
Conformational sampling and interpolation using language-based protein folding neural networks

Diego del Alamo *

Protein Design and Informatics
GSK Research & Development
Baar, ZG, Switzerland
diego.p.delalamo@gsk.com

Jeliazko R. Jeliazkov

Protein Design and Informatics
GSK Research & Development
Collegeville, PA, USA
jeliazko.r.jeliazkov@gsk.com

Daphné Truan

Protein Design and Informatics
GSK Research & Development
Stevenage, UK
daphne.x.truan@gsk.com

Joel D. Karpiak

Protein Design and Informatics
GSK Research & Development
Collegeville, PA, USA
joel.x.karpiak@gsk.com

Abstract

Protein language models (PLMs), such ESM2, learn a rich semantic grammar of the protein sequence space. When coupled to protein folding neural networks (e.g., ESMFold), they can facilitate the prediction of tertiary and quaternary protein structures at high accuracy. However, they are limited to modeling protein structures in single states. This manuscript demonstrates that ESMFold can predict alternate conformations of some proteins, including *de novo* designed proteins. Randomly masking the sequence prior to PLM input returned alternate embeddings that ESMFold sometimes mapped to distinct physiologically relevant conformations. From there, inversion of the ESMFold trunk facilitated the generation of high-confidence interconversion paths between the two states. These paths provide a deeper glimpse of how language-based protein folding neural networks derive structural information from high-dimensional sequence representations, while exposing limitations in their general understanding of protein structure and folding.

1 Introduction

Dynamics allow proteins to carry out complex biological functions [1, 2, 3, 4], but cannot be reliably predicted from sequence alone [5]. Tuning the inputs of the alignment-based protein folding neural network AlphaFold2 [6] can sometimes accurately model proteins in multiple states [7, 8, 9, 10, 11]. However, these approaches require significant manual intervention and scale poorly to the proteome-level [12, 13, 14], precluding large-scale analyses of protein dynamics akin to those recently carried out on millions of static models [15, 16, 17]. Recently, protein folding neural networks coupled to large PLMs have achieved nearly state-of-the-art performance on structure prediction, particularly on orphan proteins and *de novo* designed proteins, at far faster compute speeds [18, 19, 20, 21]. Yet despite the widespread attention given to emergent qualities of their PLMs [22, 23, 24], their suitability in modeling conformational dynamics remains, to our knowledge, unexplored.

Here we show that the language-based protein folding neural network ESMFold can sample conformational landscapes of some *de novo* designed proteins and natural proteins, pointing to a deeper understanding of protein folding and thermodynamics than previously assumed. Randomly masking

*To whom correspondence should be addressed.

amino acids prior to PLM input allows ESM2 to generate alternate residue representations, which ESMFold converted to distinct conformational states. Inversion of the folding trunk allowed interpolation of these representations along high-pLDDT transition paths, thereby generating hypotheses of conformational interconversion (Figure 1). Sequence representations that mapped to distinct conformations of natural proteins, but not *de novo* designed proteins, were highly segregated, with abrupt transitions manifested in high-RMSD structural changes that sometimes skip over high-energy transition states. These results suggest that language-based protein folding neural networks are equipped to rapidly generate hypotheses about protein dynamics. At the same time, they expose discontinuities in how ESMFold maps sequences to structures.

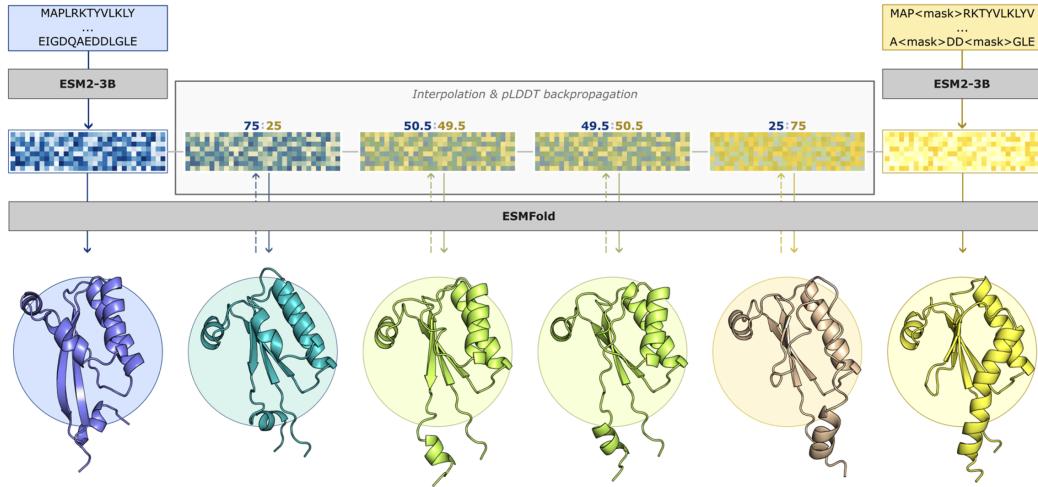


Figure 1: Overview of the conformational sampling and interpolation pipeline. Residue-level representations, shown as heat maps, for initial start and end states are generated using either an unmasked query sequence (left) or a partially masked sequence (right), thereby generating multiple states. Element-wise interpolation of these representations is combined with backpropagation through the ESMFold folding neural network to generate high-confidence transition paths.

2 Results and Discussion

2.1 *De novo* designed proteins can be modeled in apo and holo conformations

Self-supervised training of large language models, including PLMs like ESM2, relies on randomized masking of tokens in a sequence, with the objective of recovering the unmasked query sequence [25, 26, 27]. Recently, Hermosilla *et al* demonstrated that ESMFold can correctly fold *de novo* designed proteins when large fractions (80-90%) of the query sequence are masked [28]. We reasoned that this approach may be able to sample both states of a recent set of six proteins that were *de novo* designed to interconvert between apo and holo conformations [29].

Both conformations were predicted at high accuracy for two of the six proteins, with the remaining four sampling conformations with features from both states (i.e., putatively intermediate; Figures 2 and S1). For the successful cases, conformational sampling occurred even when only 10% of the sequence was masked. The ESM2 representations of the masked sequences were largely identical to those of the unmasked sequence, with high cosine similarities and low Euclidean distances (Figure S2). Dimensionality reduction of these representations using t-SNE [30] further showed how distinct conformational states were not derived from distinct representations. These results suggested that ESMFold may have learned to extrapolate conformational dynamics in some proteins absent from the training set.

2.2 Distinct conformations of natural proteins are segregated in the embedding space

Expecting these results to be recapitulated in natural monomeric proteins with well-defined dynamics, we assembled a panel of 11 fold-switching proteins, 6 transporters, and 90 ligand-binding proteins

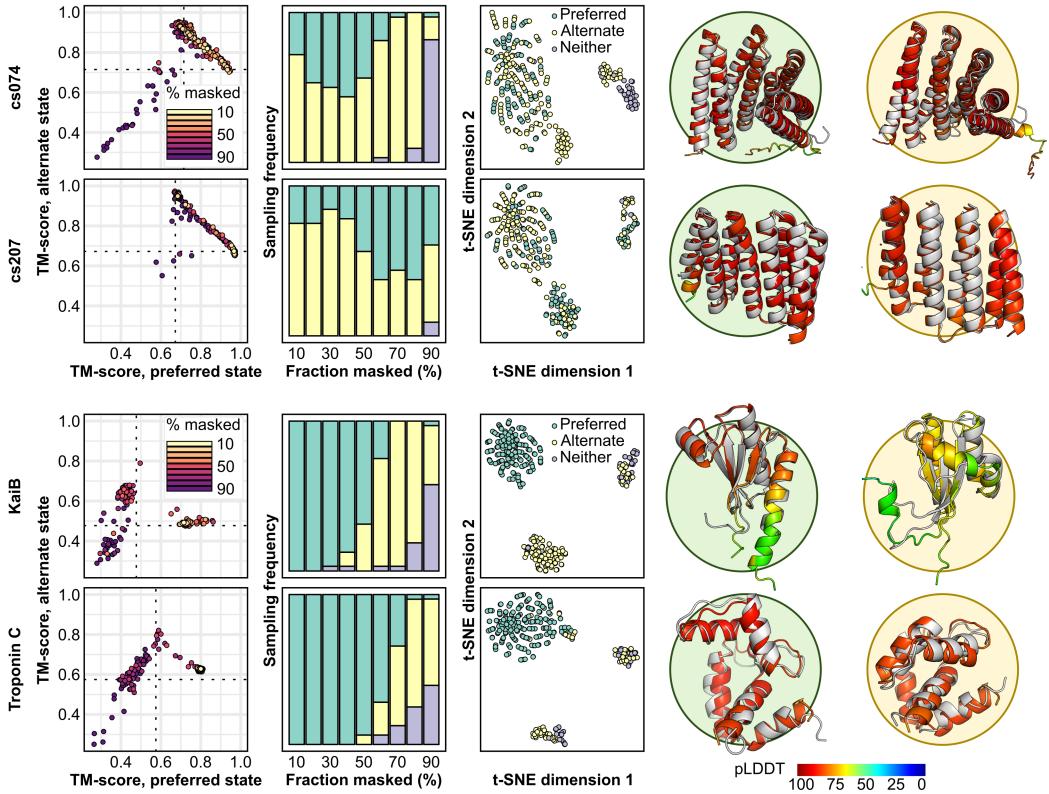


Figure 2: ESMFold predicts alternative conformations of *de novo* designed and some natural proteins. For *de novo* designed proteins (top two rows), no correlation between sequence masking and conformational sampling is observed. In contrast, natural proteins (bottom two rows) only sample alternative states when most of the sequence is masked. Dashed lines indicate TM-scores between reference structures. Cartoon diagrams depict the preferred (left) and alternative (right) states, with reference structures shown in gray. Computational models of the protein cs074 were used as references due to lack of experimental structures.

