldability of generated sequences rather than sequence recovery (Appendix A). Existing refoldability methods compare target structures to single conformations predicted by folding models (e.g., AlphaFold2 (Jumper et al., 2021)). We argue 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).

We propose using the AlphaFold initial guess (AFIG) framework (Roney & Ovchinnikov, 2022; Bennett et al., 2023), originally suggested to rank candidate protein structures, to evaluate multi-state designs (Praetorius et al., 2023). AFIG initializes the AF2 reference frames on the target backbone coordinates during inference to bias the model towards the target conformation. The similarity (Cα-RMSD) between predicted and target structures, along with AF2 confidence scores, serves 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 $X = \{X_1, X_2, \ldots, X_m\}$ and designed sequence Y, we define

Table 1: **Multi-state design benchmark.** Comparison of average sequence design performance between DynamicMPNN and ProteinMPNN (multi-state design inference strategy) on a test set of 94 common pairs.

Model	RMSD ↓					pLDDT ↑		
	Best Paired (Å)	Best Paired (Struct Norm)	Best Paired (Decoy Norm)	Best Single (Å)	All Avg. (Å)	Best Paired	Best Single	All Avg.
DynamicMPNN	**13.43**	**3.86**	**0.58**	**12.76**	17.39	**58.65**	**59.31**	50.22
ProteinMPNN-MSD	14.76	4.47	0.65	13.99	17.24	55.89	57.74	50.18
Natural sequence	16.85	5.51	0.76	16.85	**16.85**	48.51	48.51	48.51



(a) (b) (c)

Figure 3: **Benchmark performance comparison across design approaches.** All plots compare the best designs of DynamicMPNN (Our model, left) and ProteinMPNN MSD (middle), as well as Natural WT sequences (right). DynamicMPNN demonstrates statistically significant improvements across all metrics compared to ProteinMPNN MSD and natural sequences. Statistical significance is indicated by p-values calculated using the Wilcoxon signed-rank test due to non-normal distribution of the data. (a) Raw AFIG RMSD (Å) values (Table 1, column 2) measuring structural deviation between ground-truth backbones and AFIG-predicted structures, following Equation 1. (b) Decoy-normalized RMSD values (Table 1, column 4) showing RMSD divided by AFIG RMSD of random decoy structures, following Equation 3. (c) pLDDT confidence scores (Table 1, column 7) from AFIG predictions.

the AFIG RMSD for each target conformation X_k as:

$$\text{AFIG}(Y, X_k) = \text{RMSD}(\text{AF2}(Y, X_k), X_k) \quad (1)$$

where $\text{AF2}(Y, X_k)$ is the structure predicted by AlphaFold2 for sequence Y when backbone frames are initialized on coordinates of structure X_k. We define two normalization criteria to contextualize observed deviations:

Structure normalization (Struct Norm): We normalize AFIG RMSD by the maximum RMSD between target conformations to capture task difficulty; i.e. designing proteins with large conformational shifts is inherently more difficult than small rearrangements:

$$\text{RMSD}_{\text{struct}}(Y, X_k; X) = \frac{\text{AFIG}(Y, X_k)}{\max_{i,j} \text{RMSD}(X_i, X_j)} \quad (2)$$

Decoy normalization (Decoy Norm): We initialize AF2 with structurally dissimilar decoy structures (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:

$$\text{RMSD}_{\text{decoy}}(Y, X_k; D) = \frac{\text{AFIG}(Y, X_k)}{\text{AFIG}(Y, D)} \quad (3)$$

where D is a decoy that is structurally dissimilar to X_k.

Additionally, we measure pLDDT confidence scores to evaluate AFIG 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. DynamicMPNN was trained for 50 epochs, and an AFIG evaluation on the validation set was performed every third epoch to select the best-performing model. Then, DynamicMPNN and ProteinMPNN (using Multi-state Design

inference strategy) were used to sample 16 sequences for the benchmark set of 94 paired conformations, which were each run through the AFIG pipeline separately against both target states.

