
Building Confidence in Deep Generative Protein Design

Tianyuan Zheng¹ Alessandro Rondina² Pietro Liò³

Abstract

Deep generative models show promise for *de novo* protein design, but their effectiveness within specific protein families remains underexplored. In this study, we evaluate two 3D rigid-body generative methods, score matching and flow matching, to generate monomeric protein backbones in SE(3) space. Our goal is to provide new insights and build confidence in the broader applicability of deep generative models for protein design. The optimal amino acid sequences were predicted from the generated backbones, followed by side-chain homology modeling. Results demonstrated high structural integrity, with conserved key residues aligning with known proteins. Structural phylogenetic analysis shows evolutionary links between the generated samples and their protein family members. Further molecular dynamics simulations and protein-ligand docking confirm the dynamic stability and functional potential of these samples, with ligand binding inducing conformational changes consistent with those in wild-type proteins.

1. Introduction

Protein functions and specificity are determined by their complex structures. Over the past 60 years, we have progressed from viewing protein design as unattainable to achieving complete artificial design and synthesis (Koren-dovych & DeGrado, 2020; Huang et al., 2016), with expanding applications across industries (Arunachalam et al., 2021; Kingwell, 2024; Barclay & Acharya, 2023). However, precise design remains challenging due to the nonlinear complexity of protein folding and the sensitivity of function to slight changes, still demanding significant resources.

¹Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK ²Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy ³Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. Correspondence to: Tianyuan Zheng <tz365@cam.ac.uk>.

With the rapid growth of protein structure databases (Jamasb et al., 2024; Berman, 2000), various *in silico de novo* protein design methods, particularly deep generative models (Watson et al., 2023; Ingraham et al., 2023; Wu et al., 2022), have emerged. Approaches such as diffusion-based score matching (SM) (Yim et al., 2023), and flow matching (FM) (Bose et al., 2023), show competitive performance for generating monomeric protein backbones in SE(3)¹ space.

While many methods show novelty, diversity, and designability in generated structures, few studies assess their functionality or evolutionary relevance, leaving the observed novelty and diversity potentially as mere hallucinations (Ji et al., 2023). Unclear biological functionality, unknown stability, and the lack of validation connecting these structures to known functions limit the broader application and advancement of generative methods.

This study evaluates the effectiveness of SM and FM methods across four well-studied protein families (β -lactamases, cytochrome *c*, GFP, and Ras). After determining optimal amino acid sequences and conducting side-chain homology modeling, we assess the functional relevance and evolutionary potential of the design, aiming to build confidence in deep generative approaches for protein design.

Our key contributions are: (1) We present a deep generative model pipeline for early-stage *de novo* protein design, with a robust evaluation protocol adaptable to a wider range of protein structure generation methods. (2) Despite sequence dissimilarity, generated samples show structural similarity to experimentally determined proteins in their families, conserving key residues linking to functions. (3) Structural phylogenetic analysis of both experimental and generated proteins reveals evolutionary relationships, showing the potential of structural phylogenetics. (4) Molecular dynamics (MD) simulations and protein-ligand docking show the dynamic stability and functional potential of generated samples. These samples form binding pockets akin to wild-type (WT) proteins, with conserved residues that bind family-specific ligands at low energies. For instance, generated KRas-like structures exhibit switch region flexibility and GTP/GDP-dependent conformational shifts similar to WT. (5) FM samples show greater flexibility, diversity, and novelty, while SM better captures conserved regions, offering

¹Special Euclidean group. See more in Appendix A.1.

insights for model selection and optimization in future work.

2. Background: Protein Backbone Generation

In this section and Appendix A, we review the key concepts behind Yim et al. (2023) and Bose et al. (2023)’s approaches for backbone generation, using SM and FM on SE(3).

Protein backbone representations. Molecules can be intuitively represented as atomic point clouds; however, for macromolecules, this approach often leads to high dimensionality and data sparsity. A more compact, fragment-based representation involves using backbone rigid groups from AlphaFold (Jumper et al., 2021). For each amino acid residue $i \in [1, N]$, its group consists of the backbone atoms $[N, C_\alpha, C, O]_i$ (Figure 1A). The goal is to learn how the rigid transformation (or *frame*) \mathbf{T}_i acts on the idealized coordinates, so that the transformed coordinates match the actual coordinates. For simplicity, $\mathbf{T}_i \in \text{SE}(3)$ decomposes into a rotation matrix $\mathbf{R}_i \in \text{SO}(3)$ and a translation vector $\mathbf{x}_i \in \mathbb{R}^3$. An extra dihedral angle $\psi_i \in \text{SO}(2)$ is introduced between the C_α and C bond to optimize the positioning of the backbone oxygen (O).

SE(3) score matching. Let $\mathbf{T}_t = [\mathbf{T}_{1,t}, \dots, \mathbf{T}_{N,t}]$ be the manifold of N frames at time t . The *forward process* gradually perturbs the distribution $\mathbf{T}_0 \sim p_0$, pushing it toward a simple distribution p_T , such as a Gaussian, by introducing a stochastic process independently in rotation and translation spaces (Equation (4)). The *reverse process* approximately reconstructs p_0 by iteratively calculating the gradient of the log probability density (or the *score*) $\nabla \log p$ along a reverse path (Equation (5) and Equation (6)). Since this score function is typically intractable in practice, a network $s(\theta, t, \cdot)$ is trained to approximate the conditional score $\nabla \log p_{t|0}$ at time $t \sim \mathcal{U}(0, T)$ (Vincent, 2011).

SE(3) flow matching. The *flow* from p_T to p_0 is defined as a set of continuously differentiable, time-indexed vector fields $\{\mathbf{u}_t\}$, where FM learns the time evolution of \mathbf{u}_t to gradually transport the data from p_T to p_0 . Because \mathbf{u}_t is generally intractable, Lipman et al. (2022) introduced conditional FM, training a neural network $v(\theta, t, \cdot)$ to approximate $\mathbf{u}_{t|0}$ (Equation (13)). When the optimal transport condition is satisfied, it provides a smooth and efficient transformation path, improving training stability and reducing computational complexity (Pooladian et al., 2023).

3. Evaluations of Generated Structures

3.1. Data

We generated monomeric proteins across a range of families, covering diverse fold types and functions. These proteins

include both natural and engineered mutations while preserving conserved core functional regions.

β -lactamases. β -lactamases are enzymes that deactivate β -lactam antibiotics by hydrolyzing their β -lactam ring, contributing significantly to bacterial resistance (Lee et al., 2016). Using β -lactamase inhibitors can restore antibiotic efficacy by blocking this reaction (Behzadi et al., 2020). Since their discovery, β -lactamase diversity has rapidly expanded due to the evolution of bacterial resistance, making them ideal for protein modeling studies. For this study, we gathered structural data on 1,578 unique monomeric β -lactamases and their variants across Ambler classes (A, B, C, and D) from the BLDB (Naas et al., 2017) and the Protein Data Bank (PDB) (Berman, 2000). Class A β -lactamases are the most prevalent, with conserved active-site residues Ser⁷⁰, Glu¹⁶⁶, and Asn¹⁷⁰ coordinating the hydrolytic water for deacylation (Tooke et al., 2019; Brown et al., 2009). Figure 1C shows the wild-type β -lactamase TEM1 (1BTL) alongside its acylated E166N intermediate (1FQG).

Cytochrome *c*. Cytochrome *c* is a water-soluble protein (~ 12 kDa) essential for ATP synthesis in mitochondria and intrinsic apoptosis (Kashyap et al., 2021; Ow et al., 2008). It also serves as an independent marker for apoptosis in several cancers (Li et al., 2001; Way et al., 2004). Despite variations across species, its core structure and function are conserved. We obtained structural data for 498 unique cytochrome *c* proteins and variants from the PDB.

Figure 1D shows horse cytochrome *c*, where a hydrophobic shell surrounds the heme group, with only $\sim 7.5\%$ of the surface available for electron transfer (Bushnell et al., 1990). The hydrophobic environment and iron coordination by His¹⁸ and Met⁸⁰ maintain a high redox potential (260 mV) (Salemme, 1977). Phosphorylation occurs at Thr²⁸, Thr⁴⁷, Tyr⁴⁸, and Tyr⁹⁷ (Hüttemann et al., 2011), while Lys⁷², Lys⁷³, and Lys⁸⁷ bind phospholipids (Kagan et al., 2009). The ATP-binding pocket includes Glu⁶⁹, Asn⁷⁰, Lys⁸⁸, and Lys⁷², Lys⁸⁶, Lys⁸⁷ (McIntosh et al., 1996).

Green fluorescent proteins (GFP). GFP, first isolated from *Aequorea victoria*, emits green fluorescence under specific wavelengths. Its core structure is an 11-strand β -barrel that encloses the chromophore (Figure 1E) (Remington, 2011). Various mutants have been engineered to enhance or modify its properties, including enhanced GFP (Cormack et al., 1996), superfolder GFP (Pédelacq et al., 2005), and color variants like YFP (Ormö et al., 1996) and BFP (Glaser et al., 2016). We collected structural data for 448 GFPs and variants from the PDB to explore this diversity.

Ras. Ras proteins, a subgroup of the small GTPase superfamily, act as molecular switches, cycling between GTP-

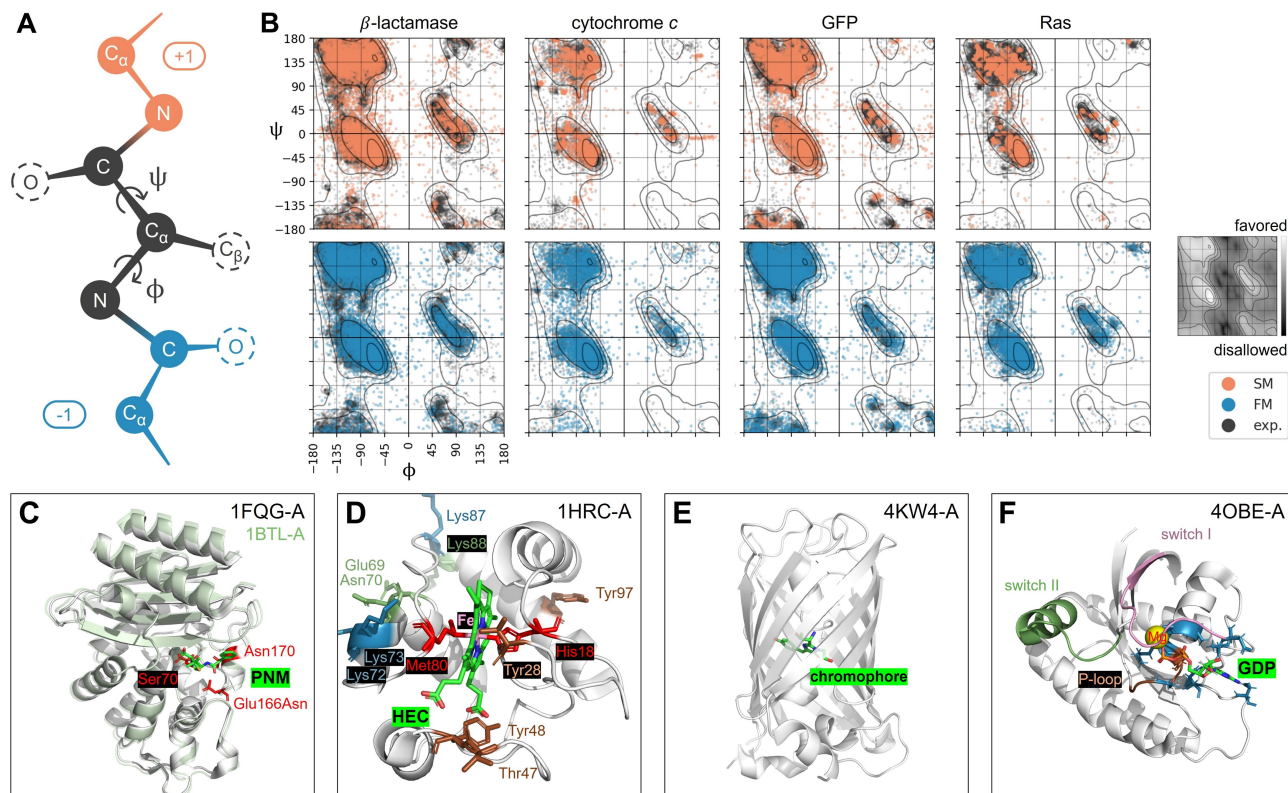


Figure 1. (A) Protein backbone and dihedral angles ψ and ϕ . (B) Ramachandran plots showing ψ and ϕ angle distributions for generated protein backbones vs. experimental proteins. The top-right inset shows conformationally favored (light) and disallowed (dark) angle regions. (C) Structural overlay of WT *E. coli* TEM1 (PDB: 1BTL; Jelsch et al. (1993); green) and its E166N acylated intermediate (PDB: 1FQG; Brown et al. (2009); white) with penicillin (PNM). (D) Structure of the WT *E. caballus* heart cytochrome *c* (PDB: 1HRC; Dickerson et al., 1967) with heme C (HEC). (E) Structure of *A. victoria* GFP and its chromophore (PDB: 4KW4; Barnard et al. (2014)). (F) Structure of GDP-bound *H. sapiens* KRas protein (PDB: 4OBE; Hunter et al., 2014).

bound (active) and GDP-bound (inactive) states to regulate cell proliferation, differentiation, migration, and apoptosis (Ladygina et al., 2011; Simanshu et al., 2017; Weinmann & Ottow, 2007). They play a key role in signaling from the cell surface to downstream pathways. Mutations that keep Ras proteins in an active state drive excessive cell growth and malignancy, making Ras inhibition a promising cancer treatment strategy (Singh & Lingham, 2002). We focused on the most common cancer-related Ras proteins (HRas, KRas, and NRas) (Cox, 2002) and obtained 511 experimental structures from the PDB.

In human KRas (Figure 1F), the switch I and II regions form the key interface for interacting with effectors and regulators (Pantsar, 2020). These regions are highly flexible, with conformations depending on GTP or GDP binding. Cancer-related mutations frequently occur in the P-loop and switch II (Pantsar, 2020).

3.2. Training and Sampling

We trained the SM and FM (with OT) models on the described protein families, using pretrained weights from Yim

et al. (2023) and Bose et al. (2023), with each model having ~ 17 million parameters. The sequence length distributions of the training proteins are shown in Figure 6. Each model generated 50 backbone structures with target lengths of 270 amino acids for β -lactamases, 113 for cytochrome *c*, 238 for GFP, and 170 for Ras proteins.

3.3. Backbone Dihedral Angles

Due to steric hindrance and spatial repulsion, not all backbone dihedral angles in proteins are physically feasible or energetically favorable. We analyzed the ϕ and ψ distribution in the generated structures using Ramachandran plots (Ramachandran et al., 1963) (Figure 1B). Most data points fall within the allowed and favored regions, with only a few in the disallowed areas, and no significant geometric clashes or unreasonable conformations were observed.

