

PROTEIN-SE(3): UNIFIED FRAMEWORK AND COMPREHENSIVE BENCHMARK FOR SE(3)-BASED PROTEIN STRUCTURE DESIGN

Anonymous authors

Paper under double-blind review

ABSTRACT

SE(3)-based generative models have shown great promise in protein geometry modeling and effective structure design. However, the field currently lacks a pipeline to support consistent re-training and fair comparison across different methods. In this paper, we propose Protein-SE(3), a unified framework accompanied by the comprehensive benchmark for SE(3)-based protein design. Protein-SE(3) integrates recent advanced methods, supports diverse evaluation metrics and also develops a mathematical decoupling toolkit. Specifically, recent advanced generative models designed for typical protein design tasks (unconditional generation and motif scaffolding), from multiple perspectives like DDPM (Genie1 and Genie2), Score Matching (FrameDiff and RfDiffusion) and Flow Matching (FoldFlow and FrameFlow) are systematically incorporated into our framework. All methods are re-trained on identical datasets and evaluated with consistent metrics, ensuring fair and reproducible comparison. Furthermore, the proposed decoupling toolkit abstracts the mathematical foundations of generative models, facilitating rapid prototyping of future algorithms without reliance on explicit protein structures. Taken together, our work establishes a standardized foundation for the advancing research field of SE(3)-based protein design.

1 INTRODUCTION

The design of protein structures is a fundamental challenge in computational biology, with far-reaching implications like drug discovery and enzyme engineering (Quijano-Rubio et al., 2020; Yang et al., 2025; Arunachalam et al., 2021). Recent advances in AI-driven methods (Jumper et al., 2021; Abramson et al., 2024; Watson et al., 2023; Lin et al., 2023) have revolutionized this field, enabling the *de novo* generation of complex, functional proteins. By operating residues in the special Euclidean group $SE(3) = \mathbb{R}^3 \rtimes SO(3)$ and respecting equivariance to rotation and translation, SE(3)-based models demonstrate remarkable quality and diversity in generating protein structures. From multiple perspectives of the diffusion process, researchers have proposed various generative models (DDPM-based (Lin & AlQuraishi, 2023; Lin et al., 2024), Score Matching-based (Yim et al., 2023; Watson et al., 2023) and Flow Matching-based (Bose et al., 2024; Yim et al., 2024) models) to design protein structures. However, due to differences in their dataset and training protocols, it remains challenging to make a fair cross-comparison of these methods. Existing benchmarks like ProteinBench (Ye et al., 2024) and Scaffold-Lab (Zheng et al., 2024) primarily focus on the inference performance, while overlooking the consistent re-training and fair comparison. Furthermore, the implementation of diffusion processes is closely tied to the processing of specific protein data, which hinders the understanding and further development of the underlying mathematical principles. All these challenges motivate us to establish Protein-SE(3), a unified training framework accompanied with comprehensive benchmark for SE(3)-based protein design.

Backend by Pytorch Lightning (Falcon et al., 2019), Protein-SE(3) is the first framework to systematically align diverse methods under identical datasets and training protocols, thereby laying a rigorous foundation for fair, apples-to-apples comparisons. Diverse evaluation metric as Quality (scTM, scRMSD), Diversity (Pairwise TM), and Novelty (Max. TM Score to PDB), are also integrated to analyze the strengths and limitations of different methods on typical protein structure design tasks.

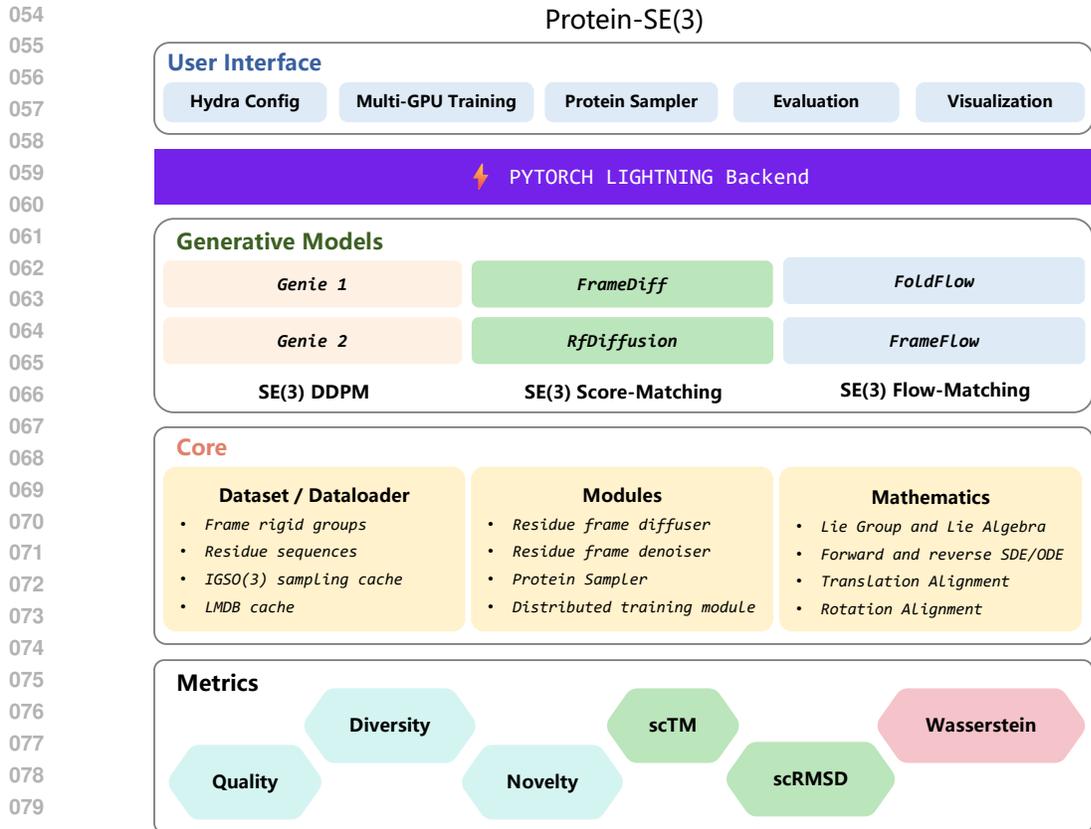


Figure 1: Overview of our proposed Protein-SE(3) benchmark.