Aggregated evaluation metrics were computed as follows. (**Best Single**) The minimum RMSD and maximum pLDDT values across all sets of 16 sequences are selected from among both states. (**Best Paired**) The RMSD and pLDDT values are averaged over the 2 states for each sequence, then the best RMSD and pLDDT across all 16 averaged values are selected. Struct Norm and Decoy Norm RMSD values are then calculated for each protein. (**All Avg.**) The RMSD and pLDDT values are averaged across all 16 sequences and both states. The natural sequences were processed identically and resulting metrics were averaged over the benchmark set and reported in Table 1.

DynamicMPNN outperforms ProteinMPNN. We see this across all metrics, with the largest gains in the "Best Paired" metrics of up to 13% - arguably the most important set of metrics for experimental design, as only the best sequences are selected for *in vitro* validation. Visual structural inspection of both states (Fig. 4) confirms successful design of the metamorphic Switch Arc protein by DynamicMPNN, compared to failure to recapitulate the central beta sheet fold by the best scoring ProteinMPNN sequence. DynamicMPNN also outperforms ProteinMPNN in the "Best Single" metrics, which is surprising since ProteinMPNN was trained on single-state design tasks. It should be noted that ProteinMPNN training dataset contains proteins within the sequence clusters of 84 out of 94 benchmark proteins. This represents a substantial data leakage advantage for ProteinMPNN, while DynamicMPNN was specifically trained to exclude any sequence similarity to the test set. We plan on training a single-state version of DynamicMPNN to more fairly demonstrate the advantage of the multi-state design approach. While DynamicMPNN AFIG RMSD values are still high, it is important to contextualize these results within our deliberately challenging benchmark: our test set comprises proteins with the largest documented conformational changes in the known proteome. This can be seen as an extrapolation beyond the overall modest conformational changes observed in most PDB structures and hence our training data.

Natural sequence vs. multi-state designs. While the best-of-16 sequence from DynamicMPNN and ProteinMPNN (Dauparas et al., 2022) outperform natural sequences, the average metrics are slightly worse than the natural sequences (Fig. 5). Dauparas et al. (2022) have presented a similar finding with respect to single-state protein design, where the ProteinMPNN designed sequences showed improved AlphaFold refoldability over the natural sequences. A possible explanation is that while both inverse folding models are trained mostly to optimize the sequence for structure stability, evolution has undergone a multi-objective optimization for different functions, which might have prevented it from reaching the energy minima for those structures. Additionally, due to graph sparsity, DynamicMPNN and ProteinMPNN emphasize local interactions, making designed sequences more foldable by AFIG - since AFIG does not use any MSA, it poorly captures long-range interactions present in natural sequences (Zhang et al., 2024).

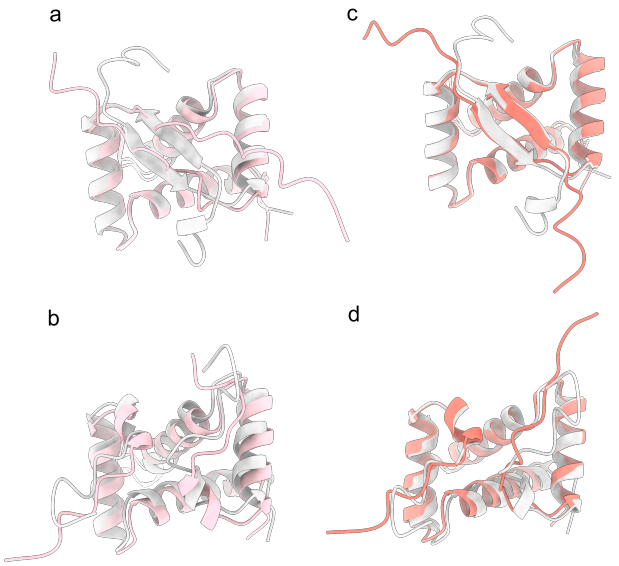


Figure 4: **Switch Arc protein case study.** (a, b) ProteinMPNN and (c,d) DynamicMPNN best design structure prediction (pink and salmon, respectively) against both Arc states from PDB ID: 1BDT and 1QTG respectively (grey). The DynamicMPNN design recapitulates the beta sheet fold (c), but the ProteinMPNN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

4. Conclusion

We present DynamicMPNN, the first explicit multi-state inverse folding model, achieving up to 13% 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.