The dihedral angle distributions of the generated structures align well with those of the experimental training proteins. For instance, cytochrome *c* structures have sparse points in the region $-180^\circ < \psi < -90^\circ$ and $45^\circ < \phi < 180^\circ$, while GFP structures form four clusters there. SM samples

are closer to the training data and concentrate in allowed regions but show less diversity than FM samples.

3.4. Conserved Residue Consistency

In protein families or across species, certain residues are highly conserved, typically to ensure structural stability and support proper folding and function.

Determining the optimal amino acid sequence. Following Yim et al. (2023), we used ProteinMPNN (Dauparas et al., 2022) to predict ten sequences for each generated backbone. We then modeled these sequences with EMSfold (Rives et al., 2019) and compared them to the original backbone using the TM_{score} (Zhang, 2005). The sequence with the highest TM_{score} was chosen as the optimal match.

Identifying conserved residues. The experimental protein sequences were aligned with each generated backbone’s optimal sequence using Clustal Omega (Sievers et al., 2011). Given these multiple sequence alignments (MSAs) and generated structures, ConSurf (Yariv et al., 2023) reconstructed phylogenetic trees and applied Rate4Site (Pupko et al., 2002) to calculate evolutionary rates for each position with an empirical Bayesian method (Ashkenazy et al., 2016; Mayrose, 2004). The optimal predicted sequences for the generated backbones showed conserved residues matching those in the experimental sequences (Figure 8). FM samples displayed greater diversity than SM samples, suggesting higher novelty, as noted in Section 3.5. Interestingly, this trend reverses in generated β -lactamases, possibly due to the greater structural variability in the experimental β -lactamase training set (Figure 2B), to which SM models are more sensitive. Section 3.6 further assesses the dynamic stability and ligand-binding properties of these generated samples.

In Figure 2A, similar to the experimental proteins, the generated structures have higher residue conservation near binding pockets. For instance, in cytochrome *c* (1HRC) and generated structures SM-1 and FM-0, conserved residues cluster around the central heme C binding site. A similar pattern is seen in 4OBE and the generated KRas proteins.

3.5. Structural Phylogenetics

In specific tasks, we aim to generate structures with dissimilar sequences while preserving similar functions. However, when sequence similarity falls below 30%, detecting homology and evolutionary relationships becomes challenging (Puente-Lelievre et al., 2023). Structural comparison, being more conserved (Illergård et al., 2009), is more effective for uncovering evolutionary connections (Flores et al., 1993; Puente-Lelievre et al., 2023; Moi et al., 2023).

Using Q_{score} in structural phylogenetics. Malik et al. (2020) proposed the Q_{score} (Krissinel & Henrick, 2004) for structural phylogenetics, as it accommodates indels and combines alignment quality with length. Q_{score} compares the positions of all C_α atoms among N_{align} comparable residues in pairwise comparisons. We construct structural phylogenetic trees using $1 - Q_{score}$ as a distance measure, with higher values indicating greater structural similarity. For any two structures with N_1 and N_2 residues, we align and calculate Q_{score} using TM-align (Zhang, 2005) as:

$$Q_{score} = \frac{N_{align}^2}{N_1 N_2} \times \frac{1}{1 + \left(\frac{RMSD}{R_0}\right)^2} \quad (1)$$

where N_{align} is the number of aligned residues, RMSD is the root-mean-square deviation of atomic positions, and R_0 (set to 4Å) balances the contributions of RMSD and N_{align} .

Using 3Di alphabet in structural phylogenetics. van Kempen et al. (2023) developed Foldseek, which represents protein tertiary interactions as sequences over a 20-state 3D interaction (3Di) alphabet, simplifying structural alignments. This 3Di alphabet reduces false positives and increases information density by encoding conserved core regions more effectively. Leveraging Foldseek’s structural divergence metrics, Moi et al. (2023) constructed structural phylogenetic trees based on rigid body alignment, local alignment without superposition, and sequence alignment with structural alphabets, showing that these trees outperform traditional sequence-based trees across varied evolutionary timescales (Moi et al., 2023).

Summary tree. Two methods were used to calculate pairwise structural distances between generated and experimental structures, creating a distance matrix for constructing structural phylogenetic trees (Figure 7). After normalizing branch lengths, we generated a summary tree using SumTrees (Moreno et al., 2024), shown in Figure 2B. To ensure reliable Q_{score} calculations, residue count differences between compared structures were limited to within 10% (Malik et al., 2020). Thus, we included proteins with residue counts of 243-297 for β -lactamases, 102-124 for cytochrome *c*, 216-261 for GFP, and 166-187 for Ras proteins, excluding class C β -lactamases (~ 360 residues) from the tree but retaining them for model training and analysis.

The summary tree illustrates the potential of structural phylogenies, with each protein family clustering on distinct branches and subtypes (such as metallo- β -lactamases) grouping closely. Generated structures mostly integrate within small branches of experimental proteins, suggesting a close structural relationship. However, the models may not fully capture the specificity of class D β -lactamases, as no generated samples clustered in this class.

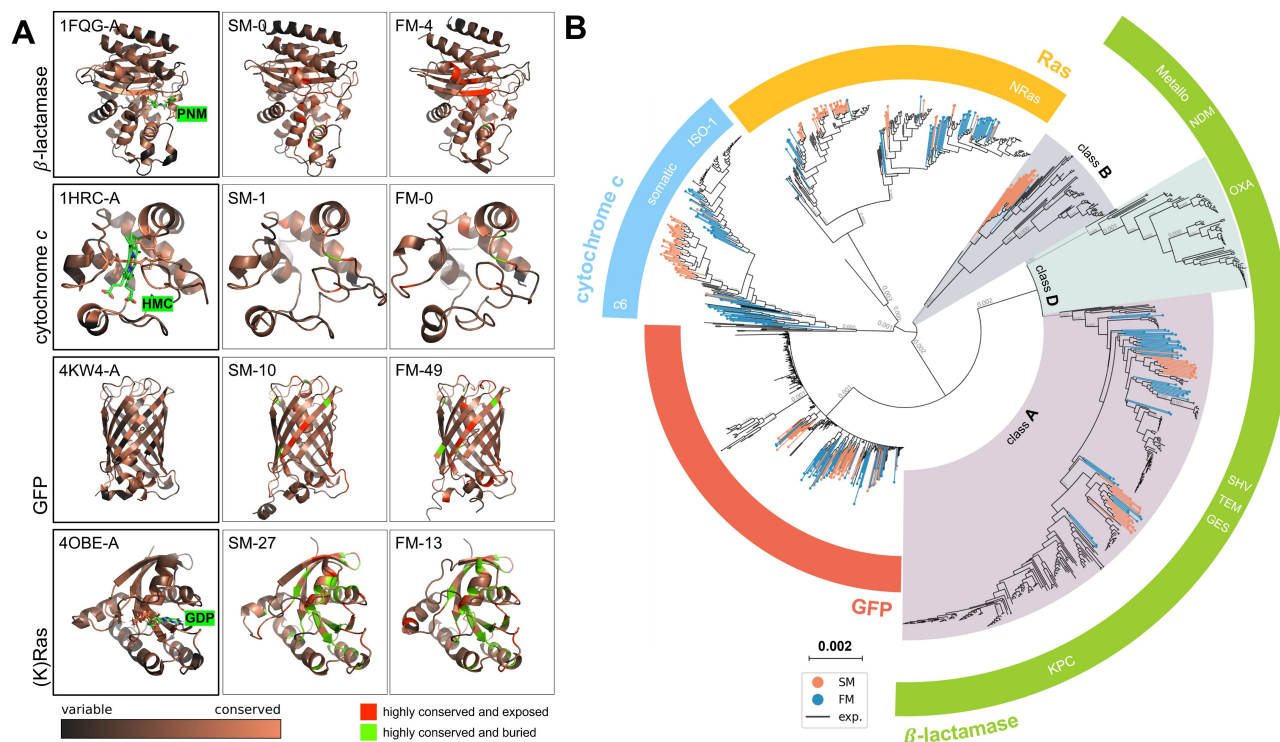


Figure 2. (A) Normalized evolutionary rates mapped onto structures, with black for rapidly evolving positions and orange for conserved ones. Experimental structures are highlighted with bold outlines, and highly conserved residues are marked in red (exposed) and green (buried). (B) Summary tree of two structural phylogenetic trees (Figure 7), constructed using Q_{score} and the 3Di alphabet. Protein family members are grouped in semi-circular rings, with specific subtypes labeled in white. In β -lactamases, different Ambler classes are shown with colored backgrounds. Colored branches and nodes indicate generated protein structures.

SM samples form dense clusters, while FM samples show greater diversity, positioning closer to the tree root but near experimental proteins without creating new branches, indicating higher novelty in FM outputs.

3.6. Molecular Dynamics

Structures that appear reasonable may, in fact, be unstable due to molecular dynamics, water interactions, and entropy, which are not considered during generation. To evaluate the dynamic stability of the generated structures, we conducted MD simulations under physiological conditions and analyzed their time-dependent behavior.

Homology modeling of side-chains. We used homology modeling to add side-chains to the protein backbone. After determining the optimal sequence (Section 3.4), we selected template proteins with at least 45% sequence identity and a TM_{score} of 0.75 or higher using FoldSeek (van Kempen et al., 2023). Sequences were aligned with Clustal Omega, and MODELLER (Šali & Blundell, 1993) generated possible side-chain conformations using statistical potentials and rotamer libraries, with backbone fixed. We chose the final side-chain arrangement based on lowest energy and minimal steric clashes, verifying model quality with PROCHECK

(Laskowski et al., 1993; 1996) and WHAT_CHECK (Hooft et al., 1996), discarding low-quality models.

Simulation. After adding hydrogen atoms with Reduce2 from the computational crystallography toolbox (Grosse-Kunstleve et al., 2002), we conducted MD simulations with GROMACS (Abraham et al., 2015). The protein was placed in an octahedral box with at least 15 Å from the edges, using the CHARMM36 force field (July 2022) (Vanommeslaeghe et al., 2009; Vanommeslaeghe & MacKerell, 2012; Yu et al., 2012; Soteras Gutiérrez et al., 2016) for intermolecular interactions. Following vacuum energy minimization, the system was solvated, neutralized to 150 mM Na^+ and Cl^- , and minimized again. The system was then heated to 310 K under NVT² conditions and equilibrated at 1 atm under NPT³ conditions, followed by a 10 ns production run. Details are provided in Appendix F.

Stability. Dynamic stability were analyzed as follows:

- (1) Root mean square deviation (RMSD) of backbone atoms over time, with stable proteins typically below 2Å, or up to 3Å for larger, flexible proteins (Burton et al., 2012; Liu

²Constant number of particles, volume, and temperature.

³Constant number of particles, pressure, and temperature.

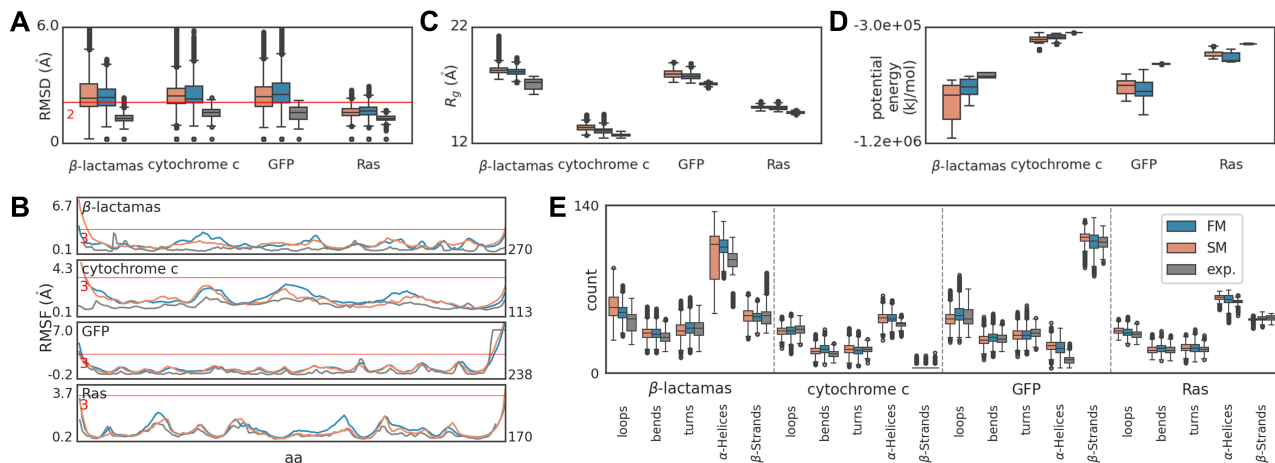


Figure 3. Stability assessment of MD simulations across proteins using various metrics. (A, C-E) Distributions of (A) RMSD, (C) radius of gyration (R_g), (D) potential energy, and (E) secondary structure counts throughout the simulation. Interquartile ranges and whiskers show metric variation; high-quality structures have medians close to experimental values with narrow ranges. (B) Residue fluctuation (RMSF) during simulation, where stable structures exhibit lower RMSF values and trends similar to the experimental structure.

et al., 2017; Wong & Wong, 2024). In Figure 3A, experimental structures have RMSD within 2Å, while generated structures average around 2.5Å, rarely exceeding 3Å. SM samples show RMSD \sim 0.3Å higher than FM samples. (2) Root mean square fluctuation (RMSF) measures the fluctuation of individual residues from their average positions, with values below 3Å considered stable (Burton et al., 2012; Oyewusi et al., 2024). Most residues in generated structures stay under 3Å, except for the first and last \sim 10 residues (Figure 3B). (3) Radius of gyration (R_g) quantifies the spatial distribution of a molecule’s atoms relative to its center of mass. The generated structures are expected to be compact, with R_g values close to experimental structures and fluctuations around 1Å (Figure 3C). (4) The DSSP algorithm assigns secondary structure to each residue based on hydrogen bonds and geometry. In Figure 3D, the stability in secondary structure elements over time suggests structural stability with no major conformational changes. (5) Lower potential energy are generally more stable, with contributions from bond, angle, dihedral, van der Waals, and electrostatic energies. In Figure 3E, most generated samples have lower potential energy than the experimental data.

3.7. Protein-ligand Docking

Using AutoDock Vina (Trott & Olson, 2009; Eberhardt et al., 2021), we predicted optimal binding modes between generated structures and their family-specific ligands. We performed blind docking, scanning the entire protein surface for potential ligand binding sites without prior knowledge of binding pockets. The grid box covered the entire protein, and identical configurations were applied to both generated and experimental structures to evaluate if deviations in generated samples fell within an acceptable range, thus assessing their functional viability. Details on simulation

settings are provided in Appendix G.