084 In addition to the unified training framework and comprehensive, Protein-SE(3) also abstracts high-
 085 level mathematical principles from protein generation models into different perspectives (DDPM,
 086 Score Matching and Flow Matching). Based on the and wasserstein distance, it enables visualization
 087 and analysis of the two marginal diffusion processes in \mathbb{R}^3 and $\text{SO}(3)$ spaces, facilitating agile
 088 prototyping of future algorithms without requiring explicit protein structure data. Taken together, our
 089 work is intended to equip researchers with fresh insights and foster further advances in the field. The
 090 key contributions of this paper can be summarized as follows:

- 091
- 092 • **Unified Training Framework:** We incorporate advanced SE(3)-based protein structure design
 093 methods into a unified training framework, which supports consistent re-training on identical
 094 datasets and protocols, and therefore enables fair comparisons.
 - 095 • **Comprehensive Benchmark:** Diverse evaluation metrics, including Quality (scTM, scRMSD),
 096 Diversity (Pairwise TM), and Novelty (maximum TM score to PDB), are incorporated to compre-
 097 hensively benchmark different methods on typical protein structure design tasks.
 - 098 • **Toolkit for Mathematical Decoupling:** We present a decoupling toolkit that abstracts the mathe-
 099 matical principles of SE(3)-based protein design methods (core formulations are summarized in
 100 the Appendix). Through perspectives of DDPM, Score Matching, and Flow Matching, intuitive
 101 demos illustrate the distribution alignment process in both \mathbb{R}^3 and $\text{SO}(3)$ spaces. This facilitates
 102 rapid prototyping and algorithm development without reliance on explicit protein structure data.

103 2 PRELIMINARIES AND PROBLEM FORMULATION

104 **Backbone parameterization** As shown in Figures 2(a) and 2(b), the parameterization of the
 105 protein backbone mainly follows two paths: the AlphaFold2 frame (Jumper et al., 2021) and the
 106 Frenet-Serret frame (Hu et al., 2011b; Chowdhury et al., 2022). Generally, each residue is associated
 107

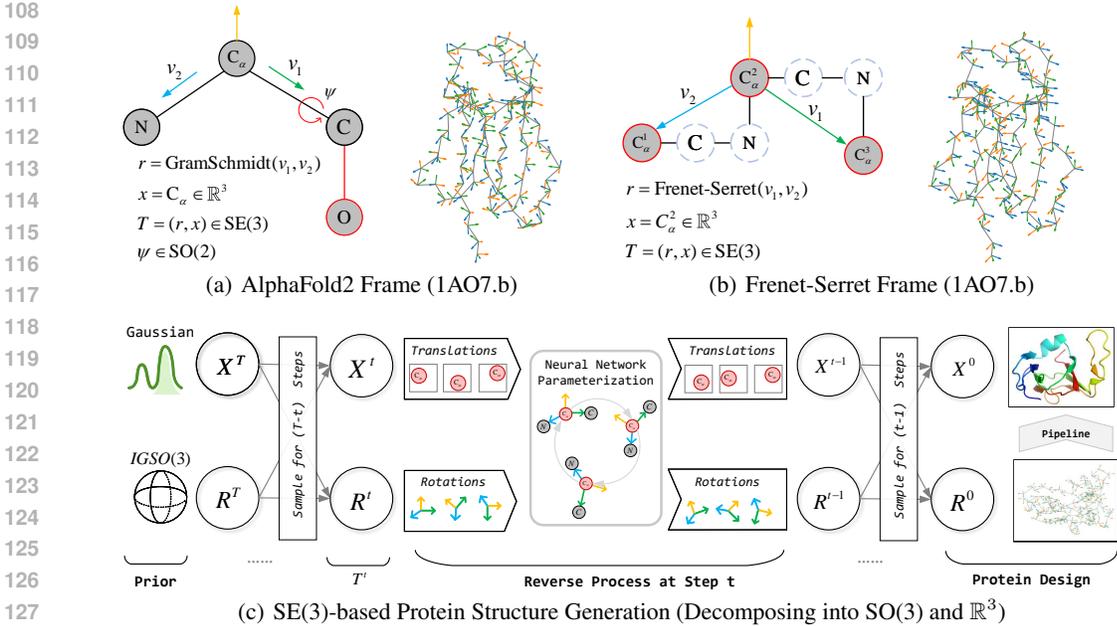


Figure 2: Protein frame parameterization and problem formulation.

with a frame, resulting in N frames that are SE(3)-equivariant for a protein of length N : (1) In the seminal work of AlphaFold2, each frame maps a rigid transformation starting from idealized coordinates of four heavy atoms $[N^*, C_\alpha^*, C^*, O^*] \in \mathbb{R}^3$, with $C_\alpha^* = (0, 0, 0)$ being centered at the origin. Thus, residue $i \in [1, N]$ is represented as an action $T^i = (r^i, x^i) \in \text{SE}(3)$ applied to the idealized frame $[N, C_\alpha, C, O]^i = T^i \circ [N^*, C_\alpha^*, C^*, O^*]$. The coordinate of backbone oxygen atom O is constructed with an additional rotation angle φ ; (2) Another way of backbone parameterization is the Frenet-Serret (FS) frame, which maps each three consecutive C_α into a FS frame. Following (Lin & AlQuraishi, 2023; Lin et al., 2024; Hu et al., 2011a), the FS frame T^i is constructed as:

$$t^i = \frac{x^{i+1} - x^i}{\|x^{i+1} - x^i\|}, b^i = \frac{t^{i-1} \times t^i}{\|t^{i-1} \times t^i\|}, n^i = b^i \times t^i; \quad r^i = [t^i, b^i, n^i], T^i = (r^i, x^i) \quad (1)$$

where the coordinate of the second element is recognized as the translation vector x^i .