from a variety of previous benchmarks [7, 31, 32] (Table S1). Although this approach successfully sampled multiple conformations of some proteins (examples shown in Figure 2), in most cases it failed to do so. Such proteins were predicted either exclusively in one state or, more frequently, in a putatively intermediate state that blended structural features from both states (Figures S3, S4, S5, S6, S7, S8, and S9).

How ESMFold mapped the sequence representations of natural proteins to structure diverged from its treatment of *de novo* designed proteins in several key respects. First, conformational sampling of natural proteins steadily increased with masking rate until a critical threshold appeared to be reached, at which point fewer native-like structures were being generated at all. Second, the ESM2 embeddings for the two states segregated into distinct clusters, with folding failures dispersed elsewhere. Finally, major changes in the per-residue representations were observed in embeddings that yielded the alternative state relative to the unmasked query sequence (Figure S2). The magnitude of these residue-level changes did not appear to correlate with either LDDT [33] (Figure S10) and only weakly correlated with residues' movement in Cartesian space (Figure S11), consistent with encoding of structural changes at the whole-sequence rather than residue level.

2.3 Tracking the determinants of conformational sampling in ESMFold

Why did sequence masking induce conformational transitions in some proteins, but not others? Structural models matching the alternate conformation could be found in the ESMFold Metagenomic Atlas [18] using Foldseek [34] for almost all proteins discussed here, indicating that this effect does

not arise from the alternate state being unreachable by the folding trunk (Figure S12). It also failed to correlate with the number of hits to the query sequence fetched from the UniRef50 and UniRef90 databases using Jackhmmer [35, 36] (Figure S13). Moreover, conformational sampling failed to correlate with the query sequence’s pseudo-perplexity, a metric capturing how well it is understood by the PLM [18] (Figure S14). These negative results, combined with those observed in four out of six *de novo* designed proteins, suggest that variations in conformational sampling may arise from the cumulative effects of idiosyncrasies and imbalances in the training set, rather than specific properties of the individual proteins being predicted.

2.4 Sampling conformational transition paths by interpolating between representations for different states

Because the representations for both end states are identical in shape, we reasoned that transition paths between these states could be generated by iteratively interpolating between their representations and passing the resulting outputs through the ESMFold folding trunk as shown in Figure 1. The full algorithm is detailed in Section 4; briefly, it proceeded recursively, with initial guesses generated by averaging the representations of high-RMSD pairs of consecutive structural models in the transition path and refining by backpropagation through an inverted ESMFold network, an approach inspired by similar methods for protein design [37, 38, 39, 40, 41]. Loss functions included pLDDT as well as Euclidean distance restraints ensuring that the ESM2 representations for new models in the transition path were equidistant from those of the two flanking models used to generate them.

Nine such trajectories were generated in the fold-switching protein KaiB and the calcium-binding protein Troponin C (Figure 3A). All eighteen trajectories limited conformational interconversion to a very minor fraction of the overall Euclidean space separating the two sequence representations, and no correlation was observed between Euclidean distance traversed in the embedding space and RMSD changes in resulting structures (Figure 3B). Most trajectories were found to skip over high-energy conformational transitions entirely in both proteins, even though the traversed distance in Euclidean terms was extremely small (Figure 3C).

Residues 62–68 of KaiB provide an illustrative example, as they undergo a sheet-to-helix transition that must, in principle, sample an intermediate loop conformation. Only one trajectory, shown in Figure 3D successfully sampled such a state. The remaining eight skipped over the state entirely, suggesting that high-energy states occupy an extremely limited fraction of the total embedding space separating the two conformations, and that the boundaries between the representations of the two states do not uniformly reflect biophysically relevant conformational transitions.

3 Conclusion

This manuscript demonstrates that language-based protein folding neural networks are primed to explore protein conformational landscapes at scale. Of particular interest was the observation that ESMFold interpreted nearly-identical language model representations of *de novo* designed proteins as distinct conformations, which may point to an emergent understanding of structural dynamics as a fundamental property of some amino acid sequences. However, although we observed success with *de novo* designed proteins that is unachievable with AlphaFold without pipeline modifications [29], the relatively low success rate in our benchmark set proves that reliable conformational sampling is not currently achievable by ESMFold, and may require alternate training schemes to fully unlock [42]. Moreover, our proof-of-concept involving the simulation of transition paths in KaiB and Troponin C demonstrated that the language model’s embedding space does not neatly map to conformational space, as indicated by unphysical transition paths between discrete conformations. Nevertheless, it provides one example of the broader range of hypotheses that could be generated by a more robust language-based protein folding neural network. Finally, while these results are limited to ESM2 and ESMFold, its two-stage training procedure closely aligns with that of other language-based protein folding models such as OmegaFold and RGN2 [20, 43], and similar results would be expected in those models as well.

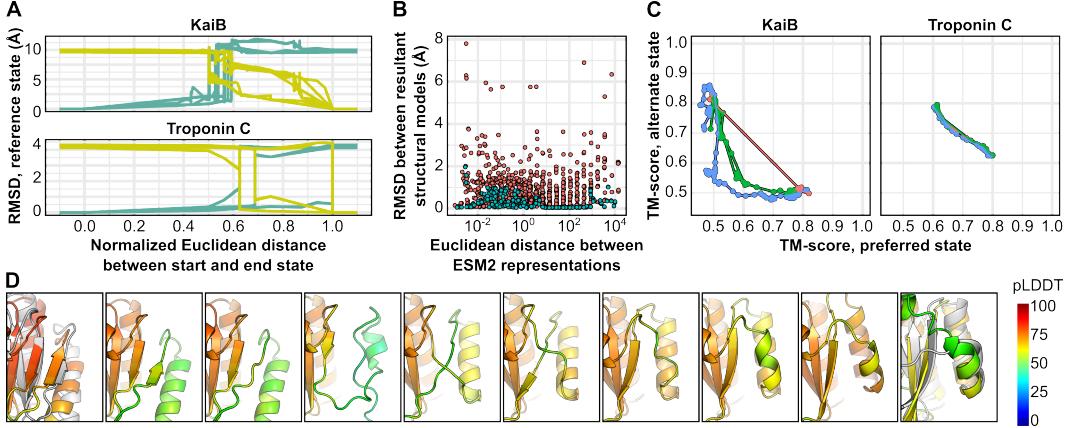


Figure 3: Conformational interpolation between the ESM2 representations of start and end states of KaiB and Troponin C using ESMFold. (A) Conformational transitions occur during very short intervals during the transition from one state to another. Start and end points were chosen based on structural similarity to either the preferred state or alternate state and are shown in either teal or yellow. (B) Changes in protein representations are not correlated with changes in RMSD in either KaiB (orange) or Troponin C (teal). (C) Example trajectories show how different sampling granularities are achieved, with some trajectories skipping over conformational transitions. (D) Example sheet-to-helix transition in KaiB residues 62–68, with residues colored by pLDDT. Crystal structures are shown in white on the leftmost and rightmost panels.

4 Methods

4.1 Sequence masking and conformational sampling

Conformational sampling was achieved by modifying a recently described sequence masking approach described in [28]. The identities of a fraction of residues 0.1, 0.2, … 0.9 are randomly masked prior to processing into embeddings by the ESM2 language model; our modification is limited to returning the full $(N_{layer}, N_{res}, N_{dim})$ embeddings for each sequence, rather than the embeddings for the last layer. We applied this to the benchmark set of protein structure pairs in Table S1, which was derived from three benchmarks for conformational change modeling using AlphaFold2 and includes recent *de novo* designed proteins that adopt multiple conformations. Recycles were set to zero.