Docking on experimental structures closely matches known binding modes, with ligand deviations (RMSD) around 1.5Å and low binding energies (Figure 4). Successful docking typically shows binding free energies (ΔG) between -7 and -10 kcal/mol, where lower values indicate stronger, more stable interactions (P. & M. K., 2021; Nguyen et al., 2019).

The generated protein structures have pockets similar to the experimental proteins (Figure 4). In most simulations, ligands bind within 4Å of experimental positions and binding energies below -6 kcal/mol. SM samples generally show lower binding energies and RMSD values than FM samples.

β -lactamases binding to penicillin. Mutations at Glu¹⁶⁶ and Asn¹⁷⁰ in class A β -lactamases (Figure 1C) can form a stable acyl-enzyme intermediate, disrupting deacylation (Chen & Herzberg, 2001). Only asparagine and glycine at position 170 preserve WT-like function (Brown et al., 2009), and this conservation is retained in generated structures like FM-4 (Figure 4A), suggesting they may retain the ability to inactivate penicillin antibiotics.

Cytochrome c binding to heme c. The generated cytochrome c-like structures retain conserved residues found in the WT (Figure 1D; 1HRC). SM-1 (Figure 4B) includes phosphorylatable residues Tyr⁵⁸, Thr⁵⁹, and Tyr¹⁰⁷, as well as lysine residues Lys⁸², Lys⁸³, and Lys⁹⁶ involved in phospholipid binding. Asn⁸⁰ and these lysines form an ATP-binding pocket-like structure. FM-0 shows heme iron coordinated by two cysteines, which may form stronger covalent bonds, potentially affecting electron transfer efficiency.

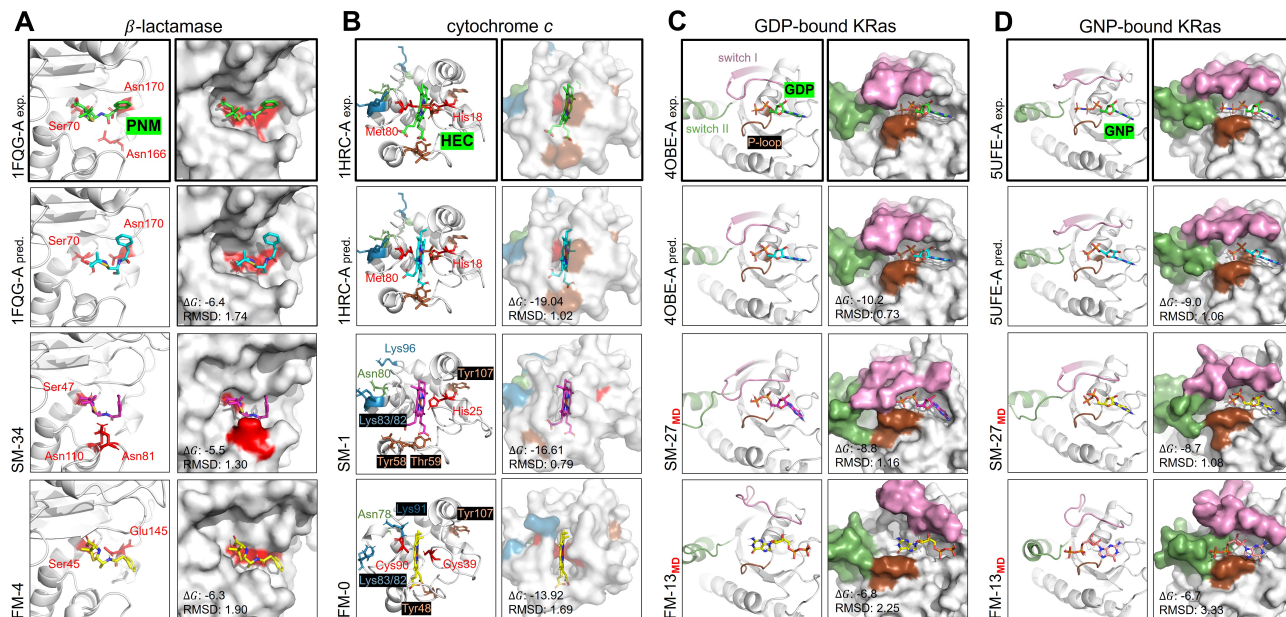


Figure 4. Comparison of experimentally determined bindings (bold boxes) with predictions from blind docking simulations, using models closest to experimental data. Ligands are shown in green (experimental data), cyan (predictions from experimental receptors), magenta (predictions from SM samples), and yellow (predictions from FM samples). ΔG is in kcal/mol, with RMSD indicating deviation from experimental ligand positions (in Å). Dockings labeled “MD” represent protein-ligand complex MD simulations after blind docking.

KRas binding to GNP/GDP. GNP, a non-hydrolyzable GTP analog, is commonly used to simulate the GTP-bound active state of Ras proteins (Pantsar, 2020). In WT KRas bound to GNP (Figure 4D; 5UFE), residues from the Switch II region fill the pocket. Removal of the γ -phosphate relaxes the protein conformation to the GDP-bound state, opening the switch II region (Figure 4C; 4OBE) (Kauke et al., 2017).

After blind docking the generated KRas-like structures with GNP/GDP, we analyzed the complexes using protein-ligand MD simulations. Allowing flexibility in both the backbone and side-chains enabled us to capture the dynamic conformational changes, especially in the switch regions, between active/inactive states. FM-generated KRas-like structures show greater flexibility. In FM-13 (Figure 4D), GNP binding causes switch II residues to fill the pocket, similar to 5UFE. In the GDP-bound state, FM-13 adopts an open conformation, with the channel between the switches and P-loop opening, as seen in 4OBE (Figure 4C). In contrast, SM-27 are more rigid, with fewer conformational changes between GDP- and GTP-bound states.

4. Discussion

This study demonstrates the potential of SM and FM methods for generating protein backbones. After determining the sequence, we performed side-chain homology modeling. Despite sequence dissimilarities, key residues are conserved, and the generated structures are highly similar. SM better captures conserved regions, producing more rigid structures,

while FM offers greater flexibility. Structural phylogenetic analysis revealed evolutionary relationships, and MD and docking simulations confirmed stability and functionality. The presented pipeline and evaluation protocols build confidence for broader applications in protein generation.

SM and FM can generate a range of monomeric protein structures and can have applications beyond protein design (Yim et al., 2023; Bose et al., 2023). However, the complexity of these tasks is often underestimated, and function verification for novel samples remains costly. Targeting specific tasks or integrating generative methods into well-established empirical knowledge may yield better results. Key factors like conformational dynamics, water interactions, and entropy have not been fully considered in generation (Du et al., 2024). Incorporating protein sequences, side-chain details, and functional annotations as context could improve model performance. Although some progress has been made (Torge et al., 2023; Jin et al., 2023; Somnath et al., 2023; Zhou et al., 2023), research gaps remain.

These generative processes differ from protein synthesis simulations and do not reflect human hierarchical understanding. Limited interpretability reduces their utility for further downstream analyses and industrial applications. Although generated proteins may resemble real ones and include similar conserved residues and conformational changes, small variations could significantly impact function. Simply aiming for diversity and novelty does not guarantee quality.

Generative models have also shown great potential as effi-

cient surrogates for MD simulations, which are often computationally intensive for complex systems or long-term behaviors. Jing et al. (2024) developed a framework to simulate molecular trajectories using generative models. Similarly, Viguera Diez et al. (2023) used generative models to enhance sampling of slow degrees of freedom, covering the sample space more effectively than traditional MD.

Software and Data

All input data and software used in this study are available from public sources or provided under academic licenses. The source code, scripts, generated samples, and curated datasets can be accessed at <https://github.com/ECburx/PROTEVAL>. The PDB structure files were downloaded on June 19, 2024, from <https://www.wwpdb.org/ftp/pdb-ftp-sites>.

Acknowledgments

The authors would like to thank Charlie Harris and Gos Micklem for their fruitful discussions and valuable feedback on this work.

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, September 2015. ISSN 2352-7110. doi: 10.1016/j.softx.2015.06.001. URL <http://dx.doi.org/10.1016/j.softx.2015.06.001>.
- Anand, N. and Achim, T. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models, 2022. URL <https://arxiv.org/abs/2205.15019>.
- Arunachalam, P. S., Walls, A. C., Golden, N., Atyeo, C., Fischinger, S., Li, C., Aye, P., Navarro, M. J., Lai, L., Edara, V. V., Röltgen, K., Rogers, K., Shirreff, L., Ferrell, D. E., Wrenn, S., Pettie, D., Kraft, J. C., Miranda, M. C., Kepl, E., Sydeman, C., Brunette, N., Murphy, M., Fiala, B., Carter, L., White, A. G., Trisal, M., Hsieh, C.-L., Russell-Lodrigue, K., Monjure, C., Dufour, J., Spencer, S., Doyle-Meyers, L., Bohm, R. P., Maness, N. J., Roy, C., Plante, J. A., Plante, K. S., Zhu, A., Gorman, M. J., Shin, S., Shen, X., Fontenot, J., Gupta, S., O’Hagan, D. T., Van Der Most, R., Rappuoli, R., Coffman, R. L., Novack, D., McLellan, J. S., Subramaniam, S., Montefiori, D., Boyd, S. D., Flynn, J. L., Alter, G., Villinger, F., Kleanthous, H., Rappaport, J., Suthar, M. S., King, N. P., Veessler, D., and Pulendran, B. Adjuvanting a subunit covid-19 vaccine to induce protective immunity. *Nature*, 594(7862):253–258, April 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03530-2. URL <http://dx.doi.org/10.1038/s41586-021-03530-2>.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, 44 (W1):W344–W350, May 2016. ISSN 1362-4962. doi: 10.1093/nar/gkw408. URL <http://dx.doi.org/10.1093/nar/gkw408>.
- Barclay, A. and Acharya, K. R. Engineering plastic eating enzymes using structural biology. *Biomolecules*, 13 (9):1407, September 2023. ISSN 2218-273X. doi: 10.3390/biom13091407. URL <http://dx.doi.org/10.3390/biom13091407>.
- Barnard, T., Yu, X., Noinaj, N., and Taraska, J. Crystal structure of green fluorescent protein, April 2014. URL <http://dx.doi.org/10.2210/pdb4KW4/pdb>.
- Behzadi, P., García-Perdomo, H. A., Karpiński, T. M., and Issakhanian, L. Metallo- β -lactamases: a review. *Molecular Biology Reports*, 47(8):6281–6294, July 2020. ISSN 1573-4978. doi: 10.1007/s11033-020-05651-9. URL <http://dx.doi.org/10.1007/s11033-020-05651-9>.
- Berman, H. M. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, January 2000. ISSN 1362-4962. doi: 10.1093/nar/28.1.235. URL <http://dx.doi.org/10.1093/nar/28.1.235>.
- Bose, A. J., Akhound-Sadegh, T., Huguet, G., Fatras, K., Rector-Brooks, J., Liu, C.-H., Nica, A. C., Korablyov, M., Bronstein, M., and Tong, A. SE(3)-stochastic flow matching for protein backbone generation, 2023. URL <https://arxiv.org/abs/2310.02391>.
- Brown, N. G., Shanker, S., Prasad, B., and Palzkill, T. Structural and biochemical evidence that a tem-1 β -lactamase n170g active site mutant acts via substrate-assisted catalysis. *Journal of Biological Chemistry*, 284(48):33703–33712, November 2009. ISSN 0021-9258. doi: 10.1074/jbc.M109.053819. URL <http://dx.doi.org/10.1074/jbc.M109.053819>.
- Burton, B., Zimmermann, M. T., Jernigan, R. L., and Wang, Y. A computational investigation on the connection between dynamics properties of ribosomal proteins and ribosome assembly. *PLoS Computational Biology*, 8(5):e1002530, May 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002530. URL <http://dx.doi.org/10.1371/journal.pcbi.1002530>.

- Bushnell, G. W., Louie, G. V., and Brayer, G. D. High-resolution three-dimensional structure of horse heart cytochrome c. *Journal of Molecular Biology*, 214(2):585–595, July 1990. ISSN 0022-2836. doi: 10.1016/0022-2836(90)90200-6. URL [http://dx.doi.org/10.1016/0022-2836\(90\)90200-6](http://dx.doi.org/10.1016/0022-2836(90)90200-6).
- Chen, C. C. H. and Herzberg, O. Structures of the acyl-enzyme complexes of the staphylococcus aureus β -lactamase mutant glu166asp asn170gln with benzylpenicillin and cephaloridine. *Biochemistry*, 40(8):2351–2358, January 2001. ISSN 1520-4995. doi: 10.1021/bi002277h. URL <http://dx.doi.org/10.1021/bi002277h>.
- Chen, R. T. Q. and Lipman, Y. Flow matching on general geometries. 2023. doi: 10.48550/ARXIV.2302.03660. URL <https://arxiv.org/abs/2302.03660>.
- Cormack, B. P., Valdivia, R. H., and Falkow, S. Facs-optimized mutants of the green fluorescent protein (gfp). *Gene*, 173(1):33–38, 1996. ISSN 0378-1119. doi: 10.1016/0378-1119(95)00685-0. URL [http://dx.doi.org/10.1016/0378-1119\(95\)00685-0](http://dx.doi.org/10.1016/0378-1119(95)00685-0).
- Cox, A. Farnesyltransferase inhibitors: promises and realities. *Current Opinion in Pharmacology*, 2(4):388–393, August 2002. ISSN 1471-4892. doi: 10.1016/S1471-4892(02)00181-9. URL [http://dx.doi.org/10.1016/S1471-4892\(02\)00181-9](http://dx.doi.org/10.1016/S1471-4892(02)00181-9).
- 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, October 2022. ISSN 1095-9203. doi: 10.1126/science.add2187. URL <http://dx.doi.org/10.1126/science.add2187>.
- De Bortoli, V., Mathieu, E., Hutchinson, M., Thornton, J., Teh, Y. W., and Doucet, A. Riemannian score-based generative modelling, 2022. URL <https://arxiv.org/abs/2202.02763>.
- Dickerson, R. E., Kopka, M. L., Borders, C. L., Varnum, J., Weinzierl, J. E., and Margoliash, E. A centrosymmetric projection at 4 Å of horse heart oxidized cytochrome c. *Journal of Molecular Biology*, 29(1):77–95, October 1967. ISSN 0022-2836. doi: 10.1016/0022-2836(67)90182-9. URL [http://dx.doi.org/10.1016/0022-2836\(67\)90182-9](http://dx.doi.org/10.1016/0022-2836(67)90182-9).
- Du, Y., Jamasb, A. R., Guo, J., Fu, T., Harris, C., Wang, Y., Duan, C., Liò, P., Schwaller, P., and Blundell, T. L. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6(6):589–604, June 2024. ISSN 2522-5839. doi: 10.1038/s42256-024-00843-5. URL <http://dx.doi.org/10.1038/s42256-024-00843-5>.
- Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. Autodock vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, July 2021. ISSN 1549-960X. doi: 10.1021/acs.jcim.1c00203. URL <http://dx.doi.org/10.1021/acs.jcim.1c00203>.
- Flores, T., Orengo, C., Moss, D., and Thornton, J. Comparison of conformational characteristics in structurally similar protein pairs. *Protein Science*, 2(11):1811–1826, November 1993. ISSN 1469-896X. doi: 10.1002/pro.5560021104. URL <http://dx.doi.org/10.1002/pro.5560021104>.
- Glaser, A., McColl, B., and Vadolas, J. Gfp to bfp conversion: A versatile assay for the quantification of crispr/cas9-mediated genome editing. *Molecular Therapy - Nucleic Acids*, 5:e334, 2016. ISSN 2162-2531. doi: 10.1038/mtna.2016.48. URL <http://dx.doi.org/10.1038/mtna.2016.48>.
- Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W., and Adams, P. D. The computational crystallography toolbox: crystallographic algorithms in a reusable software framework. *Journal of Applied Crystallography*, 35(1):126–136, January 2002. ISSN 0021-8898. doi: 10.1107/S0021889801017824. URL <http://dx.doi.org/10.1107/S0021889801017824>.
- Hoof, R. W. W., Vriend, G., Sander, C., and Abola, E. E. Errors in protein structures. *Nature*, 381(6580):272–272, May 1996. ISSN 1476-4687. doi: 10.1038/381272a0. URL <http://dx.doi.org/10.1038/381272a0>.
- Huang, P.-S., Boyken, S. E., and Baker, D. The coming of age of de novo protein design. *Nature*, 537(7620):320–327, September 2016. ISSN 1476-4687. doi: 10.1038/nature19946. URL <http://dx.doi.org/10.1038/nature19946>.
- Hunter, J. C., Gurbani, D., Ficarro, S. B., Carrasco, M. A., Lim, S. M., Choi, H. G., Xie, T., Marto, J. A., Chen, Z., Gray, N. S., and Westover, K. D. In situ selectivity profiling and crystal structure of sml-8-73-1, an active site inhibitor of oncogenic k-ras g12c. *Proceedings of the National Academy of Sciences*, 111(24):8895–8900, June 2014. ISSN 1091-6490. doi: 10.1073/pnas.1404639111. URL <http://dx.doi.org/10.1073/pnas.1404639111>.