Decomposing SE(3) into SO(3) and \mathbb{R}^3 To construct the probability path of SE(3), the definitions of an inner product and a metric on SE(3) are required to obtain a Riemannian structure (Bose et al., 2024; Yim et al., 2023). The common choices in previous studies (Yim et al., 2023; Bose et al., 2024) are:

$$\begin{aligned} \text{Inner Products: } & \langle r, r' \rangle_{\text{SO}(3)} = \text{tr}(rr'^T)/2 \quad \text{and} \quad \langle x, x' \rangle = \sum_{i=1}^3 x_i x'_i \\ \text{Metric on SE(3): } & \langle (r, x), (r', x') \rangle_{\text{SE}(3)} = \langle r, r' \rangle_{\text{SO}(3)} + \langle x, x' \rangle_{\mathbb{R}^3} \end{aligned} \quad (2)$$

As shown in Figure 2(c), the probability path of SE(3) are typically decomposed into SO(3) and \mathbb{R}^3 . The protein structure design task is to learn the reverse process from the prior distributions (Gaussian for \mathbb{R}^3 and IGSO(3) for SO(3)) to the target structure distributions of actual proteins (see Appendix C for the definition of IGSO(3)).

3 DATASETS FOR RE-TRAINING

Unconditional Scaffolding Following the filtering protocol of FrameDiff(Yim et al., 2023), we construct a subset of the Protein Data Bank (PDB) (Berman et al., 2000) by selecting monomeric proteins with sequence lengths ranging from 60 to 512 residues and a structural resolution $< 5\text{\AA}$. After filtering out proteins with more than 50% loops, the dataset is left with 19,703 proteins.

Motif Scaffolding Motif scaffolding problems consist of sequence and structure constraints on motif(s). We use the same protein subset mentioned above but with lengths range from 60 to 320 (17,576 proteins in total). The motif masks are randomly generated during training, following the mask sampling method described in Genie2 (Lin et al., 2024). For evaluation, we use a previously published motif scaffolding benchmark Design24 (Watson et al., 2023) comprising 24 tasks curated from recent publications.

Datasets for training and evaluation are all publicly available at [Harvard Dataverse](#) (organized as the LMDB cache for efficient and parallel data loading).

4 BASELINE MODELS

We ensemble recent open source protein structure design baselines (Lin & AlQuraishi, 2023; Lin et al., 2024; Yim et al., 2023; Bose et al., 2024; Yim et al., 2024) in the unified training framework. The renowned method RfDiffusion (Watson et al., 2023) has received widespread recognition for excelling in *de novo* design, but its training code is unfortunately unavailable, so we report its performance with the official checkpoint as a supplementary reference.

DDPM Methods combine aspects of the SE(3)-equivariant reasoning machinery of IPA with denoising diffusion probabilistic models to create a diffusion process (conditional or unconditional) over protein structures, including Genie1 (Lin & AlQuraishi, 2023) and Genie2 (Lin et al., 2024).

Score-Matching Methods define a forward noising process as Brownian motion on SE(3), where translations in \mathbb{R}^3 and rotations in SO(3) are treated separately but consistently. The score function is then learned to estimate gradients of log-densities on this manifold, enabling efficient reverse sampling that generates realistic protein backbones or scaffolds. Representative approaches include FrameDiff (Yim et al., 2023) and RfDiffusion (Watson et al., 2023).

Flow-Matching Methods is a family of continuous normalizing flow models tailored for distributions on $SE(3)^N$, which directly regress time-dependent vector fields that generate probability paths. Representative methods are FoldFlow (Bose et al., 2024) and FrameFlow (Yim et al., 2024).

5 METRICS

In this section, we briefly introduce the evaluation metrics that will be used to investigate different protein design methods. After re-training on identical training datasets and protocols, all methods integrated in Protein-SE(3) are fairly compared.

Quality. To test whether a model generates designable proteins, we use an *in silico* self-consistency pipeline. The generated structure is first input into an inverse folding model (ProteinMPNN (Dauparas et al., 2022)) to produce 8 candidate sequences, and their structures are then predicted with ESM-Fold (Lin et al., 2023). The consistency between the generated and predicted structure is evaluated using metrics scTM and scRMSD (Zhang & Skolnick, 2005).

Diversity. This metric measures the diversity of generated structures, ensuring the method produces varied backbones rather than replicating known folds. We use the average pairwise TM-score across designable samples and varying lengths as the diversity metric (lower is better).

Novelty. This metric evaluates a method’s ability to explore novel structural space. Novelty is measured by the maximum TM-score between each designed structure and all reference PDB structures, computed with Foldseek (Van Kempen et al., 2024) (lower is better).

Secondary Structure Distribution. For generated backbones, helix and strand percentages reflect secondary structure distribution (Pelton & McLean, 2000; Venyaminov & Yang, 1996). A reasonable distribution should resemble that of natural proteins, rather than being biased toward helices or sheets (Lin et al., 2024; Yue et al., 2025).