4.2 Interpolation between models in embedding space

Transition paths were generated using a recursive algorithm that interpolates between two sets of language model embeddings x_1, x_2 , each of which has the shape $(N_{layer}, N_{res}, N_{dim})$ and maps to structurally distinct conformers. For the ESM2 model, these representations are of shape $37 * N_{res} * 2560$, or approximately 79,000 dimensions per amino acid. At each iteration i , we seek to introduce a new representative model x_i that interpolates between the highest-RMSD pair of adjacent structures. After calculating pairwise RMSD values for all such pairs along the trajectory, a new representative model is generated by taking the midpoint between the representations for the highest-RMSD pair of consecutive structural models. The resulting calculated representations are passed through the folding trunk of ESMFold to generate a structural model, which is then optimized using two losses: a pLDDT loss equal to negative pLDDT, and a geometric loss L_{dist} that ensures the new representations are equidistant from the two seed representations:

$$L_{dist} = \left(\frac{|x_{i,n} - x_{i-1}| - |x_{i,n} - x_{i+1}|}{|x_{i,0} - x_{i-1}|} \right)^2 \quad (1)$$

Here, x_{i-1} and x_{i+1} are the flanking representations used for interpolation, and $x_{i,0}$ and $x_{i,n}$ are the initial and refined embeddings for the model whose position is being interpolated. The Adam optimizer with beta1 = 0.9, beta2 = 0.999 was used with gradient clipping set to 1, and the learning

rate was initially set to 1e-3 but was scaled with the distance between the two seed embeddings, x_{i-1} and x_{i+1} [44]. Each iteration proceeded until convergence. The algorithm terminated when repeated interpolation failed to decrease the distance between the highest-RMSD pairs of structures. All computations were carried out on an A6000 GPU.

4.3 Miscellaneous

Searches using Foldseek (v. 2-8bd520) [34] and Jackhmmer (v3.2.1) [35] were carried out using default settings. t-SNE was used as implemented in python using SciKit-Learn [45]. TM-scores and structural alignments were calculated using TM-align [46].

Acknowledgments and Disclosure of Funding

The authors would like to thank the GSK Fellows Program for research support and the ESM team at Meta for permissive licensing and development of ESMFold.

References

- [1] Eleanor Campbell, Miriam Kaltenbach, Galen J. Correy, Paul D. Carr, Benjamin T. Porebski, Emma K. Livingstone, Livnat Afriat-Jurnou, Ashley M. Buckle, Martin Weik, Florian Hollfelder, Nobuhiko Tokuriki, and Colin J. Jackson. The role of protein dynamics in the evolution of new enzyme function. *Nature Chemical Biology*, 12(11):944–950, 2016. doi: 10.1038/nchembio.2175.
- [2] Kresten Lindorff-Larsen, Robert B. Best, Mark A. DePristo, Christopher M. Dobson, and Michele Vendruscolo. Simultaneous determination of protein structure and dynamics. *Nature*, 433(7022):128–132, 2005. doi: 10.1038/nature03199.
- [3] Ferran Feixas, Steffen Lindert, William Sinko, and J. Andrew McCammon. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry*, 186:31–45, February 2014. doi: 10.1016/j.bpc.2013.10.007.
- [4] Ashutosh Srivastava, Sandhya Premnath Tiwari, Osamu Miyashita, and Florence Tama. Integrative/Hybrid Modeling Approaches for Studying Biomolecules. *Journal of Molecular Biology*, 432(9):2846–2860, 2020. doi: 10.1016/j.jmb.2020.01.039.
- [5] Thomas J. Lane. Protein structure prediction has reached the single-structure frontier. *Nature Methods*, 20(2):170–173, February 2023. doi: 10.1038/s41592-022-01760-4.
- [6] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishabh Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michał Zieliński, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. doi: 10.1038/s41586-021-03819-2.
- [7] Diego del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife*, January 2022. doi: 10.1101/2021.11.22.469536.
- [8] Richard A. Stein and Hassane S. Mchaourab. SPEACH_af: Sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLOS Computational Biology*, 18(8), August 2022. doi: 10.1371/journal.pcbi.1010483.
- [9] Hannah K. Wayment-Steele, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2, October 2022.

- [10] Lim Heo and Michael Feig. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics*, 90(11):1873–1885, November 2022. doi: 10.1002/prot.26382.
- [11] D. Sala, F. Engelberger, H.S. Mchaourab, and J. Meiler. Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*, 81:102645, August 2023. doi: 10.1016/j.sbi.2023.102645.
- [12] Bulat Faezov and Roland L. Dunbrack. AlphaFold2 models of the active form of all 437 catalytically-competent typical human kinase domains, July 2023.
- [13] Diego Del Alamo, Lillian DeSousa, Rahul M. Nair, Suhaila Rahman, Jens Meiler, and Hassane S. Mchaourab. Integrated AlphaFold2 and DEER investigation of the conformational dynamics of a pH-dependent APC antiporter. *Proceedings of the National Academy of Sciences*, 119(34):e2206129119, August 2022. doi: 10.1073/pnas.2206129119.
- [14] Bodhi P. Vani, Akashnathan Aranganathan, Dedi Wang, and Pratyush Tiwary. AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *Journal of Chemical Theory and Computation*, 19(14):4351–4354, July 2023. doi: 10.1021/acs.jctc.3c00290.
- [15] Janani Durairaj, Andrew M. Waterhouse, Toomas Mets, Tetiana Brodiazhenko, Minhal Abdullah, Gabriel Studer, Gerardo Tauriello, Mehmet Akdel, Antonina Andreeva, Alex Bateman, Tanel Tenson, Vasili Hauryliuk, Torsten Schwede, and Joana Pereira. Uncovering new families and folds in the natural protein universe. *Nature*, September 2023. doi: 10.1038/s41586-023-06622-3.
- [16] Inigo Barrio-Hernandez, Jingi Yeo, Jürgen Jänes, Milot Mirdita, Cameron L. M. Gilchrist, Tanita Wein, Mihaly Varadi, Sameer Velankar, Pedro Beltrao, and Martin Steinegger. Clustering-predicted structures at the scale of the known protein universe. *Nature*, September 2023. doi: 10.1038/s41586-023-06510-w.
- [17] Kathryn Tunyasuvunakool, Jonas Adler, Zachary Wu, Tim Green, Michal Zielinski, Augustin Žídek, Alex Bridgland, Andrew Cowie, Clemens Meyer, Agata Laydon, Sameer Velankar, Gerard J Kleywegt, Alex Bateman, Richard Evans, Alexander Pritzel, Michael Figurnov, Olaf Ronneberger, Russ Bates, Simon A A Kohl, Anna Potapenko, Andrew J Ballard, Bernardino Romera-Paredes, Stanislav Nikolov, Rishabh Jain, Ellen Clancy, David Reiman, Stig Petersen, Andrew W Senior, Koray Kavukcuoglu, Ewan Birney, Pushmeet Kohli, John Jumper, and Demis Hassabis. Highly accurate protein structure prediction for the human proteome. *Nature*, July 2021. doi: 10.1038/s41586-021-03828-1.
- [18] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.adc2574.
- [19] Robert Verkuil, Ori Kabeli, Yilun Du, Basile I. M. Wicky, Lukas F. Milles, Justas Dauparas, David Baker, Sergey Ovchinnikov, Tom Sercu, and Alexander Rives. Language models generalize beyond natural proteins, December 2022.
- [20] Ratul Chowdhury, Nazim Bouatta, Surojit Biswas, Christina Floristean, Anant Kharkar, Koushik Roy, Charlotte Rochereau, Gustaf Ahdritz, Joanna Zhang, George M. Church, Peter K. Sorger, and Mohammed AlQuraishi. Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotechnology*, 40(11):1617–1623, November 2022. doi: 10.1038/s41587-022-01432-w.
- [21] Xiaomin Fang, Fan Wang, Lihang Liu, Jingzhou He, Dayong Lin, Yingfei Xiang, Xiaonan Zhang, Hua Wu, Hui Li, and Le Song. HelixFold-Single: MSA-free Protein Structure Prediction by Using Protein Language Model as an Alternative, February 2023.
- [22] Roshan Rao, Joshua Meier, Tom Sercu, Sergey Ovchinnikov, and Alexander Rives. Transformer protein language models are unsupervised structure learners, December 2020.