Software and Data

We provide the code for our model in the following repository: `github.com/Alex-Abrudan/DynamicMPNN`.

Acknowledgements

The authors would like to thank Adam Broerman, Rohith Krishna, Kai Yi, Simon Mathis, Hannah Han, Vladimir Radenkovic, Gustavo Parisi, Sebastian Ahnert, and Pietro Liò for insightful comments and discussions. CKJ was supported by the A*STAR Singapore National Science Scholarship (PhD). AK was supported by the Graduate School of Quantitative Biosciences Munich (QBM).

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Ambroggio, X. I. and Kuhlman, B. 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.

Ambroggio, X. I. and Kuhlman, B. 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.

Bennett, N. R., Coventry, B., Goreshnik, I., et al. Improving de novo protein binder design with deep learning. *Nature Communications*, 14:2625, 2023. doi: 10.1038/s41467-023-38328-5. URL https://doi.org/10.1038/s41467-023-38328-5.

Best, R. B., Lindorff-Larsen, K., DePristo, M. A., and Vendruscolo, M. Relation between native ensembles and experimental structures of proteins. *Proceedings of the National Academy of Sciences*, 103(29):10901–10906, 2006.

Cerasoli, E., Sharpe, B. K., and Woolfson, D. N. 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.

Chakravarty, D., Schafer, J. W., Chen, E. A., Keedy, D. A., Smith, M. D., and Porter, L. L. 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.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A. K., King, N. P., and Baker, D. 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.

Dishman, A. F. and Volkman, B. F. 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.

Hrabe, T., Li, Z., Sedova, M., Rotkiewicz, P., Jaroszewski, L., and Godzik, A. 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.

Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A., and Rives, A. Learning inverse folding from millions of predicted structures. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S. (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 2022. URL https://proceedings.mlr.press/v162/hsu22a.html.

Jamasb, A. R., Morehead, A., Joshi, C. K., Zhang, Z., Didi, K., Mathis, S. V., Harris, C., Tang, J., Cheng, J., Lio, P., and Blundell, T. L. Evaluating representation learning on the protein structure universe. In *The Twelfth International Conference on Learning Representations*, 2024.

Jing, B., Eismann, S., Suriana, P., Townshend, R. J. L., and Dror, R. Learning from protein structure with geometric vector perceptrons, 2021. URL https://arxiv.org/abs/2009.01411.

Joshi, C. K., Jamasb, A. R., Viñas, R., Harris, C., Mathis, S. V., Morehead, A., Anand, R., and Liò, P. grnade: Geometric deep learning for 3d rna inverse design, 2025. URL https://arxiv.org/abs/2305.14749.

Jumper, J., Evans, R., Pritzel, A., 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.

Kalakoti, Y. and Wallner, B. 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; PMCID: PMC11882827.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. 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.

Lane, T. J. Protein structure prediction has reached the single-structure frontier. *Nature Methods*, 20(2):170–173, 2023.

Liu, Y. and Kuhlman, B. 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.

Mirarchi, A., Giorgino, T., and De Fabritiis, G. 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.

Monzon, A. M., Rohr, C. O., Fornasari, M. S., and Parisi, G. 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.

Porter, L. L. and Looger, L. L. 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.

Praetorius, F., Leung, P. J. Y., Tessmer, M. H., Broerman, A., Demakis, C., Dishman, A. F., Pillai, A., Idris, A., Juergens, D., Dauparas, J., Li, X., Levine, P. M., Lamb, M., Ballard, R. K., Gerben, S. R., Nguyen, H., Kang, A., Sankaran, B., Bera, A. K., Volkman, B. F., Nivala, J., Stoll, S., and Baker, D. 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.