- Hüttemann, M., Pecina, P., Rainbolt, M., Sanderson, T. H., Kagan, V. E., Samavati, L., Doan, J. W., and Lee, I. The multiple functions of cytochrome c and their regulation in life and death decisions of the mammalian cell: From respiration to apoptosis. *Mitochondrion*, 11(3):369–381, May 2011. ISSN 1567-7249. doi: 10.1016/j.mito.2011.01.010. URL <http://dx.doi.org/10.1016/j.mito.2011.01.010>.
- Illergård, K., Ardell, D. H., and Elofsson, A. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins: Structure, Function, and Bioinformatics*, 77(3):499–508, June 2009. ISSN 1097-0134. doi: 10.1002/prot.22458. URL <http://dx.doi.org/10.1002/prot.22458>.
- Ingraham, J. B., Baranov, M., Costello, Z., Barber, K. W., Wang, W., Ismail, A., Frappier, V., Lord, D. M., Ng-Thow-Hing, C., Van Vlack, E. R., Tie, S., Xue, V., Cowles, S. C., Leung, A., Rodrigues, J. V., Morales-Perez, C. L., Ayoub, A. M., Green, R., Puentes, K., Oplinger, F., Panwar, N. V., Obermeyer, F., Root, A. R., Beam, A. L., Poelwijk, F. J., and Grigoryan, G. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, November 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06728-8. URL <http://dx.doi.org/10.1038/s41586-023-06728-8>.
- 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, 2024. URL <https://arxiv.org/abs/2406.13864>.
- Jelsch, C., Mourey, L., Masson, J., and Samama, J. Crystal structure of escherichia coli tem1 β -lactamase at 1.8 Å resolution. *Proteins: Structure, Function, and Bioinformatics*, 16(4):364–383, August 1993. ISSN 1097-0134. doi: 10.1002/prot.340160406. URL <http://dx.doi.org/10.1002/prot.340160406>.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, March 2023. ISSN 1557-7341. doi: 10.1145/3571730. URL <http://dx.doi.org/10.1145/3571730>.
- Jin, W., Sarkizova, S., Chen, X., Hacoheh, N., and Uhler, C. Unsupervised protein-ligand binding energy prediction via neural euler’s rotation equation, 2023. URL <https://arxiv.org/abs/2301.10814>.
- Jing, B., Stärk, H., Jaakkola, T., and Berger, B. Generative modeling of molecular dynamics trajectories, 2024. URL <https://arxiv.org/abs/2409.17808>.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, July 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2. URL <http://dx.doi.org/10.1038/s41586-021-03819-2>.
- Kagan, V. E., Bayır, H. A., Belikova, N. A., Kapralov, O., Tyurina, Y. Y., Tyurin, V. A., Jiang, J., Stoyanovsky, D. A., Wipf, P., Kochanek, P. M., Greenberger, J. S., Pitt, B., Shvedova, A. A., and Borisenko, G. Cytochrome c/cardiolipin relations in mitochondria: a kiss of death. *Free Radical Biology and Medicine*, 46(11):1439–1453, June 2009. ISSN 0891-5849. doi: 10.1016/j.freeradbiomed.2009.03.004. URL <http://dx.doi.org/10.1016/j.freeradbiomed.2009.03.004>.
- Kashyap, D., Garg, V. K., and Goel, N. *Intrinsic and extrinsic pathways of apoptosis: Role in cancer development and prognosis*, pp. 73–120. Elsevier, 2021. ISBN 9780323853156. doi: 10.1016/bs.apcsb.2021.01.003. URL <http://dx.doi.org/10.1016/bs.apcsb.2021.01.003>.
- Kauke, M. J., Traxlmayr, M. W., Parker, J. A., Kiefer, J. D., Knihtila, R., McGee, J., Verdine, G., Mattos, C., and Wittrup, K. D. An engineered protein antagonist of k-ras/b-raf interaction. *Scientific Reports*, 7(1), July 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-05889-7. URL <http://dx.doi.org/10.1038/s41598-017-05889-7>.
- Kingwell, K. Ribosome-targeted antibiotic tackles antimicrobial resistance. *Nature Reviews Drug Discovery*, 23(4):249–249, March 2024. ISSN 1474-1784. doi: 10.1038/d41573-024-00040-4. URL <http://dx.doi.org/10.1038/d41573-024-00040-4>.
- Korendovych, I. V. and DeGrado, W. F. De novo protein design, a retrospective. *Quarterly Reviews of Biophysics*, 53, 2020. ISSN 1469-8994. doi: 10.1017/s0033583519000131. URL <http://dx.doi.org/10.1017/s0033583519000131>.
- Krissinel, E. and Henrick, K. Secondary-structure matching (ssm), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D Biological Crystallography*, 60(12):2256–2268,

- November 2004. ISSN 0907-4449. doi: 10.1107/s0907444904026460. URL <http://dx.doi.org/10.1107/S0907444904026460>.
- Ladygina, N., Martin, B. R., and Altman, A. *Dynamic Palmitoylation and the Role of DHHC Proteins in T Cell Activation and Anergy*, pp. 1–44. Elsevier, 2011. ISBN 9780123876645. doi: 10.1016/b978-0-12-387664-5.00001-7. URL <http://dx.doi.org/10.1016/B978-0-12-387664-5.00001-7>.
- Laskowski, R., Rullmann, J., MacArthur, M., Kaptein, R., and Thornton, J. Aqua and procheck-nmr: Programs for checking the quality of protein structures solved by nmr. *Journal of Biomolecular NMR*, 8(4), December 1996. ISSN 1573-5001. doi: 10.1007/bf00228148. URL <http://dx.doi.org/10.1007/BF00228148>.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. Procheck: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2): 283–291, April 1993. ISSN 0021-8898. doi: 10.1107/s0021889892009944. URL <http://dx.doi.org/10.1107/S0021889892009944>.
- Lee, D., Das, S., Dawson, N. L., Dobrijevic, D., Ward, J., and Orengo, C. Novel computational protocols for functionally classifying and characterising serine beta-lactamases. *PLOS Computational Biology*, 12(6): e1004926, June 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004926. URL <http://dx.doi.org/10.1371/journal.pcbi.1004926>.
- Li, M., Wu, X., and Xu, X.-C. Induction of apoptosis by cyclo-oxygenase-2 inhibitor ns398 through a cytochrome-c-dependent pathway in esophageal cancer cells. *International Journal of Cancer*, 93(2):218–223, 2001. ISSN 1097-0215. doi: 10.1002/ijc.1322. URL <http://dx.doi.org/10.1002/ijc.1322>.
- Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling, 2022. URL <https://arxiv.org/abs/2210.02747>.
- Liu, K., Watanabe, E., and Kokubo, H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *Journal of Computer-Aided Molecular Design*, 31(2):201–211, January 2017. ISSN 1573-4951. doi: 10.1007/s10822-016-0005-2. URL <http://dx.doi.org/10.1007/s10822-016-0005-2>.
- Malik, A. J., Poole, A. M., and Allison, J. R. Structural phylogenetics with confidence. *Molecular Biology and Evolution*, 37(9):2711–2726, April 2020. ISSN 1537-1719. doi: 10.1093/molbev/msaa100. URL <http://dx.doi.org/10.1093/molbev/msaa100>.
- Mayrose, I. Comparison of site-specific rate-inference methods for protein sequences: Empirical bayesian methods are superior. *Molecular Biology and Evolution*, 21(9):1781–1791, May 2004. ISSN 1537-1719. doi: 10.1093/molbev/msh194. URL <http://dx.doi.org/10.1093/molbev/msh194>.
- McIntosh, D. B., Parrish, J. C., and Wallace, C. J. Definition of a nucleotide binding site on cytochrome c by photoaffinity labeling. *Journal of Biological Chemistry*, 271(31):18379–18386, August 1996. ISSN 0021-9258. doi: 10.1074/jbc.271.31.18379. URL <http://dx.doi.org/10.1074/jbc.271.31.18379>.
- Moi, D., Bernard, C., Steinegger, M., Nevers, Y., Langleib, M., and Dessimoz, C. Structural phylogenetics unravels the evolutionary diversification of communication systems in gram-positive bacteria and their viruses. September 2023. doi: 10.1101/2023.09.19.558401. URL <http://dx.doi.org/10.1101/2023.09.19.558401>.
- Moreno, M. A., Holder, M. T., and Sukumaran, J. Dendropy 5: a mature python library for phylogenetic computing, 2024. URL <https://doi.org/10.21105/joss.06943>.
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, April 2009. ISSN 1096-987X. doi: 10.1002/jcc.21256. URL <http://dx.doi.org/10.1002/jcc.21256>.
- Naas, T., Oueslati, S., Bonnin, R. A., Dabos, M. L., Zavala, A., Dortet, L., Retailleau, P., and Iorga, B. I. Beta-lactamase database (bldb) – structure and function. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 32(1):917–919, January 2017. ISSN 1475-6374. doi: 10.1080/14756366.2017.1344235. URL <http://dx.doi.org/10.1080/14756366.2017.1344235>.
- Nguyen, N. T., Nguyen, T. H., Pham, T. N. H., Huy, N. T., Bay, M. V., Pham, M. Q., Nam, P. C., Vu, V. V., and Ngo, S. T. Autodock vina adopts more accurate binding poses but autodock4 forms better binding affinity. *Journal of Chemical Information and Modeling*, 60(1):204–211, December 2019. ISSN 1549-960X. doi: 10.1021/acs.jcim.9b00778. URL <http://dx.doi.org/10.1021/acs.jcim.9b00778>.
- Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., and Remington, S. J. Crystal structure of the aequorea victoria green fluorescent protein. *Science*, 273(5280):1392–1395, September 1996. ISSN 1095-9203.

- doi: 10.1126/science.273.5280.1392. URL <http://dx.doi.org/10.1126/science.273.5280.1392>.
- Ow, Y.-L. P., Green, D. R., Hao, Z., and Mak, T. W. Cytochrome c: functions beyond respiration. *Nature Reviews Molecular Cell Biology*, 9(7):532–542, July 2008. ISSN 1471-0080. doi: 10.1038/nrm2434. URL <http://dx.doi.org/10.1038/nrm2434>.
- Oyewusi, H. A., Wahab, R. A., Akinyede, K. A., Albadrani, G. M., Al-Ghadi, M. Q., Abdel-Daim, M. M., Ajiboye, B. O., and Huyop, F. Bioinformatics analysis and molecular dynamics simulations of azoreductases (azrbmh2) from bacillus megaterium h2 for the decolorization of commercial dyes. *Environmental Sciences Europe*, 36(1), February 2024. ISSN 2190-4715. doi: 10.1186/s12302-024-00853-5. URL <http://dx.doi.org/10.1186/s12302-024-00853-5>.
- O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. Open babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1), October 2011. ISSN 1758-2946. doi: 10.1186/1758-2946-3-33. URL <http://dx.doi.org/10.1186/1758-2946-3-33>.
- P., G. and M. K., K. Docking studies and molecular dynamics simulation of triazole benzene sulfonamide derivatives with human carbonic anhydrase ix inhibition activity. *RSC Advances*, 11(60):38079–38093, 2021. ISSN 2046-2069. doi: 10.1039/d1ra07377j. URL <http://dx.doi.org/10.1039/D1RA07377J>.
- Pantsar, T. The current understanding of kras protein structure and dynamics. *Computational and Structural Biotechnology Journal*, 18:189–198, 2020. ISSN 2001-0370. doi: 10.1016/j.csbj.2019.12.004. URL <http://dx.doi.org/10.1016/j.csbj.2019.12.004>.
- Pollard, D. *A User’s Guide to Measure Theoretic Probability*. Cambridge University Press, December 2001. ISBN 9780511811555. doi: 10.1017/cbo9780511811555. URL <http://dx.doi.org/10.1017/CBO9780511811555>.
- Pooladian, A.-A., Ben-Hamu, H., Domingo-Enrich, C., Amos, B., Lipman, Y., and Chen, R. T. Q. Multisample flow matching: Straightening flows with minibatch couplings, 2023. URL <https://arxiv.org/abs/2304.14772>.
- Puente-Lelievre, C., Malik, A. J., Douglas, J., Ascher, D., Baker, M., Allison, J., Poole, A., Lundin, D., Fullmer, M., Bouckert, R., Kim, H., Steinegger, M., and Matzke, N. Tertiary-interaction characters enable fast, model-based structural phylogenetics beyond the twilight zone. December 2023. doi: 10.1101/2023.12.12.571181. URL <http://dx.doi.org/10.1101/2023.12.12.571181>.
- Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., and Bental, N. Rate4site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, 18(suppl.1):S71–S77, July 2002. ISSN 1367-4803. doi: 10.1093/bioinformatics/18.suppl_1.s71. URL http://dx.doi.org/10.1093/bioinformatics/18.suppl_1.s71.
- Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24(1):79–88, December 2005. ISSN 1546-1696. doi: 10.1038/nbt1172. URL <http://dx.doi.org/10.1038/nbt1172>.
- Ramachandran, G., Ramakrishnan, C., and Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, July 1963. ISSN 0022-2836. doi: 10.1016/s0022-2836(63)80023-6. URL [http://dx.doi.org/10.1016/s0022-2836\(63\)80023-6](http://dx.doi.org/10.1016/s0022-2836(63)80023-6).
- Remington, S. J. Green fluorescent protein: A perspective. *Protein Science*, 20(9):1509–1519, July 2011. ISSN 1469-896X. doi: 10.1002/pro.684. URL <http://dx.doi.org/10.1002/pro.684>.
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., and Fergus, R. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS*, 2019. doi: 10.1101/622803. URL <https://www.biorxiv.org/content/10.1101/622803v4>.
- Rudolph, M., Wandt, B., and Rosenhahn, B. Same same but different: Semi-supervised defect detection with normalizing flows, 2020. URL <https://arxiv.org/abs/2008.12577>.
- Salemme, F. R. Structure and function of cytochromes c. *Annual Review of Biochemistry*, 46(1):299–330, June 1977. ISSN 1545-4509. doi: 10.1146/annurev.bi.46.070177.001503. URL <http://dx.doi.org/10.1146/annurev.bi.46.070177.001503>.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, 7(1), January 2011. ISSN 1744-4292. doi: 10.1038/msb.2011.75. URL <http://dx.doi.org/10.1038/msb.2011.75>.