- [23] Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alexander Rives. Language models enable zero-shot prediction of the effects of mutations on protein function, July 2021.
- [24] Yves Gaetan Nana Teukam, Loic Kwate Dassi, Matteo Manica, Daniel Probst, Philippe Schwaller, and Teodoro Laino. Language models can identify enzymatic active sites in protein sequences, November 2021.
- [25] Alexander Rives, Joshua Meier, Tom Sercu, Siddharth Goyal, Zeming Lin, Jason Liu, Demi Guo, Myle Ott, C Lawrence Zitnick, Jerry Ma, and Rob Fergus. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):1–12, April 2021. doi: 10.1073/pnas.2016239118.
- [26] Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, Debsindhu Bhowmik, and Burkhard Rost. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, October 2022. doi: 10.1109/TPAMI.2021.3095381.
- [27] Nadav Brandes, Dan Ofer, Yam Peleg, Nadav Rappoport, and Michal Linial. ProteinBERT: a universal deep-learning model of protein sequence and function. *Bioinformatics*, 38(8):2102–2110, April 2022. doi: 10.1093/bioinformatics/btac020.
- [28] Alvaro Martin Hermosilla, Carolin Berner, Sergey Ovchinnikov, and Anastassia Andreevna Vorobieva. Validation of de novo designed water-soluble and transmembrane proteins by in silico folding and melting, June 2023.
- [29] 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, Rianne 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, January 2023. doi: 10.1126/science.adg7731.
- [30] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.
- [31] 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, María Silvina Fornasari, and Gustavo Parisi. Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics*, 38(10):2742–2748, May 2022. doi: 10.1093/bioinformatics/btac202.
- [32] Devlina Chakravarty and Lauren L. Porter. AlphaFold2 fails to predict protein fold switching. *Protein Science*, 31(6), June 2022. doi: 10.1002/pro.4353.
- [33] Valerio Mariani, Marco Biasini, Alessandro Barbato, and Torsten Schwede. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, 29(21):2722–2728, November 2013. doi: 10.1093/bioinformatics/btt473.
- [34] Michel Van Kempen, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L.M. Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, February 2022. doi: 10.1101/2022.02.07.479398.
- [35] L Steven Johnson, Sean R Eddy, and Elon Portugaly. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, 11(1):431, December 2010. doi: 10.1186/1471-2105-11-431.
- [36] Baris E. Suzek, Hongzhan Huang, Peter McGarvey, Raja Mazumder, and Cathy H. Wu. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23(10):1282–1288, May 2007. doi: 10.1093/bioinformatics/btm098.

- [37] Jeliazko R. Jeliazkov, Diego Del Alamo, and Joel D. Karpiak. ESMFold Hallucinates Native-Like Protein Sequences, May 2023.
- [38] Sumit Kumar Jha, Arvind Ramanathan, Rickard Ewetz, Alvaro Velasquez, and Susmit Jha. Protein Folding Neural Networks Are Not Robust, September 2021.
- [39] Ismail Alkhouri, Sumit Jha, Andre Beckus, George Atia, Alvaro Velasquez, Rickard Ewetz, Arvind Ramanathan, and Susmit Jha. On the Robustness of AlphaFold: A COVID-19 Case Study, January 2023.
- [40] Zhongju Yuan, Tao Shen, Sheng Xu, Leiyu Yu, Ruobing Ren, and Siqi Sun. AF2-Mutation: Adversarial Sequence Mutations against AlphaFold2 on Protein Tertiary Structure Prediction, May 2023.
- [41] Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Joseph L Watson, Karla M Castro, Robert Ragotte, Amijai Saragovi, Lukas F Milles, Minkyung Baek, Ivan Anishchenko, Wei Yang, Derrick R Hicks, Marc Expòsit, Thomas Schlichthaerle, Jung-Ho Chun, Justas Dauparas, Nathaniel Bennett, Basile I M Wicky, Andrew Muenks, Frank DiMaio, Bruno Correia, Sergey Ovchinnikov, and David Baker. Scaffolding protein functional sites using deep learning. *Science*, 377:387–394, 2022. doi: 10.1126/science.abn2100.
- [42] Patrick Bryant. Structure prediction of alternative protein conformations, September 2023. URL <http://biorxiv.org/lookup/doi/10.1101/2023.09.25.559256>.
- [43] Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, Jianzhu Ma, and Jian Peng. High-resolution *de novo* structure prediction from primary sequence, July 2022.
- [44] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, January 2017.
- [45] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [46] Yang Zhang and Jeffrey Skolnick. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7):2302–2309, 2005. ISSN 03051048. doi: 10.1093/nar/gki524.

5 Supplementary Tables

Table S1: **List of PDB pairs used in this study.** One of the reference structures for transporter LmrP was derived from CASP14 (T1024TS4271A).

2qkeA 5jytA	2hdmA 2n54B	2hdmA 2n54B	1uxmK 2namA
1x0gA 1x0gD	1xjtA 1xjuB	2frhA 1fzpD	2k0qA 2lelA
2nxqb 1jfka	2ougC 6c6sD	3hdeA 3hdfA	3qy2A 1qb3A
3zwgN 4tsyD	4gqcC 4gqcB	4qhfA 4qhhA	4zrbC 4zrbH
5fluE 2uy7D	6z4uA 7kdtB	1aelA 1ureA	1rzlA 1uvbA
1c54A 1rghB	1ealA 1eioA	1f3yA 1jkna	1fmfA 1id8A
1gh1A 1cz2A	1gqnA 1qfeB	1gqza 2gkeA	1hsIB 1hshD
1i56A 1el1A	1igpA 2au6A	1jfjA 1jfka	1k2hA 1zfsB
1lipA 1jtbA	1mutA 1punA	1mx7A 1mx8A	1ntrA 1krxA
1o1uA 1o1vA	1pdbA 1yhoA	1tfuA 3uc5A	1tjdA 1eejB
1xsaA 1xscA	1z15A 1z17A	2cjoA 1roeA	2d9eA 2rs9B
2f63A 1eqmA	2fhmA 2hltA	2in2A 2b0fA	2ju3A 2ju8A
2jwwA 1rtpl	2k43A 2k8rA	2l50A 2l51B	2l68A 2lkka
2laoA 1lahE	2lhsA 2bemA	2nlnA 1rroA	2p3mA 2vbtA
2uz5A 2vcda	4akeB 2eckB	1ijaa 2kidA	1jm4B 1wumA
1mo7A 1mo8A	1skta 1tnqa	1symB 1xydB	1w4uA 1ur6A
2kxlA 2k0gA	2lkcA 2lkdA	1lmza 1p7mA	2cg7A 2rkzC
1urpD 2driA	1ormA 1qj8A	1pfIA 1filA	1fsfA 1fqoB
1akza 1sspE	2ai6A 2ozwa	1vr6A 1rzma	1y3qa 1y3na
1gudA 1rpja	1ex6B 1ex7A	1wd7B 1wcwA	1w0jE 1e1rF
1rf5A 1rf4A	1s2oA 1tj5A	1k5hA 1q0qa	1zola 1o03A
1viyC 1vhla	1za1A 1q95A	1jejA 1jg6A	1hooB 1cg0A
1e5lA 1e5qa	1hw1B 1h9gA	1l0wB 1g51A	1otjd 1gy9A
1evkA 1evlA	1g6wD 1k0bC	1njgB 1njfa	1rkaA 1gqtA
1k6wA 1k70A	1yl5B 1yl7A	6oy9B 1a0rB	1a6dA 1a6ea
2rcsH 1aj7H	1aw2A 1aw1B	1rkmA 1b6hA	7azpA 4pj1A
4lp5A 4p2yA	1cfca 5dowA	2kq2A 2kw4A	1dmoA 3clnA
6t1zA T1024TS4271A	6irsB 7dsqb	6rvxa 7bcqa	6xpfa 6xpdA

6 Supplementary Figures

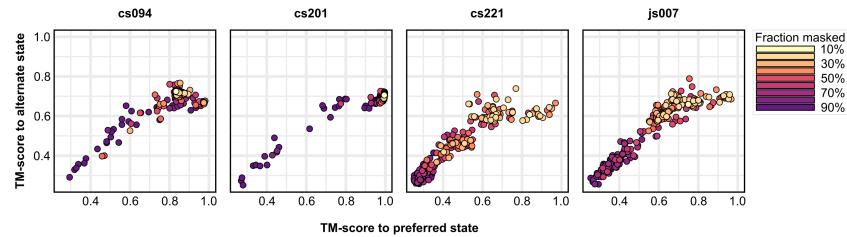


Figure S1: ESMFold predicts intermediate states of some *de novo* designed proteins. Computational models were used for all proteins.