Quijano-Rubio, A., Yeh, H.-W., Park, J., 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.

Roney, J. P. and Ovchinnikov, S. State-of-the-art estimation of protein model accuracy using alphafold. *Phys. Rev. Lett.*, 129:238101, Nov 2022. doi: 10.1103/PhysRevLett.129.238101. URL https://link.aps.org/doi/10.1103/PhysRevLett.129.238101.

Saldaño, T., Escobedo, N., Marchetti, J., Zea, D. J., Mac Donagh, J., Velez Rueda, A. J., Gonik, E., García Melani, A., Novomisky Nechcoff, J., Salas, M. N., Peters, T., Demitroff, N., Fernandez Alberti, S., Palopoli, N., Fornasari, M. S., and Parisi, G. Impact of protein conformational diversity on alphafold predictions. *Bioinformatics*, 38(10):2742–2748, 04 2022.

Stein, V. and Alexandrov, K. 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.

Sumida, K. H., Núñez-Franco, R., Kalvet, I., Pellock, S. J., Wicky, B. I. M., Milles, L. F., Dauparas, J., Wang, J., Kipnis, Y., Jameson, N., Kang, A., De La Cruz, J., Sankaran, B., Bera, A. K., Jiménez-Osés, G., and Baker, D. Improving protein expression, stability, and function with proteinmpnn. *Journal of the American Chemical Society*, 146(3):2054–2061, 2024. doi: 10.1021/jacs.3c10941. URL https://doi.org/10.1021/jacs.3c10941. PMID: 38194293.

Tompa, P. and Fuxreiter, M. 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.

Vander Meersche, Y., Cretin, G., Gheeraert, A., Gelly, J.-C., and Galochkina, T. 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.

Wang, C., Zhong, B., Zhang, Z., Chaudhary, N., Misra, S., and Tang, J. Pdb-struct: A comprehensive benchmark for structure-based protein design. *arXiv preprint arXiv:2312.00080*, 2023.

Watson, J. L., Juergens, D., Bennett, N. R., et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620:1089–1100, 2023. doi: 10.1038/s41586-023-06415-8. URL https://doi.org/10.1038/s41586-023-06415-8.

Wei, K. Y., Moschidi, D., Bick, M. J., Nerli, S., McShan, A. C., Carter, L. P., Huang, P.-S., Fletcher, D. A., Sgourakis, N. G., Boyken, S. E., and Baker, D. 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.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., and Smola, A. J. Deep sets. *Advances in neural information processing systems*, 30, 2017.

Zhang, J. Z., Yeh, H. W., Walls, A. C., 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.

Zhang, Y. and Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.

Zhang, Z., Wayment-Steele, H. K., Brixi, G., Wang, H., Kern, D., and Ovchinnikov, S. 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.