6 COMPREHENSIVE BENCHMARK

In this section, we present the benchmark developed to systematically evaluate the design performance of the baseline methods. The experiments are organized as follows:

- **Benchmarking unconditional protein structure design across varying lengths.** We benchmark models for unconditional protein structure design with the aforementioned metrics including Quality, Diversity and Novelty. With length-based performance analysis, this benchmark could serve as a standardized assessment for future work.
- **Benchmarking motif scaffolding with Design24.** Motif scaffolding problems consist of structure and sequence constraints on motifs, along with length constraints on target scaffolds. We evaluate baseline models (RfDiffusion, Genie2, Frameflow) on previously published benchmark Design24 (Watson et al., 2023) with common metrics (scTM and MotifRMSD).
- **In-distribution analysis on secondary structure.** Given the protein backbones generated with different baselines, we record the percentages of the helix and strand, respectively, and visualize the distribution with respect to the percentages. In-distribution analysis assesses how well the model generalizes to natural proteins, rather than biased to a specific structural class.
- **Efficiency Analysis.** To measure the GPU memory and time needed to train an SE(3)-based generative model or generate protein structures, we report training time, inference time, step count, and model size across different methods. Although efficiency may be less critical compared to other metrics, it remains a useful metric to assess the model scalability and practicality.

6.1 LENGTH-BASED PERFORMANCE ANALYSIS FOR UNCONDITIONAL SCAFFOLDING

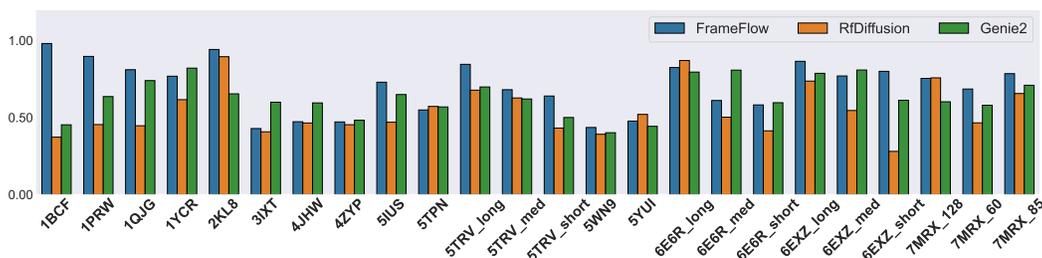
Table 1: Performance of generative models for unconditional structure design across varying lengths. We highlight the **best performance** in bold and the second-best with the underline. RfDiffusion is evaluated with the official checkpoint (since the training code is not available). The superscript * indicates that the generation quality does not meet the standard (scTM \geq 0.5), so Novelty and Diversity are excluded from the comparison.

Method	Length 100				Length 200			
	Quality		Novelty		Quality		Diversity	
	scTM \uparrow	scRMSD \downarrow	Max TM \downarrow	pairwise TM \downarrow	scTM \uparrow	scRMSD \downarrow	Max TM \downarrow	pairwise TM \downarrow
Genie1	0.89 \pm 0.11	1.25 \pm 0.98	<u>0.30\pm0.09</u>	<u>0.35\pm0.06</u>	0.72 \pm 0.23	5.27 \pm 4.60	0.23 \pm 0.12	<u>0.32\pm0.04</u>
Genie2	0.91 \pm 0.08	1.04 \pm 0.64	0.29\pm0.07	0.39 \pm 0.05	0.77 \pm 0.19	4.01 \pm 3.48	0.21 \pm 0.09	0.33 \pm 0.03
FrameDiff	0.92\pm0.04	0.93\pm0.39	0.39 \pm 0.13	0.37 \pm 0.06	0.81 \pm 0.16	3.11 \pm 3.08	0.34 \pm 0.13	0.35 \pm 0.07
RfDiffusion	0.97 \pm 0.01	0.52 \pm 0.10	0.34 \pm 0.13	0.39 \pm 0.07	0.97 \pm 0.02	0.63 \pm 0.15	0.31 \pm 0.10	0.35 \pm 0.06
FrameFlow	0.90 \pm 0.10	1.13 \pm 1.03	0.38 \pm 0.14	0.33\pm0.07	0.94\pm0.04	1.24\pm0.43	0.28 \pm 0.13	0.30\pm0.04
FoldFlow-Base	<u>0.92\pm0.05</u>	<u>0.99\pm0.36</u>	0.36 \pm 0.12	0.46 \pm 0.06	0.91 \pm 0.04	<u>1.50\pm0.49</u>	0.21 \pm 0.05	0.34 \pm 0.04
FoldFlow-OT	0.91 \pm 0.07	1.17 \pm 0.88	0.34 \pm 0.08	0.43 \pm 0.06	<u>0.91\pm0.03</u>	1.63 \pm 0.60	0.20\pm0.04	0.34 \pm 0.05
FoldFlow-SFM	0.87 \pm 0.06	1.39 \pm 0.55	0.33 \pm 0.10	0.44 \pm 0.06	0.82 \pm 0.22	3.76 \pm 3.12	<u>0.21\pm0.05</u>	0.35 \pm 0.04
Method	Length 300				Length 500			
	Quality		Novelty		Quality		Diversity	
	scTM \uparrow	scRMSD \downarrow	Max TM \downarrow	pair TM \downarrow	scTM \uparrow	scRMSD \downarrow	Max TM \downarrow	pair TM \downarrow
Genie1	0.68 \pm 0.19	6.57 \pm 3.80	0.37 \pm 0.22	0.33 \pm 0.07	0.47 \pm 0.12*	14.39 \pm 4.22	0.11 \pm 0.03	0.31 \pm 0.04
Genie2	0.64 \pm 0.20	7.56 \pm 4.33	0.15\pm0.04	0.35 \pm 0.03	0.46 \pm 0.08*	14.28 \pm 2.92	0.13 \pm 0.04	0.35 \pm 0.04
FrameDiff	0.72 \pm 0.13	5.40 \pm 2.97	0.34 \pm 0.15	0.36 \pm 0.09	0.64\pm0.19	9.68\pm5.17	0.20 \pm 0.08	<u>0.34\pm0.05</u>
RfDiffusion	0.94 \pm 0.04	1.04 \pm 0.89	0.32 \pm 0.12	0.38 \pm 0.04	0.90 \pm 0.11	3.65 \pm 2.95	0.24 \pm 0.08	0.37 \pm 0.06
FrameFlow	0.90\pm0.06	2.00\pm0.89	0.30 \pm 0.12	<u>0.32\pm0.09</u>	0.56 \pm 0.20	<u>11.10\pm6.12</u>	<u>0.17\pm0.08</u>	0.35 \pm 0.07
FoldFlow-Base	<u>0.87\pm0.12</u>	<u>2.82\pm2.71</u>	0.16 \pm 0.04	0.33 \pm 0.03	<u>0.57\pm0.23</u>	11.67 \pm 6.70	0.13\pm0.03	0.29\pm0.03
FoldFlow-OT	0.70 \pm 0.21	6.01 \pm 4.43	<u>0.16\pm0.03</u>	0.34 \pm 0.04	0.38 \pm 0.05*	13.35 \pm 3.25	0.12 \pm 0.05	0.32 \pm 0.04
FoldFlow-SFM	0.68 \pm 0.23	7.54 \pm 6.53	0.17 \pm 0.04	0.30\pm0.03	0.37 \pm 0.08*	15.46 \pm 3.70	0.12 \pm 0.03	0.32 \pm 0.05