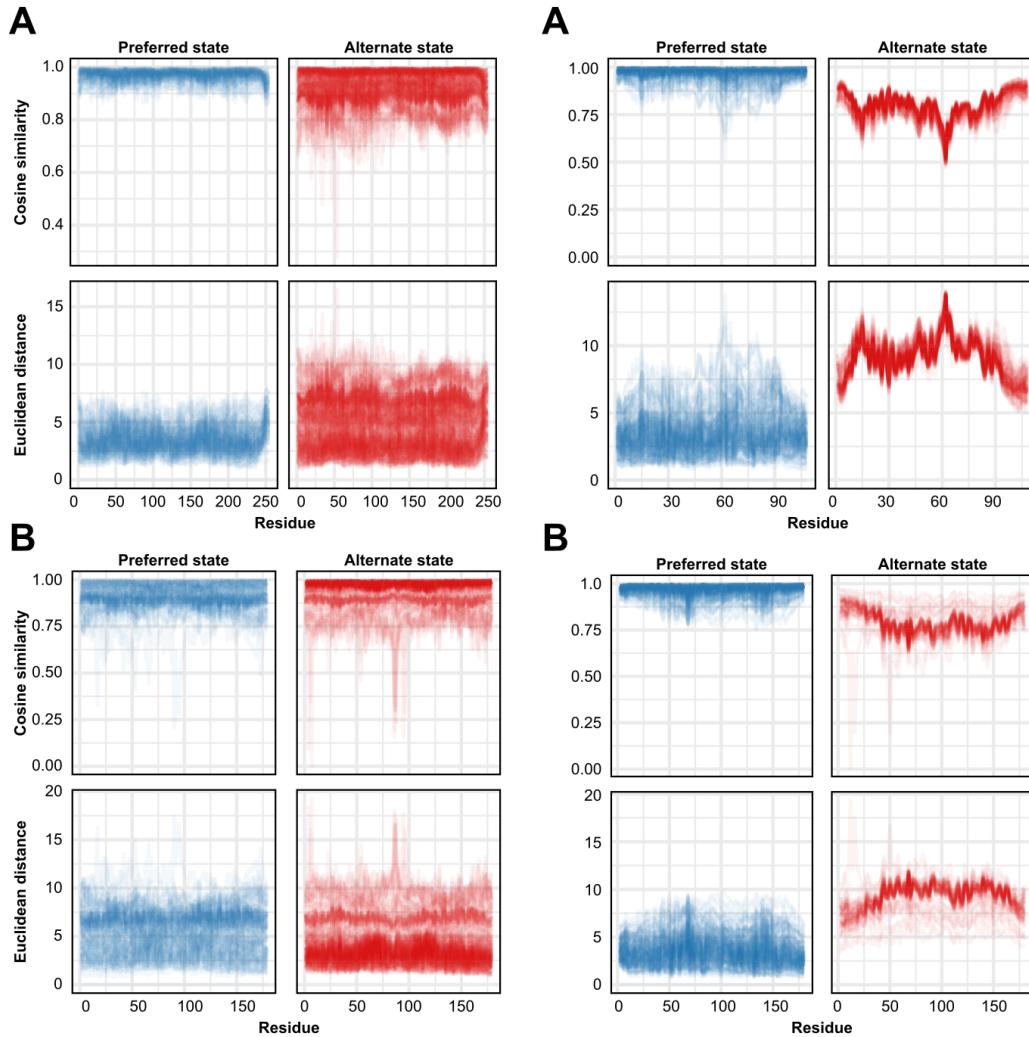


Figure S2: Distinctions in residue-level language model embeddings of designed and natural proteins. The plots show the difference between the unmasked query sequence and those mapping to either the preferred or alternate states of *de novo* designed proteins cs074 (top left) and cs207 (bottom left), and natural proteins KaiB (top right) and Troponin C (bottom right). The representations for alternate conformations of natural proteins on the right show greater dissimilarity from those of the unmasked query sequence than the *de novo* designed proteins.