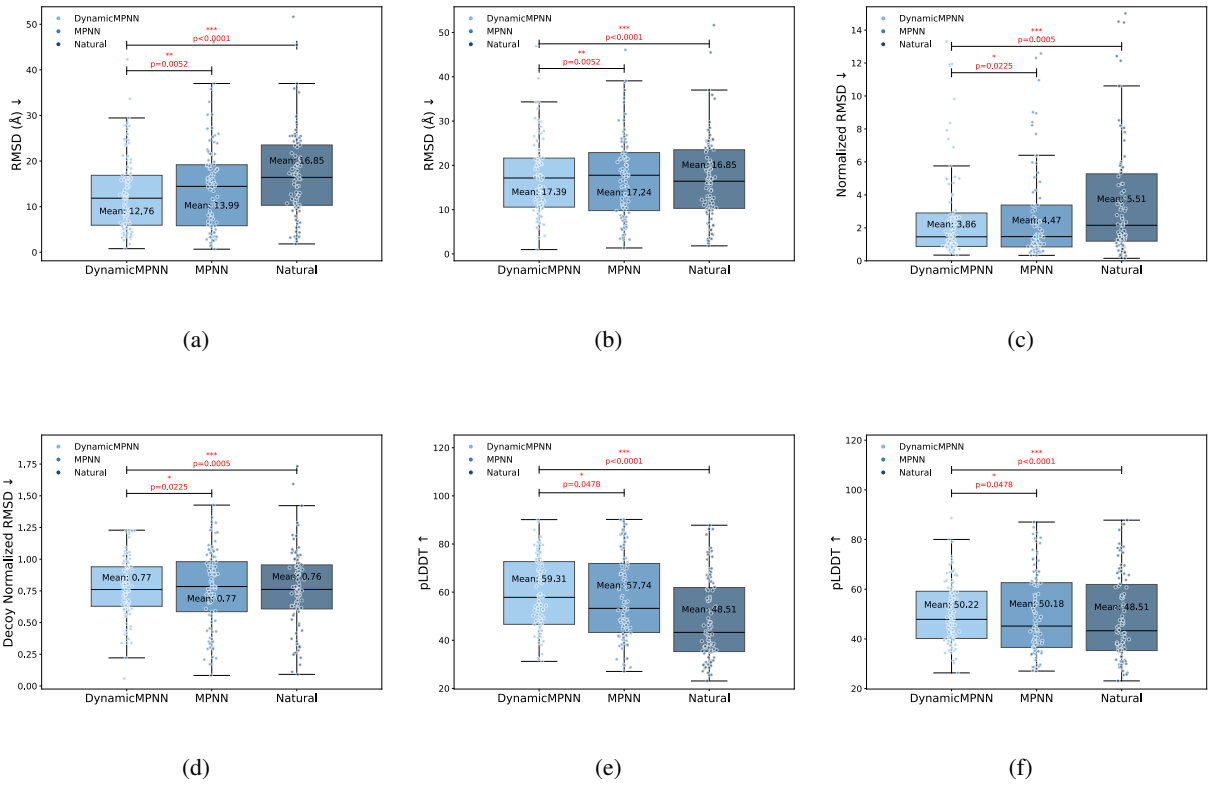
A. Supplementary Results

A.1. Additional Performance Analysis

For our statistical analysis, we first evaluated the normality of the paired differences between model performances (RMSD values) using the Shapiro-Wilk test. The test results ($W = 0.913$, $p < 0.001$) strongly rejected the null hypothesis of normality for the distribution of differences.

Therefore, we selected the Wilcoxon signed-rank test as the appropriate non-parametric alternative to the paired t-test for assessing statistical significance between model performances. This test makes no assumptions about the underlying distribution shape, making it robust for our analysis of paired protein structure comparisons across the different design approaches.

Fig. 5 (a), (e) shows how DynamicMPNN maintains its superiority for the single-state best sequence design for the RMSD and pLDDT metrics. However, averaging the performance of all designed sequences erases the advantage of DynamicMPNN over ProteinMPNN and natural sequences (Fig. 5 (b), (c), (d), (f)).



(a) (b) (c)

(d) (e) (f)

Figure 5: **Extended performance metrics across design approaches.** All box plots compare DynamicMPNN (Our model, left), ProteinMPNN MSD (middle), and Natural sequences (right). Statistical significance between DynamicMPNN and other models is indicated by p-values calculated using the Wilcoxon signed-rank test due to non-normal distribution of the data ($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). (a) Best Single RMSD values (Table 1, column 5) measuring structural deviation for the best individual conformation of each protein. (b) All Avg. RMSD values (Table 1, column 6) showing average structural deviation across all 16 sampled sequences for each protein. (c) Best Paired normalized RMSD values ((Table 1, column 3)) showing the 2-state AFIG RMSD divided by the WT inter-state RMSD according to Equation 2 (d) All Avg. decoy-normalized RMSD values (not shown in table) averaged across all sampled sequences. (e) Best Single pLDDT scores (Table 1, column 8) showing confidence in AFIG prediction for the best individual conformation. (f) All Avg. pLDDT scores (Table 1, column 9) showing the average prediction confidence across all sampled sequences.