- Simanshu, D. K., Nissley, D. V., and McCormick, F. Ras proteins and their regulators in human disease. *Cell*, 170(1):17–33, June 2017. ISSN 0092-8674. doi: 10.1016/j.cell.2017.06.009. URL <http://dx.doi.org/10.1016/j.cell.2017.06.009>.
- Singh, S. B. and Lingham, R. B. Current progress on farnesyl protein transferase inhibitors. *Current Opinion in Drug Discovery and Development*, 5(2):225 – 244, 2002.
- Somnath, V. R., Pariset, M., Hsieh, Y.-P., Martinez, M. R., Krause, A., and Bunne, C. Aligned diffusion schrödinger bridges, 2023. URL <https://arxiv.org/abs/2302.11419>.
- Soteras Gutiérrez, I., Lin, F.-Y., Vanommeslaeghe, K., Lemkul, J. A., Armacost, K. A., Brooks, C. L., and MacKerell, A. D. Parametrization of halogen bonds in the charmm general force field: Improved treatment of ligand–protein interactions. *Bioorganic & Medicinal Chemistry*, 24(20):4812–4825, October 2016. ISSN 0968-0896. doi: 10.1016/j.bmc.2016.06.034. URL <http://dx.doi.org/10.1016/j.bmc.2016.06.034>.
- Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport, 2023. URL <https://arxiv.org/abs/2302.00482>.
- Tooke, C. L., Hinchliffe, P., Bragginton, E. C., Colenso, C. K., Hirvonen, V. H., Takebayashi, Y., and Spencer, J. β -lactamases and β -lactamase inhibitors in the 21st century. *Journal of Molecular Biology*, 431(18):3472–3500, August 2019. ISSN 0022-2836. doi: 10.1016/j.jmb.2019.04.002. URL <http://dx.doi.org/10.1016/j.jmb.2019.04.002>.
- Torge, J., Harris, C., Mathis, S. V., and Lio, P. Diffhopp: A graph diffusion model for novel drug design via scaffold hopping, 2023. URL <https://arxiv.org/abs/2308.07416>.
- Trott, O. and Olson, A. J. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, June 2009. ISSN 1096-987X. doi: 10.1002/jcc.21334. URL <http://dx.doi.org/10.1002/jcc.21334>.
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L. M., Söding, J., and Steinegger, M. Fast and accurate protein structure search with foldseek. *Nature Biotechnology*, 42(2):243–246, May 2023. ISSN 1546-1696. doi: 10.1038/s41587-023-01773-0. URL <http://dx.doi.org/10.1038/s41587-023-01773-0>.
- Vanommeslaeghe, K. and MacKerell, A. D. Automation of the charmm general force field (cgenff) i: Bond perception and atom typing. *Journal of Chemical Information and Modeling*, 52(12):3144–3154, November 2012. ISSN 1549-960X. doi: 10.1021/ci300363c. URL <http://dx.doi.org/10.1021/ci300363c>.
- Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell, A. D. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):671–690, July 2009. ISSN 1096-987X. doi: 10.1002/jcc.21367. URL <http://dx.doi.org/10.1002/jcc.21367>.
- Vanommeslaeghe, K., Raman, E. P., and MacKerell, A. D. Automation of the charmm general force field (cgenff) ii: Assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling*, 52(12):3155–3168, November 2012. ISSN 1549-960X. doi: 10.1021/ci3003649. URL <http://dx.doi.org/10.1021/ci3003649>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need, 2017. URL <https://arxiv.org/abs/1706.03762>.
- Viguera Diez, J., Romeo Atance, S., Engkvist, O., and Olsson, S. Generation of conformational ensembles of small molecules via surrogate model-assisted molecular dynamics. November 2023. doi: 10.26434/chemrxiv-2023-sx61w. URL <http://dx.doi.org/10.26434/chemrxiv-2023-sx61w>.
- Vincent, P. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, July 2011. ISSN 1530-888X. doi: 10.1162/neco_a_00142. URL http://dx.doi.org/10.1162/NECO_a_00142.
- Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., Lauko, A., De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T. S., DiMaio, F., Baek, M., and Baker, D. 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>.
- Way, T.-D., Kao, M.-C., and Lin, J.-K. Degradation of her2/neu by apigenin induces apoptosis through cy-

- tochrome c release and caspase-3 activation in her2/neu-overexpressing breast cancer cells. *FEBS Letters*, 579(1):145–152, November 2004. ISSN 1873-3468. doi: 10.1016/j.febslet.2004.11.061. URL <http://dx.doi.org/10.1016/j.febslet.2004.11.061>.
- Weinmann, H. and Ottow, E. *Recent Development in Novel Anticancer Therapies*, pp. 221–251. Elsevier, 2007. ISBN 9780080450445. doi: 10.1016/b0-08-045044-x/00210-8. URL <http://dx.doi.org/10.1016/B0-08-045044-X/00210-8>.
- Wong, H. Y. and Wong, K.-B. *Using AlphaFold2 and Molecular Dynamics Simulation to Model Protein Recognition*, pp. 49–66. Springer US, 2024. ISBN 9781071640593. doi: 10.1007/978-1-0716-4059-3_4. URL http://dx.doi.org/10.1007/978-1-0716-4059-3_4.
- Wu, K. E., Yang, K. K., Berg, R. v. d., Zou, J. Y., Lu, A. X., and Amini, A. P. Protein structure generation via folding diffusion, 2022. URL <https://arxiv.org/abs/2209.15611>.
- Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A. B., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. Using evolutionary data to make sense of macromolecules with a “face-lifted” consurf. *Protein Science*, 32(3), February 2023. ISSN 1469-896X. doi: 10.1002/pro.4582. URL <http://dx.doi.org/10.1002/pro.4582>.
- Yim, J., Trippe, B. L., De Bortoli, V., Mathieu, E., Doucet, A., Barzilay, R., and Jaakkola, T. SE(3) diffusion model with application to protein backbone generation. *ICML*, 2023. doi: 10.48550/ARXIV.2302.02277. URL <https://arxiv.org/abs/2302.02277>.
- Yu, W., He, X., Vanommeslaeghe, K., and MacKerell, A. D. Extension of the charmm general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *Journal of Computational Chemistry*, 33(31):2451–2468, July 2012. ISSN 1096-987X. doi: 10.1002/jcc.23067. URL <http://dx.doi.org/10.1002/jcc.23067>.
- Zhang, Y. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic Acids Research*, 33(7): 2302–2309, April 2005. ISSN 1362-4962. doi: 10.1093/nar/gki524. URL <http://dx.doi.org/10.1093/nar/gki524>.
- Zhou, B., Zheng, L., Wu, B., Yi, K., Zhong, B., Tan, Y., Liu, Q., Liò, P., and Hong, L. A conditional protein diffusion model generates artificial programmable endonuclease sequences with enhanced activity. August 2023. doi: 10.1101/2023.08.10.552783. URL <http://dx.doi.org/10.1101/2023.08.10.552783>.
- Šali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, December 1993. ISSN 0022-2836. doi: 10.1006/jmbi.1993.1626. URL <http://dx.doi.org/10.1006/jmbi.1993.1626>.

A. Protein Backbone Generation

A.1. SE(3) Decomposition into SO(3) and \mathbb{R}^3

The Special Euclidean group SE(3) describes the rotations and translations in 3D space. An element of the SE(3) can be represented by a 4×4 matrix:

$$\mathbf{T} = \begin{pmatrix} \mathbf{R} & \mathbf{x} \\ \mathbf{0}_{1 \times 3} & 1 \end{pmatrix} \quad (2)$$

where \mathbf{R} is a 3×3 rotation matrix belonging to the Special Orthogonal group SO(3), and $\mathbf{x} = [x_x \ x_y \ x_z] \in \mathbb{R}^3$ is the translational component. Since SE(3) can be viewed as the semidirect product of SO(3) and \mathbb{R}^3 , denoted as $\text{SE}(3) \cong \text{SO}(3) \ltimes \mathbb{R}^3$, one option is to naturally treat SO(3) and \mathbb{R}^3 as independent for simplicity (Yim et al., 2023).

A.2. Protein Backbone Representations

Molecules can be intuitively represented as 3D atomic point clouds. However, macromolecules like proteins may contain thousands or tens of thousands of atoms, with variation in the atom types and quantities among different amino acids (for instance, sulfur atoms are present only in a few amino acids like cysteine). Representing proteins as unordered 3D atomic point clouds significantly increases data dimensionality and sparsity, requiring far more training data than is typically available.

Following the work of Yim et al. (2023) and Bose et al. (2023), we adopt the more compact backbone rigid groups from AlphaFold (Jumper et al., 2021) to represent protein backbone structures in 3D space. A backbone rigid group consists of the main chain atoms (N, C_α , C, O) within a single residue (Figure 1A), where their geometric relationships (relative positions and orientations) are highly stable. The position and orientation of the group is transformed as a whole, without accounting for individual atomic movements, simplifying the computation and reducing structural errors caused by excessive degrees of freedom.

Assuming experimentally measured ideal chemical bond angles and lengths, models learn how the rigid transformation (or *frame*) \mathbf{T}_i of each residue $i \in [1, N]$ acts on idealized coordinates $[\mathbf{N}^*, C_\alpha^*, C^*] \in \mathbb{R}^3$ (centered at C_α^*), so that the transformed coordinates match the actual coordinates as closely as possible:

$$[\mathbf{N}, C_\alpha, C]_i = \mathbf{T}_i \cdot [\mathbf{N}^*, C_\alpha^*, C^*] \quad (3)$$

where $\mathbf{T}_i \in \text{SE}(3)$ can be decomposed into a rotation matrix $\mathbf{R}_i \in \text{SO}(3)$ and a translation vector $\mathbf{x}_i \in \mathbb{R}^3$. An additional torsion angle $\psi_i \in \text{SO}(2)$ is introduced between the bond of C_α and C for a more accurate construction of the backbone oxygen atom O.

A.3. SE(3) Score Matching

Let $\mathbf{T}_t = [\mathbf{T}_{1,t}, \dots, \mathbf{T}_{N,t}] \in \text{SE}(3)^N$ denote the manifold of N frames at time t , where each frame can independently rotate and translate. Correspondingly, define $\mathbf{R}_t = [\mathbf{R}_{1,t}, \dots, \mathbf{R}_{N,t}]$ and $\mathbf{X}_t = [\mathbf{x}_{1,t}, \dots, \mathbf{x}_{N,t}]$. By treating SO(3) and \mathbb{R}^3 as two independent stochastic processes, a *forward process* gradually perturbs the initial data distribution p_0 . Following the approach of Yim et al. (2023), this process is described by the Stochastic Differential Equation (SDE) for $\mathbf{T}_t \sim p_t$ and any arbitrary time $t \in [0, T]$:

$$d\mathbf{T}_t = \left[0, -\frac{1}{2}\mathbf{X}_t \right] dt + \left[d\mathbf{B}_t^{\text{SO}(3)}, d\mathbf{B}_t^{\mathbb{R}^3} \right] \quad (4)$$

where $\mathbf{B}_t^{\text{SO}(3)}$ and $\mathbf{B}_t^{\mathbb{R}^3}$ are Brownian motions on the SO(3) and \mathbb{R}^3 , respectively. Invariant density $p_{\text{inv}}^{\text{SE}(3)}(\mathbf{T}) \propto \mathcal{U}^{\text{SO}(3)}(\mathbf{R}) \mathcal{N}(\mathbf{x}; 0, \mathbf{I})$ is chosen for $\mathbf{T} = (\mathbf{R}, \mathbf{x})$.

Let $(\overleftarrow{\mathbf{T}}_t)_{t \in [0, T]} = (\mathbf{T}_{T-t})_{t \in [0, T]}$, the corresponding *time-reverse process* (De Bortoli et al., 2022) is given by

$$d\overleftarrow{\mathbf{R}}_t = \nabla_{\mathbf{R}} \log p_{T-t}(\overleftarrow{\mathbf{T}}_t) dt + d\mathbf{B}_t^{\text{SO}(3)} \quad (5)$$

$$d\overleftarrow{\mathbf{X}}_t = \left\{ \frac{\overleftarrow{\mathbf{X}}_t}{2} + \nabla_{\mathbf{x}} \log p_{T-t}(\overleftarrow{\mathbf{T}}_t) \right\} dt + d\mathbf{B}_t^{\mathbb{R}^3} \quad (6)$$

where $\nabla \log p$ is the gradient of the log-probability density function (also known as the Stein *score*). However, this gradient is typically intractable in practice because the exact form of $p_t(\mathbf{T}_t)$ is unknown at any given time t .