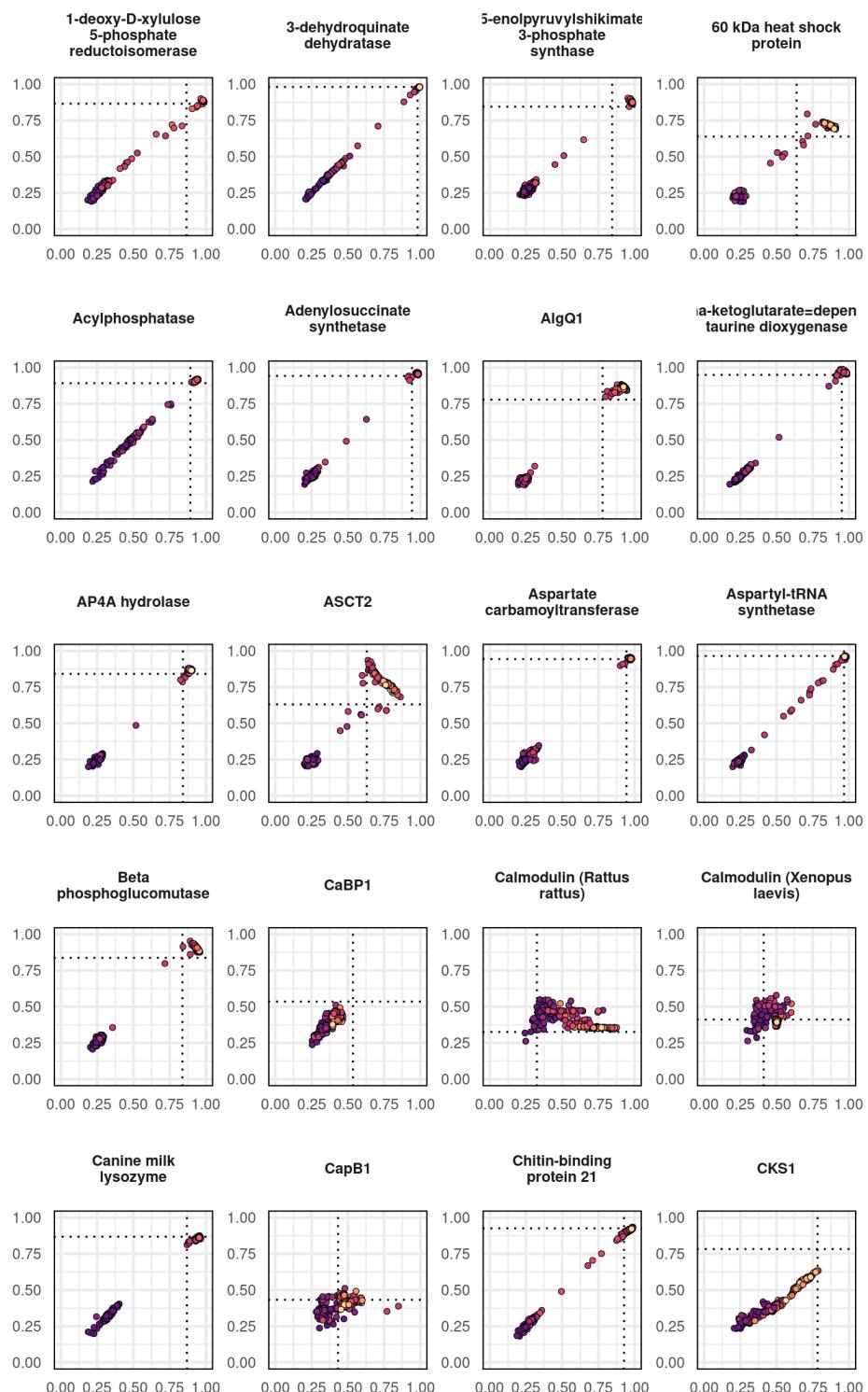


Figure S3: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

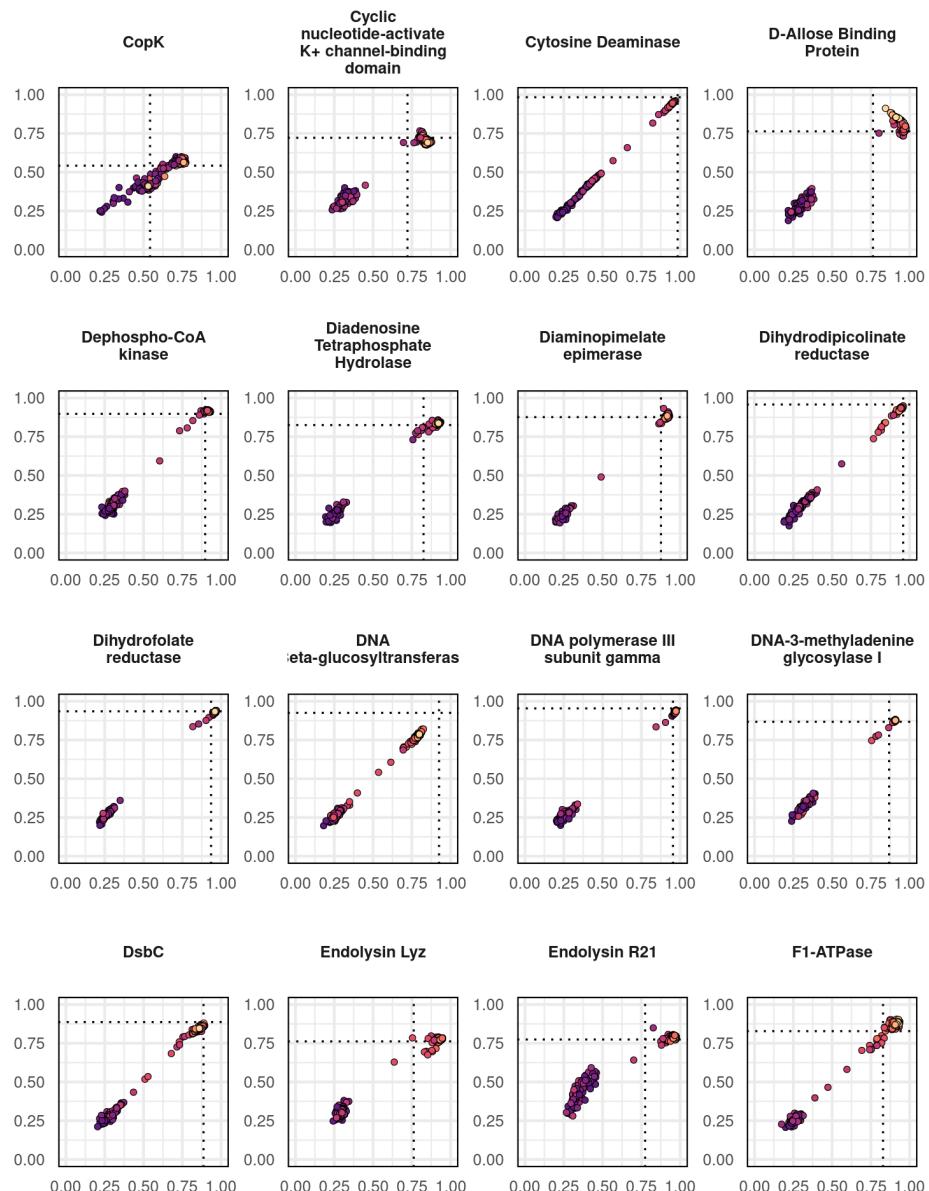


Figure S4: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

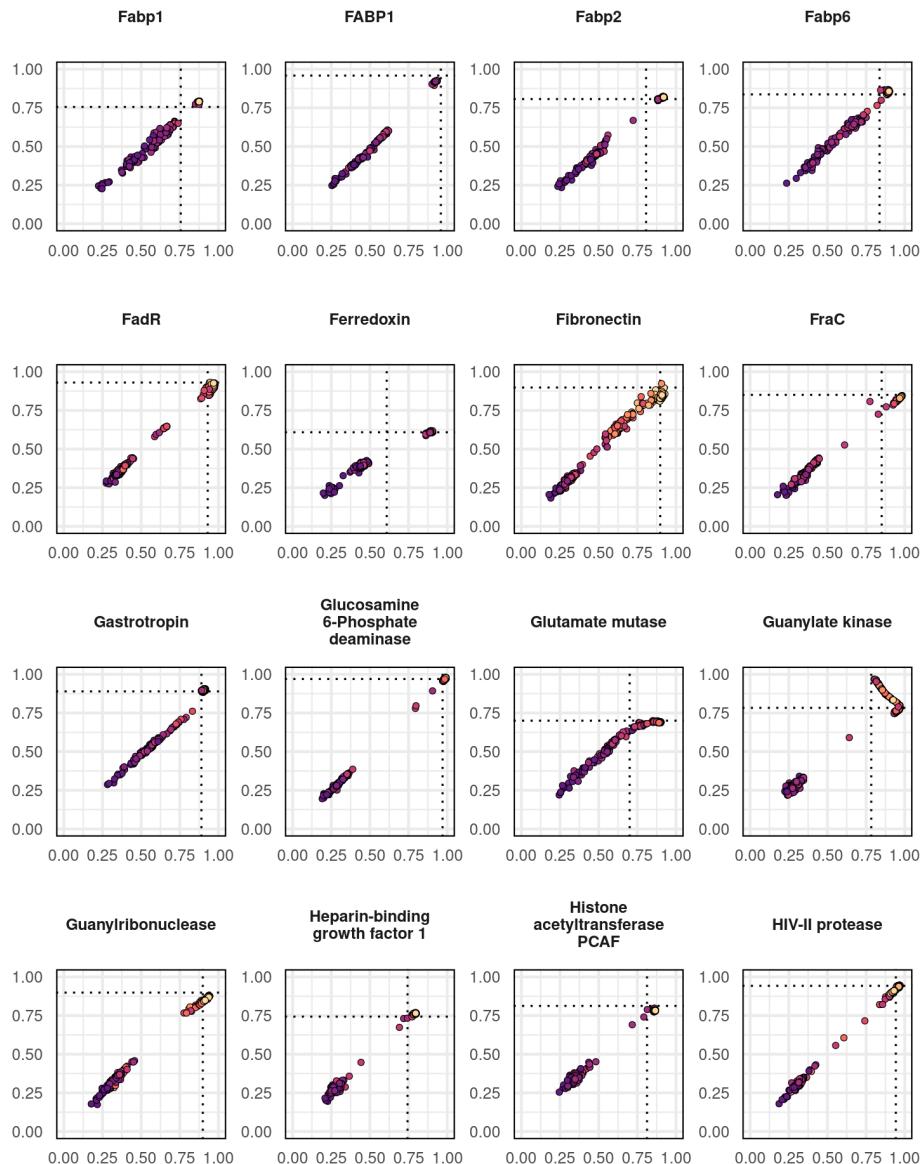


Figure S5: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

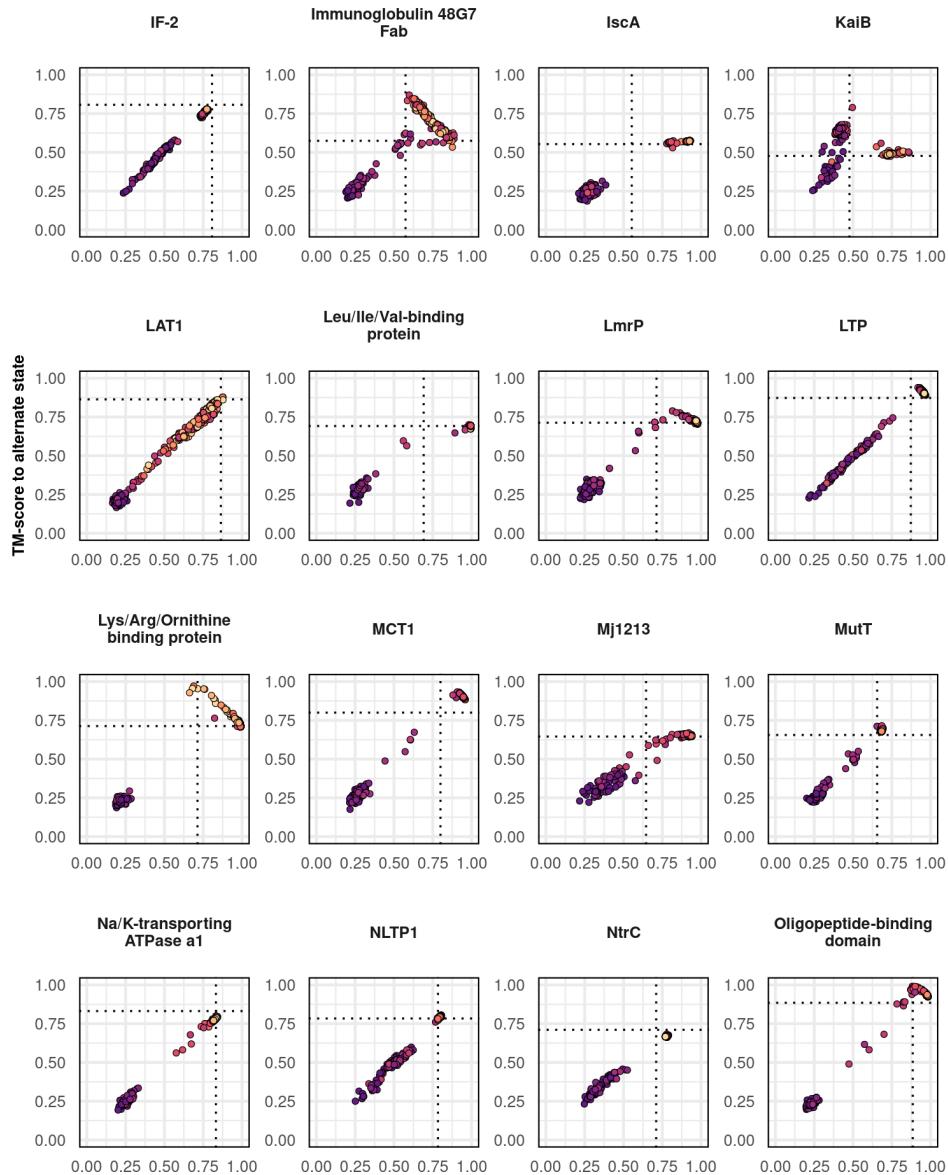


Figure S6: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

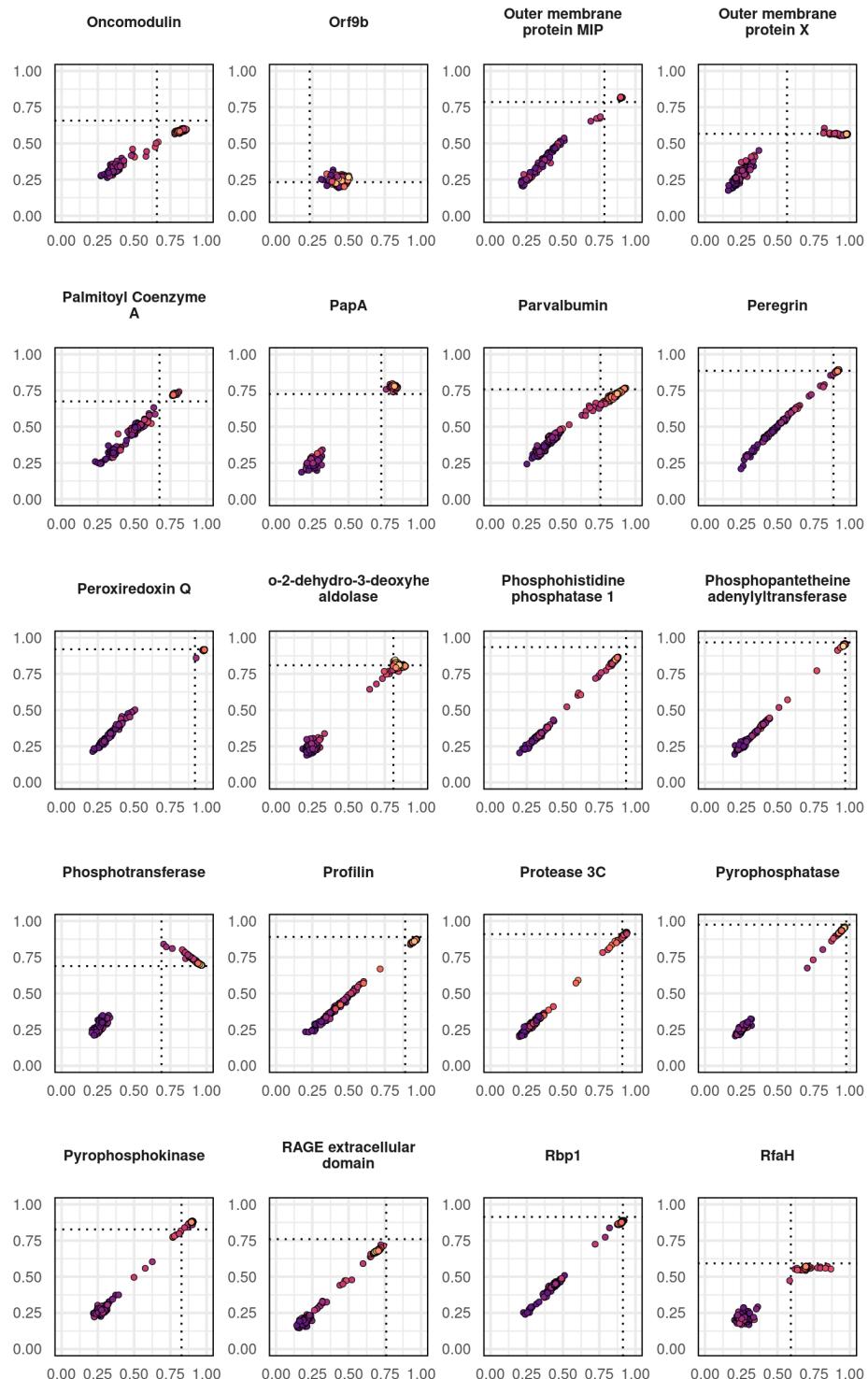


Figure S7: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

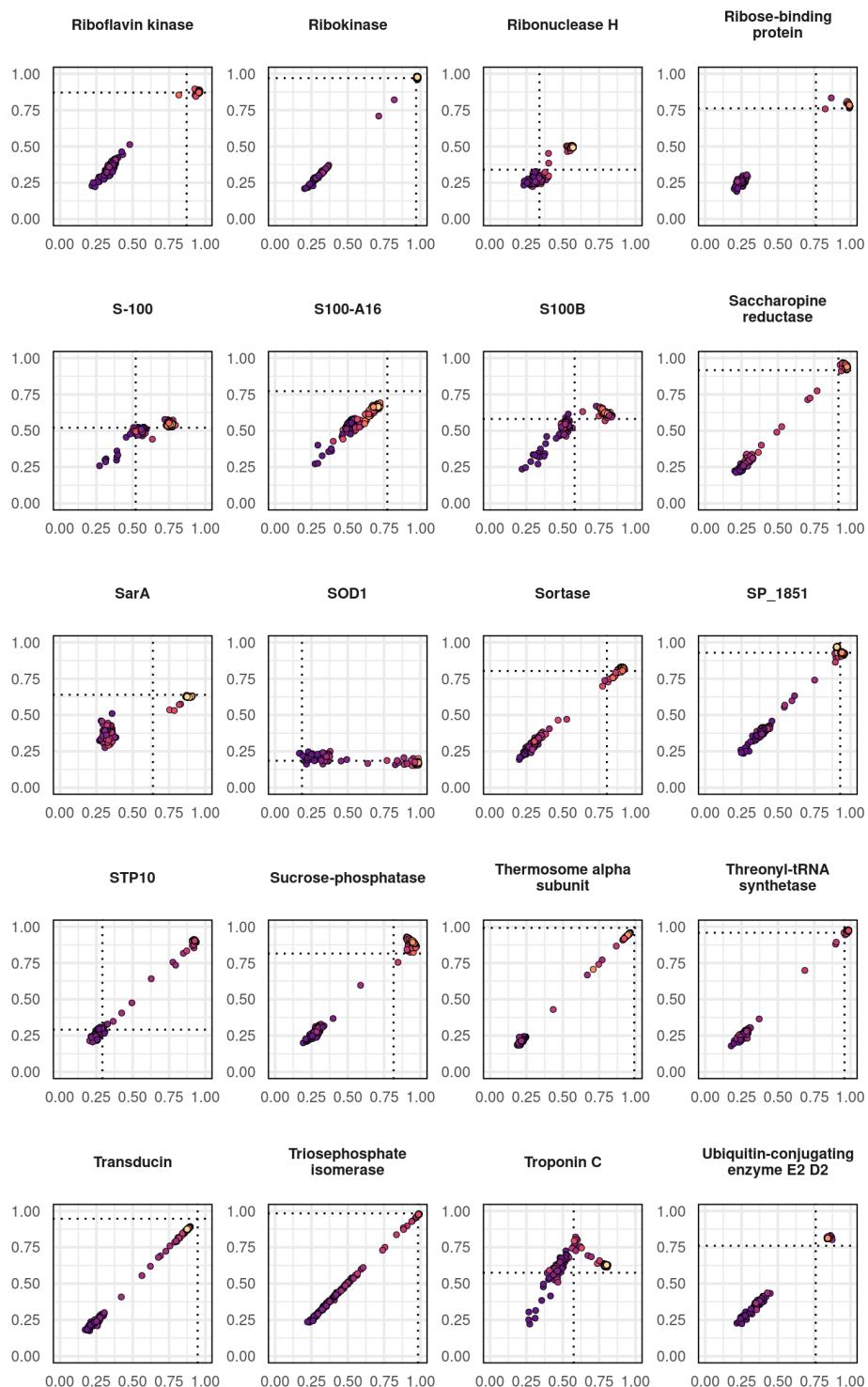


Figure S8: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

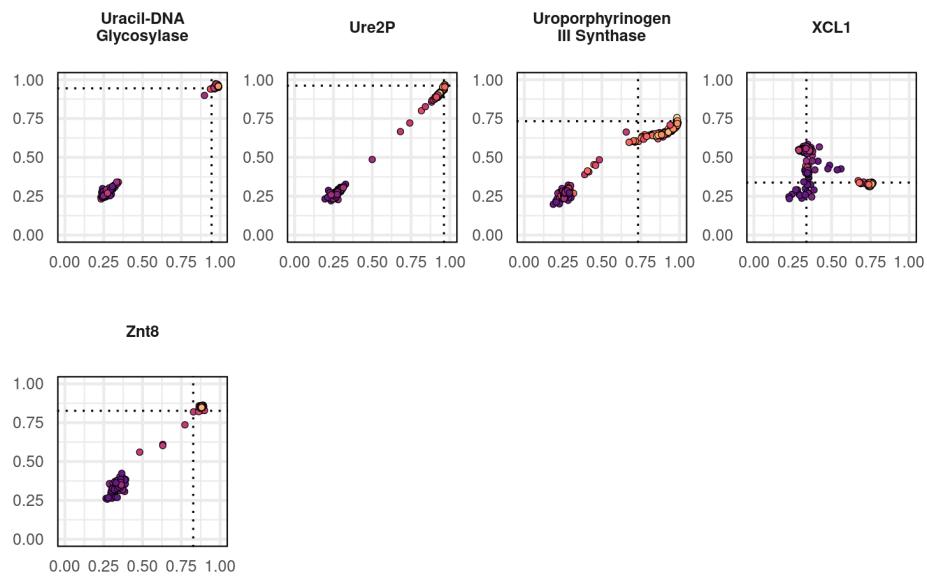