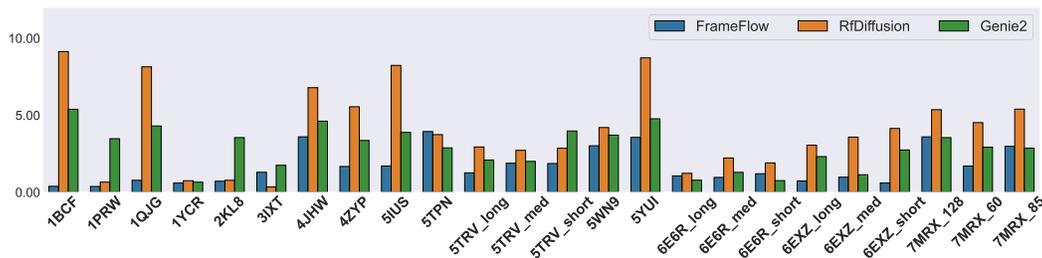
We first retrain baseline models (except RfDiffusion), and then evaluate them for unconditional scaffolding across varying lengths. The results are presented in Table 1, which is based on our previously introduced metrics (Quality, Novelty and Diversity).

In terms of the Quality metric (scTM and scRMSD), flow-matching based methods (FrameFlow and Foldflow) demonstrates relatively better performance, which is in line with the mathematical analysis in Section 7. Novelty is also an essential metric to evaluates a method’s capacity to explore new protein structures. With quality constraint (scTM>0.5), FoldFlow and Genie2 demonstrate strong performance in generating novel structures. Based on the structural diversity metric, flow-matching models still demonstrates impressive performance across the different chain lengths. It’s noteworthy that the performance of all methods across various metrics shows a decline trend as the protein length increases, which suggests that these models generally struggle to create larger proteins due to the increased conformational complexity and diversity.

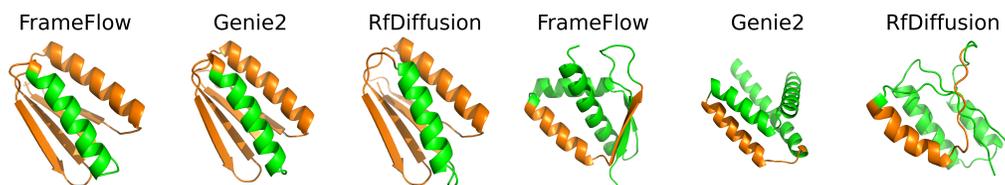
6.2 MOTIF SCAFFOLDING RESULTS ON DESIGN24



(a) Average scTM of Motif Scaffolding Designs (Higher is Better)



(b) Average MotifRMSD of Motif Scaffolding Designs (Lower is Better)



(c) Motif Scaffolding Cases (2KL8)

(d) Motif Scaffolding Cases (5TRV_med)

Figure 3: Motif scaffolding results based on Design24. In the case visualization, **motif regions** are colored orange and **scaffolding regions** green.

We consider three baselines that support the motif scaffolding task. RfDiffusion is evaluated with the official checkpoint (no source code for re-training), while FrameFlow and Genie2 are re-trained with our processed protein dataset with the length ranging from 60 to 320. with 4 scaffolding samples (using the same pipeline for quality evaluation) for each motif, the average scTM and MotifRMSD (scRMSD that only considers the motif region) of three methods are illustrated in Figure 3. Frameflow achieves the most designable scaffolds (highest scTM) among all methods in 13 out of 24 test motifs compared to Genie2’s 7/24 and RfDiffusion’s 6/24. For the MotifRMSD metric, FrameFlow still demonstrates superior performance by generating 19 proteins with the lowest MotifRMSD out of 24 tests, compared to the Rfdiffusion’s 1/24 and Genie2’s 4/24.