Instead, score-based models estimates tractable conditional score $\nabla \log p_{t|0}$ through SM (Vincent, 2011), using a neural network $s(\theta, t, \cdot)$ trained by minimizing both⁴:

$$\mathcal{L}_{\text{SM}}^{\mathbf{R}}(\theta) = \mathbb{E} [\|\nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}_t|\mathbf{R}_0) - s(\theta, t, \mathbf{R}_t)\|^2] \quad (7)$$

$$\mathcal{L}_{\text{SM}}^{\mathbf{X}}(\theta) = \mathbb{E} [\|\nabla_{\mathbf{X}} \log p_{t|0}(\mathbf{X}_t|\mathbf{X}_0) - s(\theta, t, \mathbf{X}_t)\|^2] \quad (8)$$

with $t \sim \mathcal{U}(0, T)$ and

$$\nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}_t|\mathbf{R}_0) = \frac{\mathbf{R}_t}{\omega(\mathbf{R}_{0 \rightarrow t})} \log\{\mathbf{R}_{0 \rightarrow t}\} \frac{\partial_{\omega} f(\omega(\mathbf{R}_{0 \rightarrow t}), t)}{f(\omega(\mathbf{R}_{0 \rightarrow t}), t)} \quad (9)$$

$$\nabla_{\mathbf{x}} \log p_{t|0}(\mathbf{x}_t|\mathbf{x}_0) = \frac{e^{-t/2} \mathbf{x}_0 - \mathbf{x}_t}{1 - e^{-t}} \quad (10)$$

where ω represents the rotation angle, $\mathbf{R}_{0 \rightarrow t} = \mathbf{R}_0^{\top} \mathbf{R}_t$, and

$$f(\omega, t) = \sum_{\ell \in \mathbb{N}} (2\ell + 1) e^{-\ell(\ell+1)t/2} \frac{\sin((\ell + \frac{1}{2})\omega)}{\sin(\frac{\omega}{2})} \quad (11)$$

is an auxiliary function for the heat kernel⁵ of the Brownian motion on $\text{SO}(3)$.

A.4. SE(3) Flow Matching

FM is a simulation-free method for training vector fields to follow a prescribed conditional probability path (Lipman et al., 2022). Formally, for $t \in [0, 1]$, let $\mathbf{U} = \{\mathbf{u}_t\}$ be a *flow* which is a set of time-indexed vector fields that describe the paths along which data points move from an initial distribution p_1 to a target distribution p_0 . Each vector field $\mathbf{u}_t(\mathbf{T}_t)$ represents the rate of change of \mathbf{T}_t which is typically the solution to the Ordinary Differential Equation (ODE) $\frac{d}{dt} \mathbf{T}_t = \mathbf{u}_t(\mathbf{T}_t)$. FM approximates $\mathbf{u}_t(\mathbf{T}_t)$ with a network $v(\theta, t, \cdot)$ by minimizing $\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E} \|\mathbf{u}_t(\mathbf{T}_t) - v(\theta, t, \mathbf{T}_t)\|^2$ with $t \sim \mathcal{U}(0, 1)$.

Similarly, independent flows can be built on $\text{SO}(3)$ and \mathbb{R}^3 . Computing \mathbf{u}_t , however, is also intractable due to the complex integrals involved in defining the marginal probability path and vector field. By showing $\nabla_{\theta} \mathcal{L}_{\text{FM}}(\theta) = \nabla_{\theta} \mathcal{L}_{\text{CFM}}(\theta)$, Lipman et al. (2022) suggested the tractable conditional FM objective on \mathbb{R}^3 :

$$\mathcal{L}_{\text{CFM}}^{\mathbf{X}}(\theta) = \mathbb{E} \|\mathbf{u}_t(\mathbf{X}_t|\mathbf{X}_0) - v(\theta, t, \mathbf{X}_t)\|^2 \quad (12)$$

with the Gaussian path $p_t(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; t\mathbf{x}_0, (t\sigma - t + 1)^2)$ generated by

$$\mathbf{u}_t(\mathbf{x}_t|\mathbf{x}_0) = \frac{\mathbf{x}_0 - (1 - \sigma)\mathbf{x}_t}{1 - (1 - \sigma)t} \quad (13)$$

where $\sigma > 0$ is a smoothing constant.

For flows on $\text{SO}(3)$, Bose et al. (2023) set

$$\mathcal{L}_{\text{CFM}}^{\mathbf{R}}(\theta) = \mathbb{E} \|\mathbf{u}_t(\mathbf{R}_t|\mathbf{R}_0, \mathbf{R}_1) - \mathbf{v}(\theta, t, \mathbf{R}_t)\|^2 \quad (14)$$

and define the geodesic interpolant between $\mathbf{R}_1 \sim p_1$ and $\mathbf{R}_0 \sim p_0$ as $\mathbf{R}_t = \exp_{\mathbf{R}_1}(t \log_{\mathbf{R}_1}(\mathbf{R}_0))$. Let Ψ_t be a flow that connects \mathbf{R}_1 to \mathbf{R}_0 , computing $\mathbf{u}_t(\mathbf{R}_t|\mathbf{R}_0, \mathbf{R}_1)$ simplifies to determining \mathbf{R}_t along $\frac{d}{dt} \Psi_t(\mathbf{R}) = \dot{\mathbf{R}}_t$ (Chen & Lipman, 2023) and then taking its time derivative. Thus, we have

$$\mathbf{u}_t(\mathbf{R}_t|\mathbf{R}_0, \mathbf{R}_1) = \frac{\log_{\mathbf{R}_t}(\mathbf{R}_0)}{t} \quad (15)$$

⁴One adds weights $1/\mathbb{E}[\|\nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}_t|\mathbf{R}_0)\|^2]$ to Equation (7) and $(1 - e^{-t})/e^{-t/2}$ to Equation (8) for simplicity.

⁵The heat kernel on a manifold is the fundamental solution to the heat equation, representing the probability density function of a Brownian particle diffusing from one point to another over time.

Optimal transport. Optimal Transport (OT) conditions hold when the probability paths between two distributions are defined by a displacement map that linearly interpolates between them (Pooladian et al., 2023).

Tong et al. (2023) views the OT problem as finding a mapping that minimizes the 2-Wasserstein distance between two distributions p_1 and p_0 on \mathbb{R}^3 , using the Euclidean distance $\|\mathbf{x}_0 - \mathbf{x}_1\|$ as the displacement cost:

$$W(p_0, p_1)_2^2 = \inf_{\pi \in \Pi} \int_{\mathbb{R}^3 \times \mathbb{R}^3} \|\mathbf{x}_0 - \mathbf{x}_1\|^2 d\pi(\mathbf{x}_0, \mathbf{x}_1) \quad (16)$$

where Π denotes the set of all joint probability measures on $\mathbb{R}^3 \times \mathbb{R}^3$ with marginals p_1 and p_0 . By setting $p(\mathbf{x}_0, \mathbf{x}_1) = \pi(\mathbf{x}_0, \mathbf{x}_1)$ and a Gaussian conditional probability path with mean $\mu_t = t\mathbf{x}_0 + (1-t)\mathbf{x}_1$, we have

$$\mathcal{L}_{\text{OT}}^{\mathbf{X}}(\theta) = \mathbb{E}_{\pi} \|\mathbf{u}_t(\mathbf{X}_t | \mathbf{X}_0, \mathbf{X}_1) - v(\theta, t, \mathbf{X}_t)\|^2 \quad (17)$$

$$\mathbf{u}_t(\mathbf{x}_t | \mathbf{x}_0, \mathbf{x}_1) = \mathbf{x}_0 - \mathbf{x}_1 \quad (18)$$

with $p_t(\mathbf{x}_t) = \int \mathcal{N}(\mathbf{x}_t | t\mathbf{x}_0 + (1-t)\mathbf{x}_1, \sigma^2) \pi(\mathbf{x}_0, \mathbf{x}_1) d\mathbf{x}_0 d\mathbf{x}_1$.

Inspired by this, Bose et al. (2023) extended Equation (14) and Equation (15) to SO(3) using Riemannian optimal transport, with $\bar{\pi}$ being the projection of π on SO(3):

$$\mathcal{L}_{\text{OT}}^{\mathbf{R}}(\theta) = \mathbb{E}_{\bar{\pi}} \left\| \frac{\log_{\mathbf{R}_t}(\mathbf{R}_0)}{t} - v(\theta, t, \mathbf{R}_t) \right\|^2 \quad (19)$$

A.5. SE(3) Invariance

SE(3) invariance can be achieved by consistently positioning the model at the origin (Yim et al., 2023; Bose et al., 2023; Rudolph et al., 2020).

In the context of SM, to ensure translation invariance on \mathbb{R}^3 , one apply a projection matrix $\mathbf{P} \in \mathbb{R}^{3N \times 3N}$ that removes the center of mass $\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$. It results in an invariant measure on $\text{SE}(3)^N$, denoted as $\text{SE}(3)_0^N$. Since the Brownian motion on SO(3) and the score $\nabla_{\mathbf{R}} \log p_{T-t}$ are both rotation-invariant, Equation (5) is SO(3)-invariant. Consequently, Yim et al. (2023) derive the following SE(3)-invariant forward process

$$d\mathbf{T}_t = \left[0, -\frac{1}{2} \mathbf{P} \mathbf{X}_t \right] dt + \left[d\mathbf{B}_t^{\text{SO}(3)^N}, \mathbf{P} d\mathbf{B}_t^{\mathbb{R}^{3N}} \right] \quad (20)$$

and its corresponding time-reverse process

$$d\hat{\mathbf{R}}_t = \nabla_{\mathbf{R}} \log p_{T-t}(\hat{\mathbf{T}}_t) dt + d\mathbf{B}_t^{\text{SO}(3)^N} \quad (21)$$

$$d\hat{\mathbf{X}}_t = \mathbf{P} \left\{ \frac{\hat{\mathbf{X}}_t}{2} + \nabla_{\mathbf{x}} \log p_{T-t}(\hat{\mathbf{T}}_t) \right\} dt + d\mathbf{P} \mathbf{B}_t^{\mathbb{R}^{3N}} \quad (22)$$

The same approach can be applied to FM. After centering and decoupling the flow on $\text{SE}(3)_0^N$, a separate SE(3)-invariant flow can be constructed for each residue in backbone⁶, in which each SE(3)-invariant measure is decomposed into a measure that is proportional to the Lebesgue measure on \mathbb{R}^3 (Pollard, 2001) and an SO(3)-invariant measure (Bose et al., 2023).

A.6. Additional Losses

To prevent unrealistic fine-grained features such as steric clashes or chain breaks when learning the torsion angle ψ , Yim et al. (2023) proposed adding two additional loss functions. The first is the mean squared error (MSE) on backbone atom positions:

$$\mathcal{L}_{\text{bb}} = \frac{1}{4N} \sum_{n=1}^N \sum_{a \in A} \|a_n - \hat{a}_n\|^2 \quad (23)$$

⁶As the product group of N copies of SE(3), $\text{SE}(3)_0^N$ has a geometric structure that allows global geometric operations (such as geodesic distance, exponential maps, and logarithmic maps) to be decomposed into operations on each of the N SE(3) groups.

where $A = \{N, C, C\alpha, O\}$. a_n and \hat{a}_n are the true and predicted coordinates of backbone atom a at residue n .

The second loss penalizes deviations in local pairwise atomic distances:

$$\mathcal{L}_{2D} = \frac{\sum_{n=1}^N \sum_{m=1}^N \sum_{a,b \in A} \mathbb{1}\{d_{ab}^{nm} < 6\text{\AA}\} \|d_{ab}^{nm} - \hat{d}_{ab}^{nm}\|^2}{\left(\sum_{n=1}^N \sum_{m=1}^N \sum_{a,b \in A} \mathbb{1}\{d_{ab}^{nm} < 6\text{\AA}\}\right) - N} \quad (24)$$

where $d_{ab}^{nm} = \|a_n - b_m\|$ and \hat{d}_{ab}^{nm} are the true and predicted distances between atoms a and b in residues n and m , respectively. The indicator function $\mathbb{1}\{d_{ab}^{nm} < 6\text{\AA}\}$ limits the loss to atom pairs within 6 Å.

The complete training loss is given by

$$\mathcal{L} = \mathcal{L}^{\mathbf{R}}(\theta) + \mathcal{L}^{\mathbf{X}}(\theta) + \mathbb{1}\{t < T/4\} (\mathcal{L}_{\text{bb}} + \mathcal{L}_{2D}) \quad (25)$$

where $T = 1$ in the case of FM.

A.7. Model Architecture

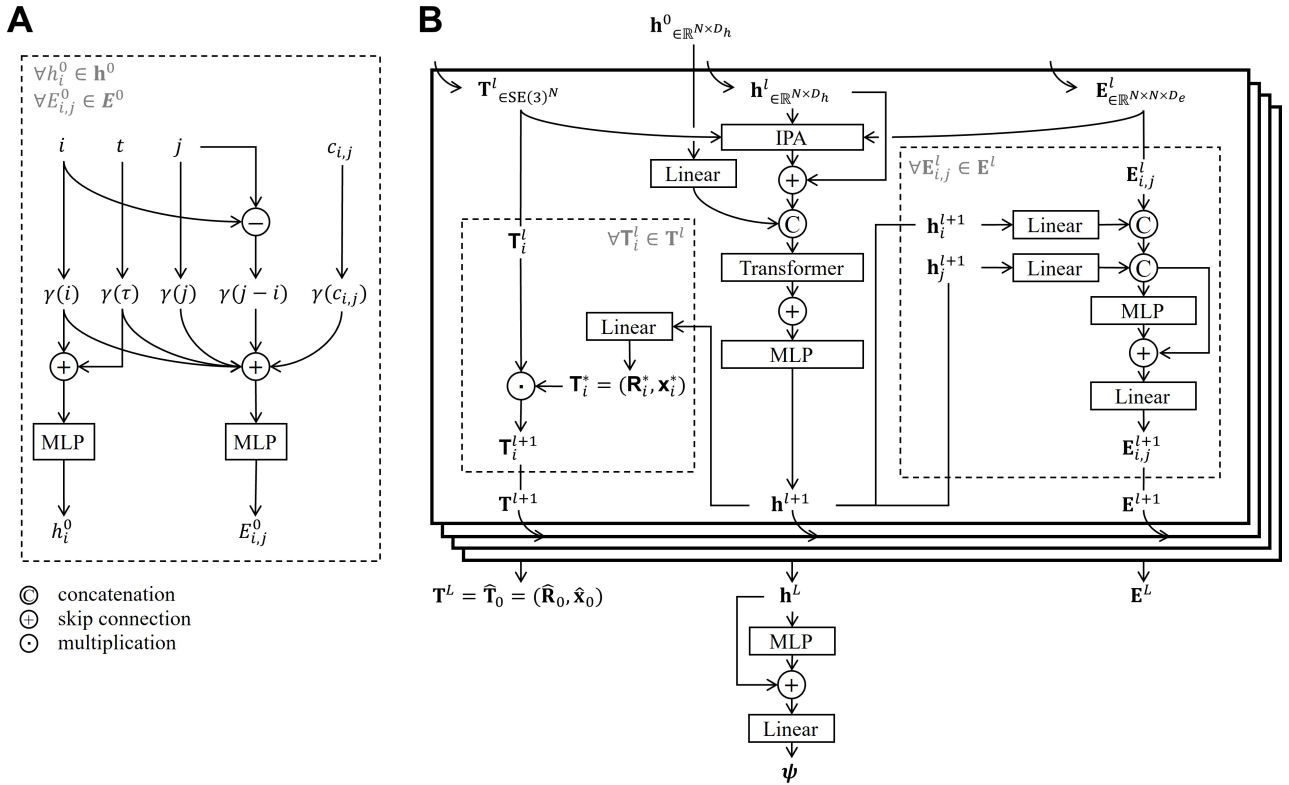


Figure 5. Overview of the (A) embedding module and (B) multi-layer network architecture.