Figure S9: Conformational sampling in benchmark proteins. Preferred and alternate conformations shown on the X- and Y-axes, respectively. Dashed lines indicate TM-scores between reference structures, with a TM-score of 1 indicating that the structures are identical. Colors correspond to the fraction of the amino acid sequence that was masked.

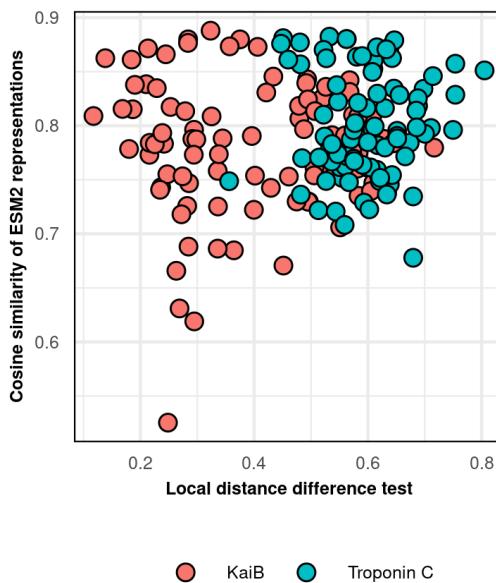


Figure S10: Residue-level structural changes (here, local distance difference test, or LDDT) do not correlate with residue-level differences in the ESM2 representations encoding those conformations.

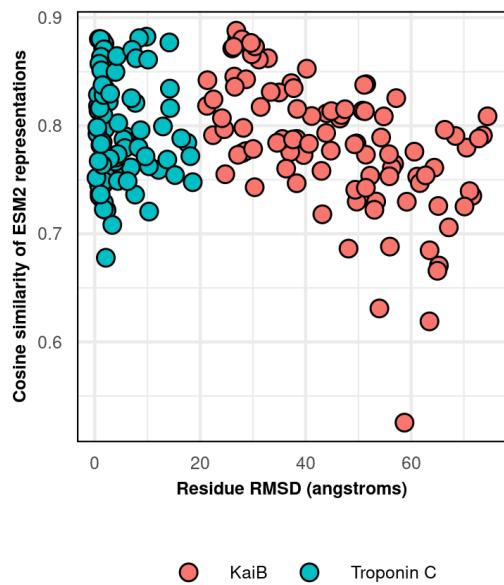


Figure S11: Residue-level structural changes (here, root mean squared deviation, or RMSD) weakly correlate with residue-level differences in the ESM2 representations encoding those conformations in KaiB and not in Troponin C.