6.3 IN-DISTRIBUTION ANALYSIS ON THE SECONDARY STRUCTURE

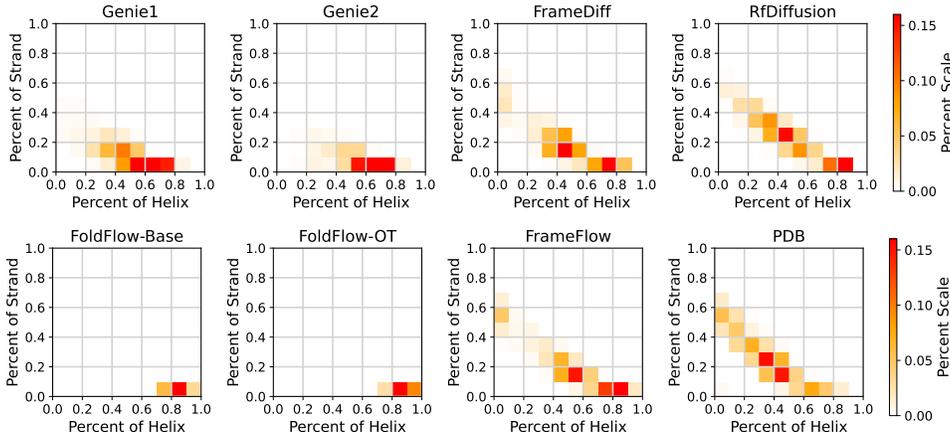


Figure 4: The distribution of secondary structure on unconditionally generated protein structures.

Given 5 unconditionally generated protein structures with every length ranging from 60 to 320 (1,040 structures in total), we report the secondary structure distribution (the percentages of helix and strand) of each method in Figure 4. The proteins generated by official RfDiffusion, re-trained FrameFlow and FrameDiff have more reasonable distributions, which is similar to those randomly sampled from the PDB (bottom right). However, the distributions of FoldFlow, Genie1 and Genie2 differ significantly from the PDB distribution, indicating the risk to generate protein backbones consistently dominated by helical structures. This explains why the performance of these methods is inferior to that of the others in the Quality and Diversity, as shown in the Table 1.

6.4 COMPUTATIONAL EFFICIENCY

Table 2: Efficiency comparison for protein generation models.

Method	Epochs	Training Time	GPUs	Model Size	Sample Steps	Inference Time
Genie1 (Lin & AlQuraishi, 2023)	100	~3.0 days	2×A100	4.10M	1k	40 hours
Genie2 (Lin et al., 2024)	100	~3.0 days	4×A100	15.7M	1k	26 hours
FrameDiff (Yim et al., 2023)	150	~4.2 days	2×L20	18.8M	100	31 hours
FoldFlow (Bose et al., 2024)	100	~3.5 days	2×L20	17.4M	100	3.8 hours
FrameFlow (Yim et al., 2024)	800	~2.0 days	2×L20	11.3M	100	1.9 hours

To encourage developing more efficient and scalable models in the future, we also benchmark the training cost, evaluation cost, model size, and sample steps in Table 2. Note that the inference time is the total time for generating 1040 proteins with lengths ranging from 60 to 320 (following the setup for in-distribution analysis in Section 6.3). Here we found that:

- Genie1 and Genie2 consume the most GPU memory during training and time for inference, primarily due to the $O(N^3)$ scaling of triangular multiplicative update layers (Lin et al., 2024).
- As for inference efficiency, Flow-Matching methods (FrameFlow and FoldFlow) significantly outperform DDPM (Genie1 and Genie2) and Score-Matching (FrameDiff) approaches. This is primarily because flow matching leverages ordinary differential equations (ODEs) to model probability paths (Chen et al., 2023; Frans et al., 2024), allowing fewer sample steps for generation.

7 TOOLKIT FOR MATHEMATICAL DECOUPLING

SE(3)-based methods model and align protein structures on both the translational \mathbb{R}^3 and rotational SO(3) spaces. To abstract and abstract the mathematical principles underlying protein structure design

378 models, we developed a mathematical decoupling toolkit to visualize and analyze the distribution
 379 alignment process with the measurement of 1st-order Wasserstein distance (see [Appendix C](#)
 380 for the definition). This toolkit is built upon simple MLP layers training on the synthetic data (3D
 381 coordinates and rotation matrices represented as Euler angles), enabling systematic study on how
 382 various generative models (DDPM (Ho et al., 2020; Leach et al., 2022), Score Matching (Song
 383 et al., 2020), and Flow Matching (Bose et al., 2024)) align distributions in both the translational and
 384 rotational spaces.