The networks $s(\theta, t, \cdot)$ involved in SM and $v(\theta, t, \cdot)$ involved in FM models, as reviewed in Section 2, can share a common high-level architecture (Yim et al., 2023; Bose et al., 2023; Anand & Achim, 2022).

Embeddings. Given node embedding dimensions D_h and edge embedding dimensions D_e , node embeddings $\mathbf{h} \in \mathbb{R}^{N \times D_h}$ are derived from residue indices $\mathbf{i} = \{1, \dots, N\}$ and time-step information $\mathbf{t} = \{0, \Delta t, \dots, T\}$, while edge embeddings $\mathbf{E} \in \mathbb{R}^{N \times N \times D_e}$ integrate additional features, such as relative sequence distances $j - i$ for any $i, j \in [1, N]$ (Figure 5A). Self-conditioning on the predicted C_α displacements is also applied:

$$c_{i,j} = \sum_{b=1}^B \mathbb{1}\{|\mathbf{x}_i^* - \mathbf{x}_j^*| < v_b\} \quad (26)$$

where \mathbf{x}^* denotes the coordinates for C_α predicted through self-conditioning, and v_1, \dots, v_B are bins spaced uniformly from 0 and B angstroms. These initial features are encoded using multilayer perceptrons (MLPs) along with sinusoidal embeddings (Vaswani et al., 2017).

Multi-layer network. Figure 5B shows the architecture of the multi-layer neural network ($L = 4$ layers used in our experiments). At each layer l , the network takes node embeddings \mathbf{h}^l , edge embeddings \mathbf{E}^l , and rigid transformations \mathbf{T}^l as input, applying the Invariant Point Attention (IPA) introduced by Jumper et al. (2021) to enable spatial attention. Transformer from Vaswani et al. (2017) models interactions along the chain structure. The network’s update procedure remains invariant under $SE(3)$ transformations due to the inherent $SE(3)$ -invariance of the IPA.

The output \mathbf{T}^L from the final layer serves as the predicted frame, denoted as $\hat{\mathbf{T}}_0 = (\hat{\mathbf{R}}_0, \hat{\mathbf{x}}_0)$. Consequently, for SM, we have the following scores predictions based on Equation (9) and Equation (10):

$$\forall s(\theta, t, \mathbf{R}_t) \in s(\theta, t, \mathbf{R}_t), \quad s(\theta, t, \mathbf{R}_t) = \nabla_{\mathbf{R}} \log p_{t|0}(\mathbf{R}_t | \hat{\mathbf{R}}_0) \quad (27)$$

$$= \frac{\mathbf{R}_t}{\omega(\hat{\mathbf{R}}_0^\top \mathbf{R}_t)} \log \{ \hat{\mathbf{R}}_0^\top \mathbf{R}_t \} \frac{\partial_\omega f(\omega(\mathbf{R}_0^\top \mathbf{R}_t), t)}{f(\omega(\mathbf{R}_0^\top \mathbf{R}_t), t)} \quad (28)$$

$$\forall s(\theta, t, \mathbf{x}_t) \in s(\theta, t, \mathbf{X}_t), \quad s(\theta, t, \mathbf{x}_t) = \nabla_{\mathbf{x}} \log p_{t|0}(\mathbf{x}_t | \hat{\mathbf{x}}_0) \quad (29)$$

$$= \frac{e^{-t/2} \hat{\mathbf{x}}_0 - \mathbf{x}_t}{1 - e^{-t}} \quad (30)$$

From Equation (13) and Equation (15), we have the following for FM with OT:

$$\forall v(\theta, t, \mathbf{R}_t) \in v(\theta, t, \mathbf{R}_t), \quad v(\theta, t, \mathbf{R}_t) = \mathbf{u}_t(\mathbf{R}_t | \hat{\mathbf{R}}_0, \mathbf{R}_1) \quad (31)$$

$$= \frac{\log_{\mathbf{R}_t}(\hat{\mathbf{R}}_0)}{t} \quad (32)$$

$$\forall v(\theta, t, \mathbf{x}_t) \in v(\theta, t, \mathbf{X}_t), \quad v(\theta, t, \mathbf{x}_t) = \mathbf{u}_t(\mathbf{x}_t | \hat{\mathbf{x}}_0) \quad (33)$$

$$= \frac{\hat{\mathbf{x}}_0 - (1 - \sigma)\mathbf{x}_t}{1 - (1 - \sigma)t} \quad (34)$$

Torsion angle $\hat{\psi} = \{\hat{\psi}_1, \dots, \hat{\psi}_N\} = \psi / \|\psi\| \in SO(2)^N$ is predicted with \mathbf{h}^L and \mathbf{E}^L .

B. Amino Acid Count Distributions in Experimental Proteins Used for Training

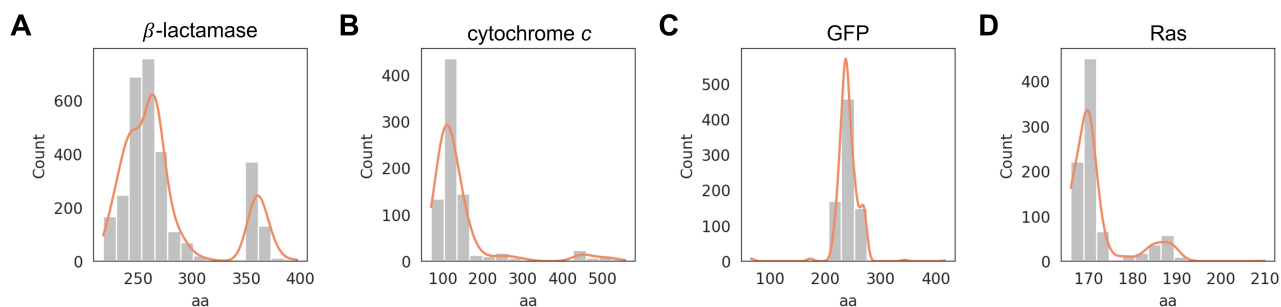


Figure 6. (A-D) Distribution of amino acid sequence lengths (aa) in experimentally determined protein structures used for training.

C. Structural Phylogenetic Trees

Figure 7 shows structural phylogenetic trees generated using Q_{score} and the 3Di alphabet. Both approaches reveal evolutionary relationships between the generated and experimental structures. However, compared to the Q_{score} -based tree, the 3Di alphabet tree shows closer and less distinguishable evolutionary relationships among samples from different protein families, especially between the GFP and cytochrome c samples.

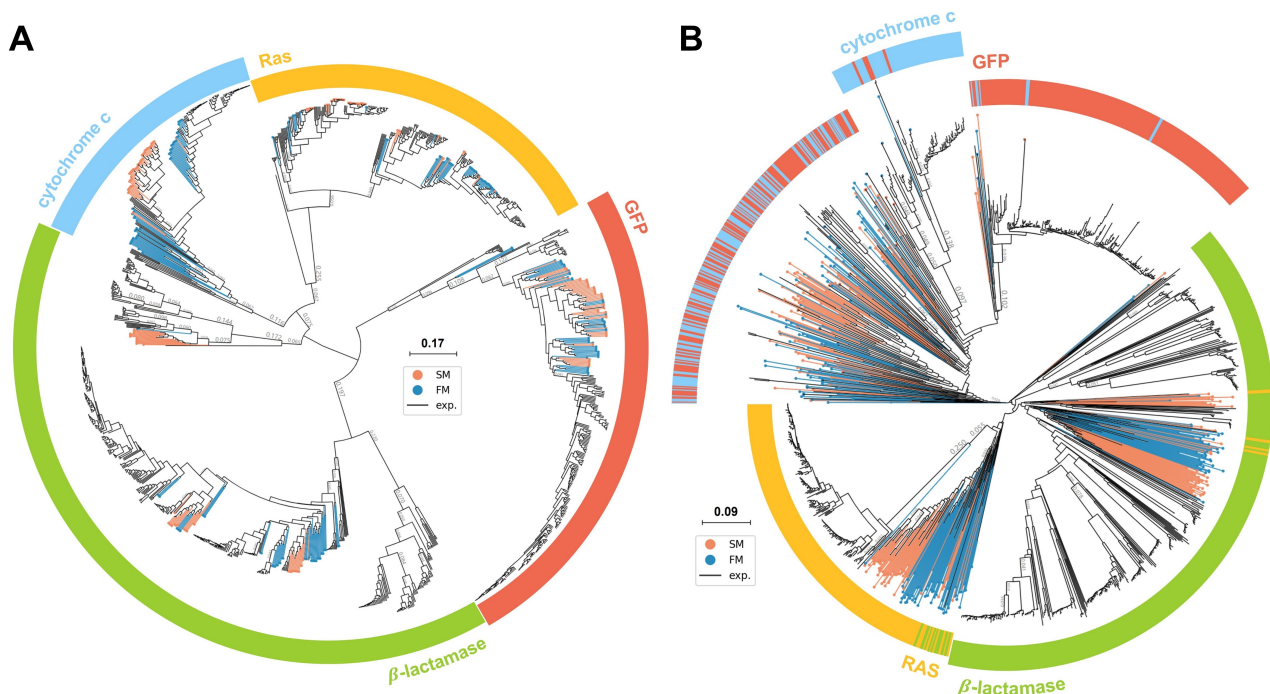


Figure 7. Structural phylogenetic trees constructed using (A) Q_{score} and (B) 3Di alphabet. Members of different protein families are grouped into semi-circular rings. The colored branches and nodes represent the generated protein structures.

D. Conserved Residue Consistency

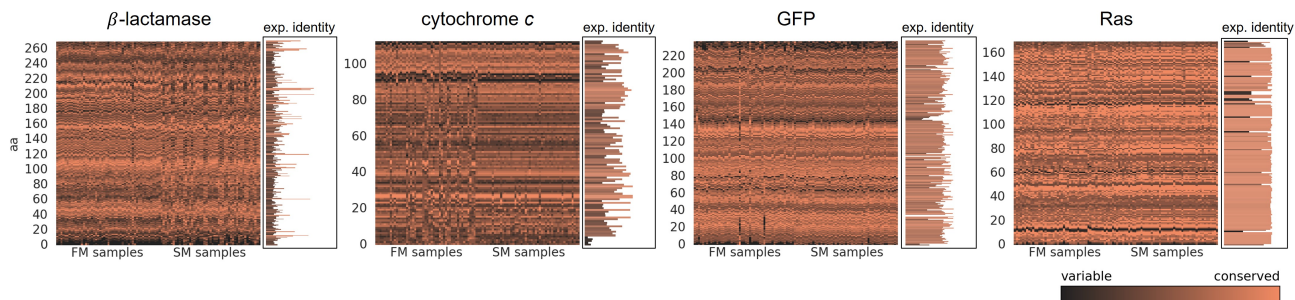


Figure 8. Normalized evolutionary rates were mapped onto sequences, with black indicating rapidly evolving positions and orange indicating conserved positions. The heat map displays each column as the optimal predicted sequence of a generated sample, with rows representing specific positions in these sequences. The bar chart (exp. identity) shows the average pairwise sequence identity at each position from the MSAs of experimental proteins. Taller bars signify positions where most experimental protein share the same residue. Consistency between generated and experimental sequences at conserved positions appears as uniform orange coloring in the heat map, aligning with the taller bars.

E. Evaluation of Side-Chain Homology Modeling

We used homology modeling to add side-chains to the generated protein backbones, evaluating them using PROCHECK (Laskowski et al., 1993; 1996) and WHAT_CHECK (Hooft et al., 1996) to correct or exclude those not meeting the expectations. Specifically:

Planarity. Planar side-chains, such as those in phenylalanine, tyrosine, tryptophan, and histidine, are essential for stability and function. Conformations lacking expected planarity were discarded.

Asparagine, glutamine, histidine flips. Asparagine, glutamine, and histidine side-chains can experience terminal flips, altering key interactions. WHAT_CHECK was used to evaluate and, if necessary, adjust side-chain orientations to have more stable interactions.

Torsion angles. Side-chain torsion angles (χ angles) were assessed, focusing on χ_1 (rotation around C_α to the first side-chain atom) and χ_2 (rotation to the second side-chain atom) to prevent spatial clashes. Conformations in uncommon χ -angle regions were excluded.

Bond lengths and angles. Unusual bond lengths and angles may indicate strain and modeling errors, potentially disrupting interactions. Conformations with such issues were discarded.

Other parameters. Side-chains with abnormal torsion angles, atypical aromatic bonding angles, or unusual proline puckering were also discarded.

F. Molecular Dynamics Simulations

For experimental structures, crystallographic water and unnecessary small molecules were removed. For generated structures, missing side-chains were added via homology modeling (Section 3.6). After adding hydrogen atoms using Reduce2 (Grosse-Kunstleve et al., 2002) and confirming no missing atoms, each protein was centered at the origin.

Simulations were performed with GROMACS (Abraham et al., 2015), using the all-atom CHARMM36 force field (July 2022 version) (Vanommeslaeghe et al., 2009; Vanommeslaeghe & MacKerell, 2012; Vanommeslaeghe et al., 2012; Yu et al., 2012; Soteras Gutiérrez et al., 2016).

F.1. Molecular Dynamics Setup for Stability Assessment of Generated Structures

Proteins were placed in an octahedral simulation box with a minimum distance of 1.5 nm between the protein and the box boundaries. Prior to solvation, energy minimization was performed in vacuum using the steepest descent method (max 30,000 steps, step size 0.01 nm, convergence 2 kJ/(mol-nm)) to resolve steric clashes and geometric inconsistencies. Neighbor searching used a grid-based method with a search radius of 1.2 nm.

In accordance with GROMACS 2024 documentation, we applied the following configurations in the MD parameter (.mdp) files. Van der Waals interactions were handled using a cutoff method, while long-range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method.

```
constraints      = h-bonds
cutoff-scheme    = Verlet
vdwtype         = cutoff
vdw-modifier     = force-switch
rlist           = 1.2
rvdw           = 1.2
rvdw-switch     = 1.0
coulombtype     = PME
rcoulomb        = 1.2
DispCorr        = no
```

After solvating the system with water using the TIP3P model, we added Na^+ and Cl^- ions to achieve a physiological concentration of 150 mM and to neutralize the system's total charge. Energy minimization was then conducted to resolve steric clashes and optimize the geometry, with potential energy and maximum force monitored to ensure they reached acceptable thresholds.