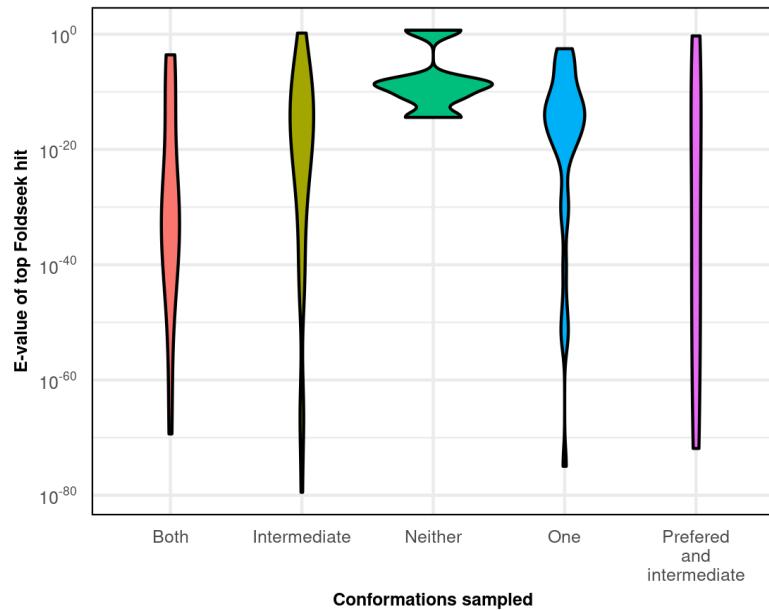


Figure S12: No correlation was observed between conformational sampling success in natural proteins and the presence or absence of similar protein structural models in the ESM Metagenomic Atlas.

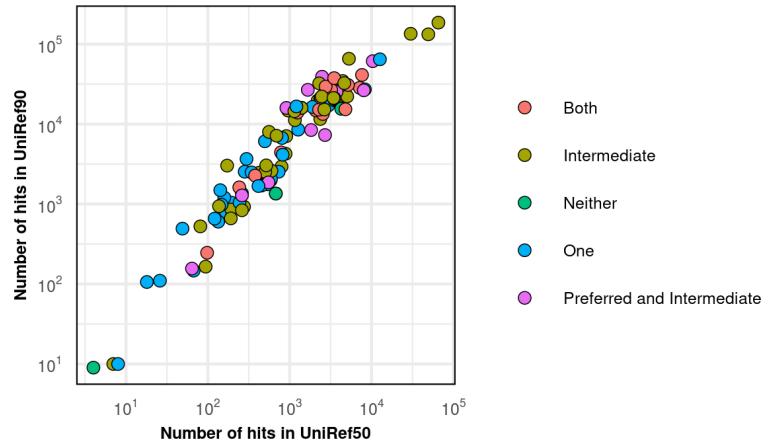


Figure S13: No correlation was observed between conformational sampling using language model masking and the number of hits in either the UniRef50 or UniRef90 sequence databases.

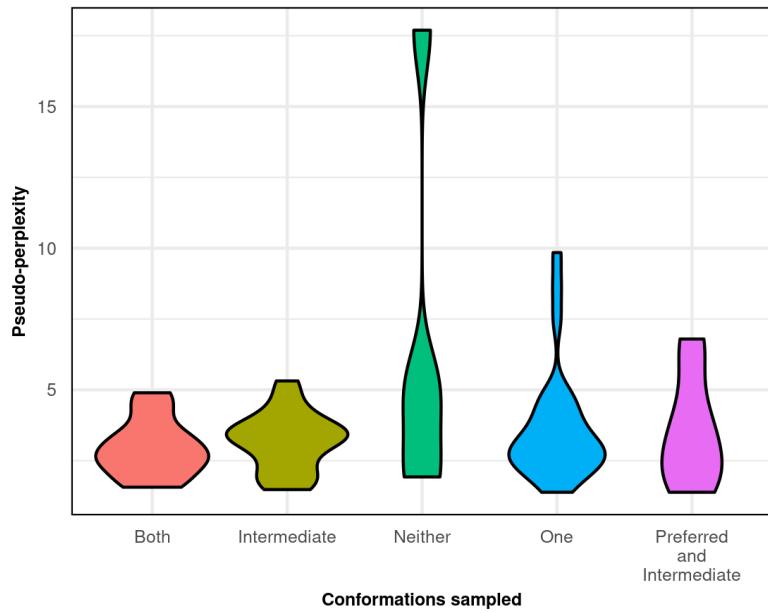


Figure S14: No correlation was observed between conformational sampling using language model masking and the pseudo-perplexity of the query sequences.