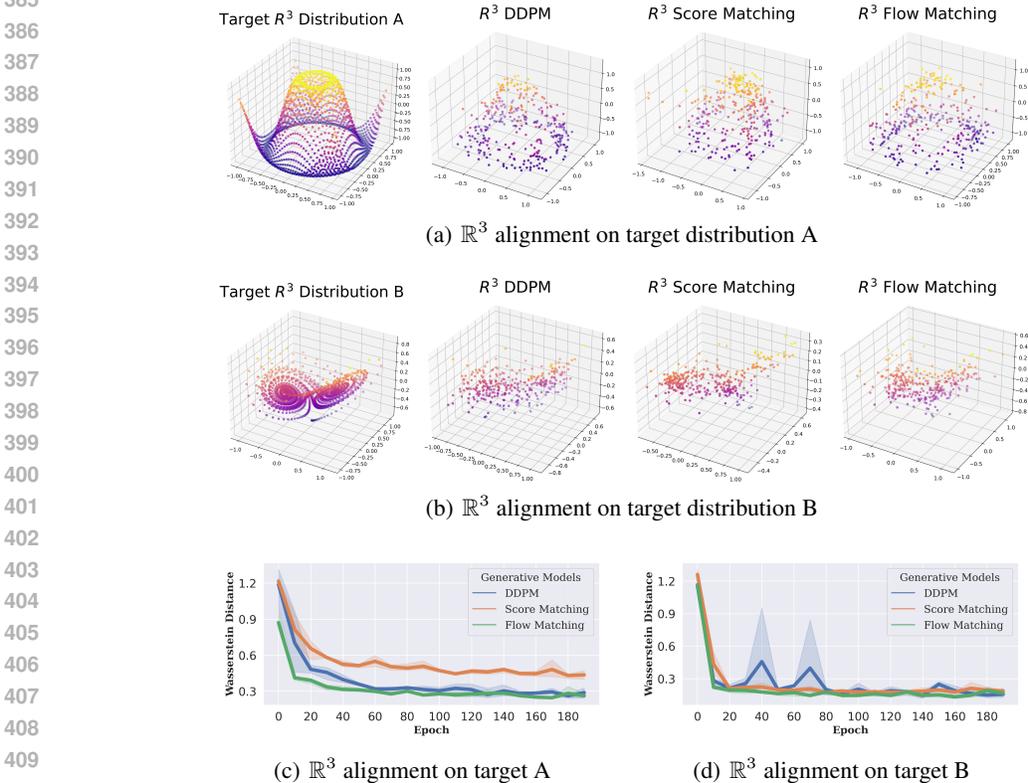
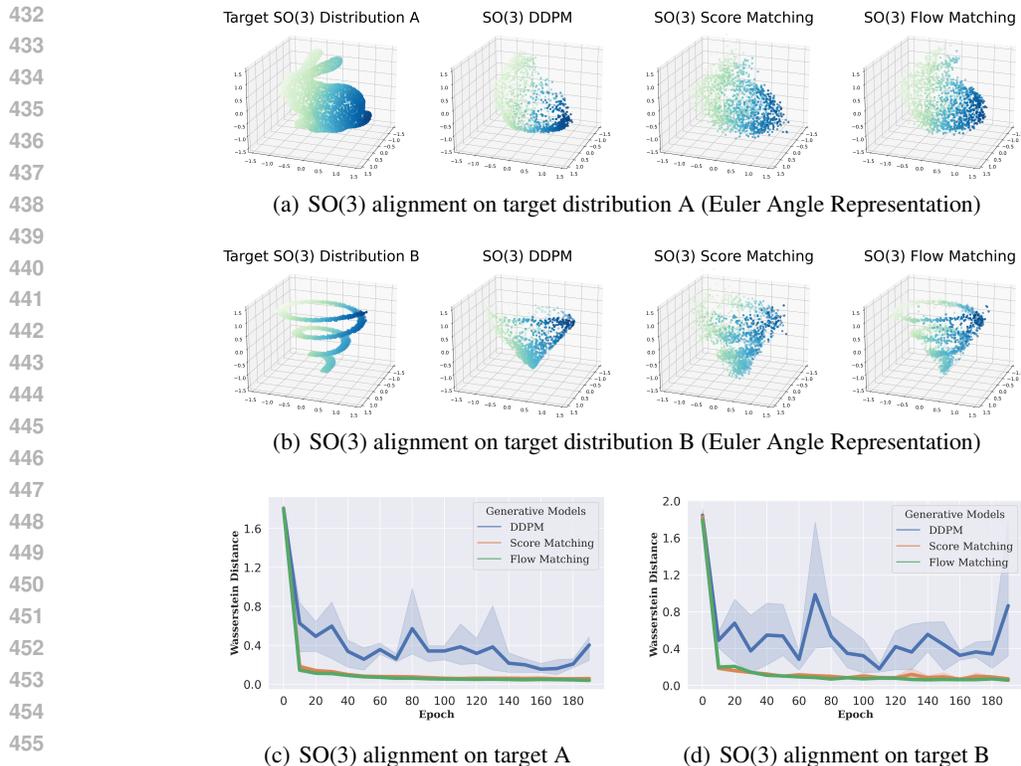


Figure 5: Experiments on \mathbb{R}^3 Alignment with different generative modeling methods.

Translation alignment in \mathbb{R}^3 space The translations (C_α positions) of proteins are defined in the standard \mathbb{R}^3 space, where the probability path can be constructed easily through the previously derived closed form equations (Luo & Hu, 2021). Detailed formulations of translation alignment, as derived from different generative models, are presented in [Appendix D](#). While figure 5 provides a visualization of the translation alignment: as training epochs increase, the \mathbb{R}^3 distributions sampled by three generative models become progressively closer to the target distribution, as evidenced by the decreasing Wasserstein Distance, ultimately achieving alignment with the target distribution.

Rotation alignment on $SO(3)$ manifold Different from Euclidean space, $SO(3)$ is a Riemannian manifold (Lee, 2018). Therefore, the probability path on $SO(3)$ has been the focus of previous studies, requiring thoughtful design for the following reasons: (1) Arithmetic operations on Riemannian manifolds are not linearly defined; (2) Modeling noise with IG $SO(3)$ rather than a Gaussian distribution; To abstract the mathematical principles behind rotation alignment from different perspectives, we first provide definitions and formulations in [Appendix E](#). Then we use two sets of synthetic rotation matrices as the target $SO(3)$ distributions, which is visualized using the Euler-angle representation in Figure 6 (the range of three Euler angles is set to $[-\pi/2, \pi/2]$). With increasing training epochs, the $SO(3)$ distributions sampled by the generative models gradually converge to the target distribution, as confirmed by the decreasing 1st-order Wasserstein Distance.

Compared to DDPM and Score Matching, the curves of Flow Matching on \mathbb{R}^3 and $SO(3)$ alignment exhibit better convergence, indicating the superior design performance evaluated in Section 6.