While ProteinMPNN outperforms DynamicMPNN on sequence recovery and perplexity (Table 2), Wang et al. (2023) argue

that refoldability - emphasised in the present work - is more direct, structure-grounded, and superior to sequence recovery and perplexity when it comes to evaluating inverse folding models.

Table 2: Comparison of sequence recovery and perplexity between models, averaged over all sequences

Model	Sequence Recovery ↑	Perplexity ↓
DynamicMPNN	0.29	10.2
ProteinMPNN MSD	0.38	5.12

B. 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. The 95% similarity threshold accounts for soluble tags or point mutations/alterations performed in different experiments so that 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). Since the proteins targeted by this project explore two functional states, this choice aims to minimize alignment artifacts and maximize conformational signal during training - chosen pairs represent the most distinct conformational states. We will include conformational information of the whole cluster in future work.

C. Multi-conformational inverse folding

Given a set of target conformations $X_1, \ldots, X_m$ we seek to model the conditional probability distribution over amino acid sequences Y that can adopt all specified structures. The challenge lies in learning a sequence distribution that simultaneously satisfies multiple structural constraints. We decompose this joint conditional probability using the autoregressive factorization:

$$p(Y|X_1, ..., X_m) = \prod_{i=1}^{n} p(y_i|y_{i-1}, ..., y_1; X_1, ..., X_m) \tag{4}$$

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

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, $\tilde{x_i} \in \mathbb{R}^3$. **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 $\tilde{v_{e^{ij}}} = \tilde{x_i} - \tilde{x_j}$. To further prevent overfitting on crystallisation artifacts, random Gaussian noise ($\bar{x} = 0, \sigma = 0.1$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:

$$\mathbf{m}_i, \tilde{\mathbf{m}}_i := \sum_{j \in N_i} \text{GVP}\left((\mathbf{s}_i, \tilde{\mathbf{v}}_i), (\mathbf{s}_j, \tilde{\mathbf{v}}_j), \mathbf{e}_{ij}, \tilde{\mathbf{e}}_{ij}\right), \qquad \text{(Message \& aggregate steps)} \tag{5}$$

$$\mathbf{s'}_i, \tilde{\mathbf{v}'}_i := \text{GVP}\left((\mathbf{s}_i, \tilde{\mathbf{v}}_i), (\mathbf{m}_i, \tilde{\mathbf{m}}_i)\right). \qquad \text{(Update step)} \tag{6}$$

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): $\mathbf{S} \in \mathbb{R}^{n \times k \times m}$ 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, m is the number of scalar features, and m' 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.

E. Further Discussion

E.1. Motivation

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 - it requires the protein to form a single, diffractible crystal (Dishman & Volkman, 2018). This led to a bias in the Protein Data Bank towards globular, thermodynamically stable chains. However, the dynamic nature of protein NMR and the advancements of Cryo-EM as well as the redundancy among PDB entries (i.e. multiple entries that contain almost identical protein sequences) have offered valuable insights into protein conformational landscape (Hrabe et al., 2015; Monzon et al., 2016). A series of MD datasets simulating diverse protein structures have been released (Vander Meersche et al., 2023; Mirarchi et al., 2024), albeit they have limited sequence variability and only a fraction of the simulations provide valuable information into dynamics of large molecular conformational changes.

The large number of high-resolution deposited structures in the PDB allowed for the training of large geometric deep learning models that boosted the prediction accuracy of protein structure (Jumper et al., 2021; Kryshtafovych et al., 2019). Remarkable breakthroughs in protein backbone inverse folding (Dauparas et al., 2022) and *de novo* backbone generation have followed suit (Watson et al., 2023). ProteinMPNN (Dauparas et al., 2022) in particular has streamlined single-state protein inverse folding due to much smaller computational and time costs coupled with increased sequence recovery, when compared to physics-based methods (Liu & Kuhlman, 2006) that treat sequence design as an energy optimization problem. While these models have been successfully used for *de novo* protein design for improved function (Sumida et al., 2024), they mostly rely on enhancing protein stability and solubility rather than dealing with dynamics and interaction mechanisms.