The next step involved equilibrating the solvent and ions around the protein. We chose the leap-frog integrator for the simulations and applied the LINCS algorithm to constrain hydrogen bonds. Equilibration involved two stages. In the first stage, we performed a 500 ps NVT equilibration (250,000 steps with a 2 fs time step). Temperature control was managed using the V-rescale thermostat, with the system divided into two groups: (1) protein and (2) water + ions, both set to a target temperature of 310 K to simulate physiological conditions. In the second stage, we carried out a 500 ps NPT equilibration with pressure coupling enabled. The pressure was regulated using the C-rescale method with isotropic coupling. The target pressure was 1.0 bar, with a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$ and a pressure coupling time constant of 0.5 ps.

Following equilibration, we conducted a 10 ns production simulation (5,000,000 steps with a 2 fs time step), during which all position restraints were removed. This allowed us to observe and analyze the system's dynamic behavior over time, in order to access its stability. Full details of the MD parameter files can be found in Software and Data.

F.2. Molecular Dynamics Setup for Conformational Analysis of Protein-Ligand Complexes

The receptor and ligand were saved as separate coordinate files to prepare their respective topologies. The receptor topology was prepared as in Appendix F.1. For the ligand, hydrogen atoms were added using OpenBabel (O'Boyle et al., 2011), and topology was generated via the CGenFF server (Vanommeslaeghe et al., 2009; Vanommeslaeghe & MacKerell, 2012; Vanommeslaeghe et al., 2012). The receptor and ligand topologies, along with force-field-compatible coordinate files, were then combined to construct the complete complex system.

The MD workflow for complexes followed the same steps in Appendix F.1. Complexes were placed in an octahedral simulation box, energy-minimized in vacuum, solvated in water, and neutralized with Na⁺ and Cl⁻ ions to 150 mM. A second energy minimization was then performed on the solvated system.

During equilibration, positional restraints were applied to the ligand to prevent unnecessary displacement in the initial stages of the simulation. Additionally, to minimize interference from temperature fluctuations of the ligand on the overall simulation, we defined two temperature coupling groups: (1) the receptor and ligand as one group, and (2) the solvent and ions as the other. Other equilibration settings followed Appendix F.1.

After equilibration, restraints were removed, and a 10 ns production simulation was conducted to analyze the dynamic behavior and conformational changes in the complexes.

G. Protein-ligand Blind Docking

Similarly, crystallographic water and unwanted molecules were removed from experimental structures, and missing side-chains were added to generated structures via homology modeling. Receptor structures were prepared using AutoDock Tools (Morris et al., 2009), with polar hydrogens added, Kollman charges assigned, and any missing atoms repaired. For receptors within the same family, we prepared a shared ligand file, adding hydrogen atoms and assigning Gasteiger charges. A large grid box, typically 80 to 110 Å per side, was defined to cover the entire protein surface.

Using these settings, we performed blind docking with AutoDock Vina (Trott & Olson, 2009; Eberhardt et al., 2021), generating up to 25 binding modes with a maximum energy difference of 5 kcal/mol and an exhaustiveness level of 20. The binding mode with the lowest binding free energy was selected as the final result.

H. Generated Structures

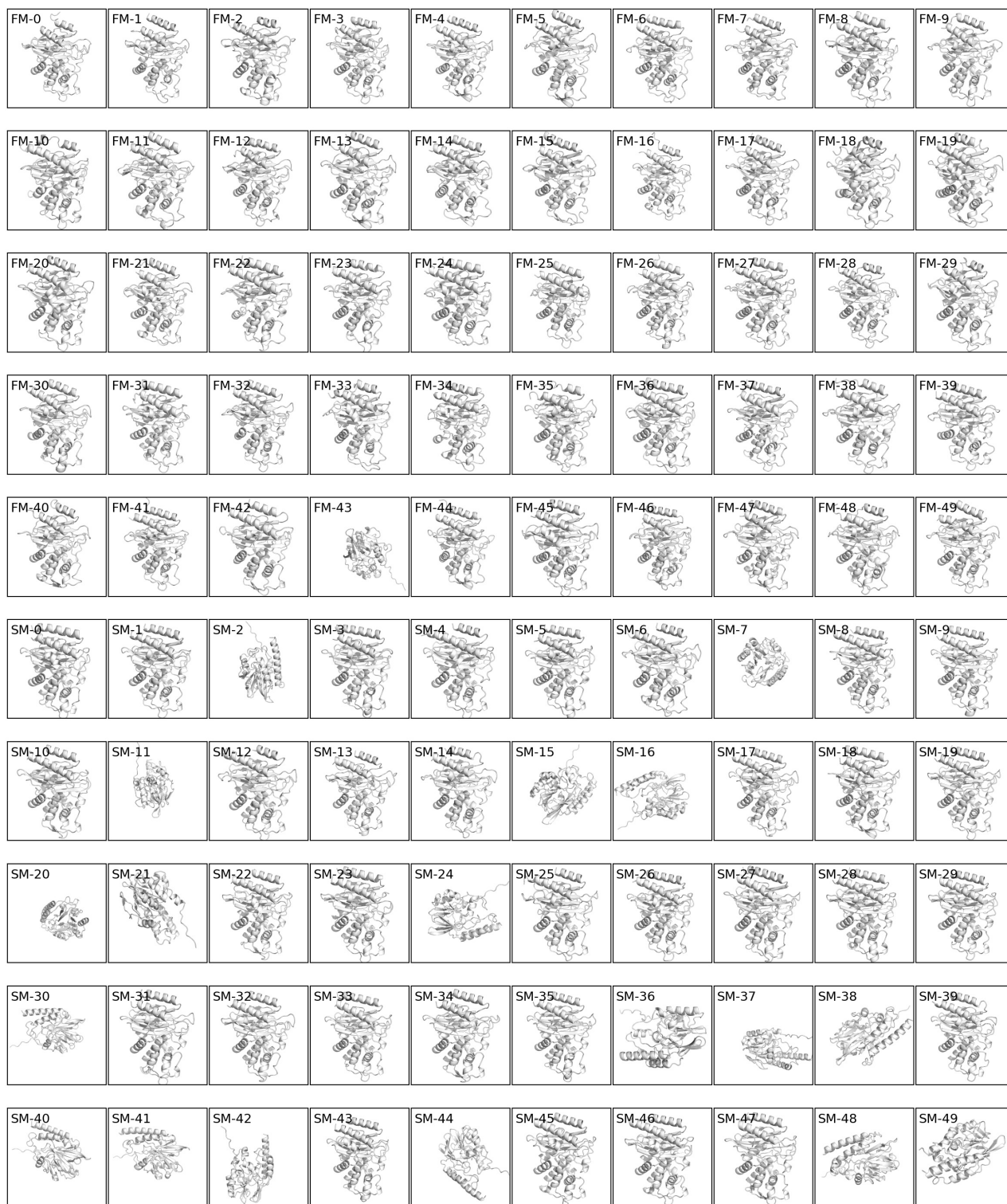


Figure 9. 50 β -lactamase-like protein backbones generated using score matching and 50 using flow matching.



Figure 10. 50 cytochrome *c*-like protein backbones generated using score matching and 50 using flow matching.

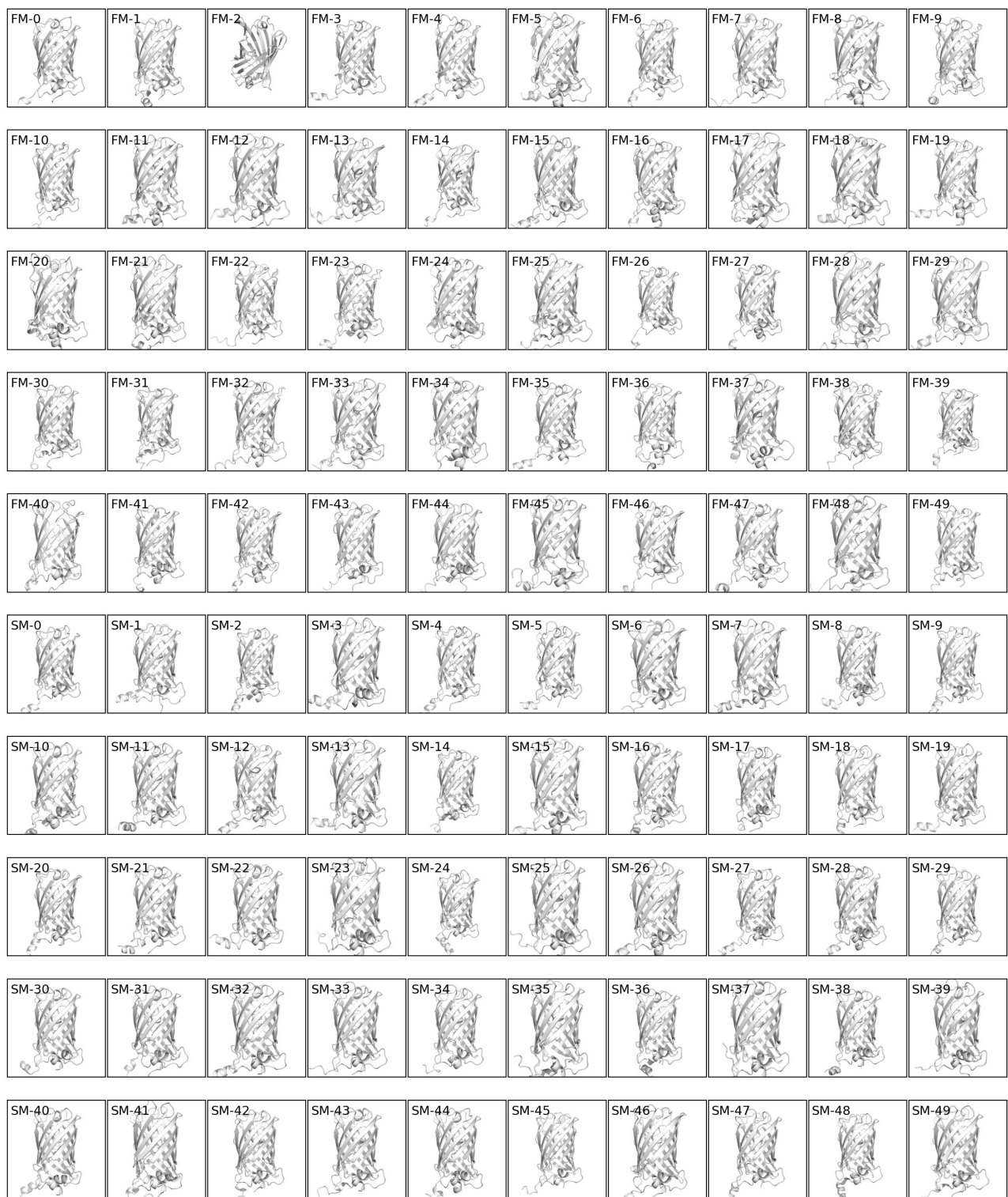


Figure 11. 50 GDP-like protein backbones generated using score matching and 50 using flow matching.

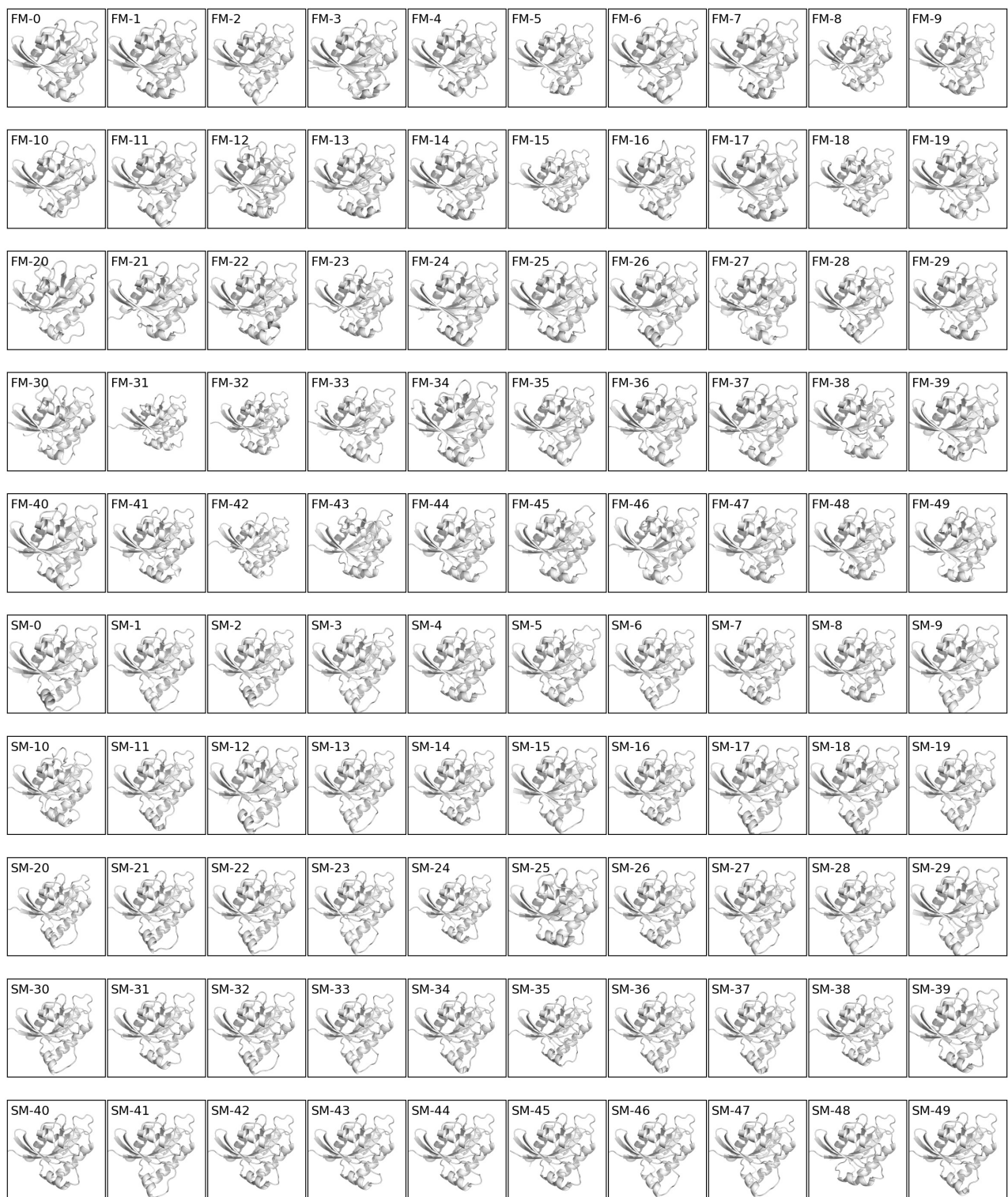


Figure 12. 50 Ras-like protein backbones generated using score matching and 50 using flow matching.