457 Figure 6: Experiments on SO(3) alignment with different generative modeling methods.

460 8 INSIGHTS FOR FUTURE METHOD DESIGN

461
462 To conclude our experimental results, we summarize several key insights that may inspire the design
463 of future methods, with the aim of facilitating broader progress in the community:

- 465 • **Flow Matching as a Superior Choice:** As revealed by our mathematical decoupling toolkit,
466 the alignment process of Flow Matching in the \mathbb{R}^3 and SO(3) space exhibits superior efficiency
467 and stability. This suggests flow matching generally outperforms other models for modeling the
468 generation process, and recent progress in the Flow Matching community (MeanFlow (Geng
469 et al., 2025), Shortcut Model (Frans et al., 2024)) can be leveraged for future method design.
- 470 • **Distribution-level Guidance Control:** Despite most methods achieving remarkable results
471 on certain evaluation metrics, the secondary structure distributions of their generated proteins
472 differ substantially from that of native PDBs. This observation reveals the need to introduce
473 distribution-level guidance into the training process (in line with ProtFID (Faltings et al., 2025)).
- 474 • **SE(3) Invariant Encoder:** Genie1 and Genie2 consume the most GPU memory during training
475 and the most time for inference, primarily due to the scaling of triangular multiplicative update
476 layers (Jumper et al., 2021), which indicates the need for a more efficient SE(3)-invariant encoder.

478 9 CONCLUSION AND FUTURE WORK

479
480 We propose Protein-SE(3), a unified framework and comprehensive benchmark for SE(3)-based
481 protein design approaches, which enables fair comparison with consistent re-training on identical
482 datasets and protocols. The developed mathematical toolkit provide intuitive demos to interpret the
483 distribution alignment process in the \mathbb{R}^3 and SO(3) space. In the future, we will gradually broaden
484 our scope beyond SE(3)-based structure generation algorithms (Geffner et al., 2025; Campbell et al.,
485 2024; Wang et al., 2024), and will keep Protein-SE(3) updated in accordance with the latest research
(some preliminary results using the official checkpoints are presented in [Appendix B](#))

486 ETHICS STATEMENT
487

488 We confirm that our research adheres to the Code of Ethics. This work does not involve human
489 subjects, personal data, or any proprietary or sensitive information. The source of constructed
490 datasets are publicly available and widely adopted in the community, and we have followed best
491 practices regarding licensing, documentation, and responsible use. The methods proposed are
492 intended for advancing scientific understanding and educational purposes, without foreseeable
493 harmful applications. We are committed to principles of fairness, transparency, and research integrity,
494 and we have carefully considered potential concerns related to bias, privacy, and security in our study.
495

496 REPRODUCIBILITY STATEMENT
497

498 We have made every effort to ensure the reproducibility of our work. The dataset construction
499 process is described in detail in Section 3. Furthermore, we include the complete source code in
500 the supplementary material, allowing others to reproduce our unified framework and comprehensive
501 benchmark. Together, these resources are intended to facilitate independent verification of our results
502 and foster transparent scientific progress.
503

504 REFERENCES
505

- 506 Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf
507 Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure
508 prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.
- 509 Prabhu S Arunachalam, Alexandra C Walls, Nadia Golden, Caroline Atyeo, Stephanie Fischinger,
510 Chunfeng Li, Pyone Aye, Mary Jane Navarro, Lilin Lai, Venkata Viswanadh Edara, et al. Adju-
511 vanting a subunit covid-19 vaccine to induce protective immunity. *Nature*, 594(7862):253–258,
512 2021.
- 513 Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, T N Bhat, Helge Weissig, Ilya N
514 Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242,
515 2000.
- 516 Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrid Rector-Brooks, Cheng-
517 Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong.
518 SE(3)-stochastic flow matching for protein backbone generation. In *The Twelfth International*
519 *Conference on Learning Representations*, 2024. URL [https://openreview.net/forum?](https://openreview.net/forum?id=kJFIH23hXb)
520 [id=kJFIH23hXb](https://openreview.net/forum?id=kJFIH23hXb).
- 521 Andrew Campbell, Jason Yim, Regina Barzilay, Tom Rainforth, and Tommi Jaakkola. Generative
522 flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design.
523 *arXiv preprint arXiv:2402.04997*, 2024.
- 524 Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, and Adil Salim. The probability flow
525 ode is provably fast. *Advances in Neural Information Processing Systems*, 36:68552–68575, 2023.
- 526 Ratul Chowdhury, Nazim Bouatta, Surojit Biswas, Christina Floristean, Anant Kharkar, Koushik Roy,
527 Charlotte Rochereau, Gustaf Ahdritz, Joanna Zhang, George M Church, et al. Single-sequence
528 protein structure prediction using a language model and deep learning. *Nature Biotechnology*, 40
529 (11):1617–1623, 2022.
- 530 Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles,
531 Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning-based
532 protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.
- 533 William Falcon et al. Pytorch lightning. *GitHub*. Note: <https://github.com/Lightning-AI/lightning>,
534 2019.
- 535 Felix Faltings, Hannes Stark, Tommi Jaakkola, and Regina Barzilay. Protein fid: Improved evaluation
536 of protein structure generative models. *arXiv preprint arXiv:2505.08041*, 2025.
537
538
539

- 540 Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut
541 models. *arXiv preprint arXiv:2410.12557*, 2024.
- 542
- 543 Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario
544 Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, et al. Proteina: Scaling flow-based
545 protein structure generative models. *arXiv preprint arXiv:2503.00710*, 2025.
- 546 Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for
547 one-step generative modeling. *arXiv preprint arXiv:2505.13447*, 2025.
- 548
- 549 Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in*
550 *Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.
- 551 Shuangwei Hu, Martin Lundgren, and Antti J Niemi. Discrete frenet frame, inflection point solitons,
552 and curve visualization with applications to folded proteins. *Physical Review E—Statistical,*
553 *Nonlinear, and Soft Matter Physics*, 83(6):061908, 2011a.
- 554
- 555 Shuangwei Hu, Martin Lundgren, and Antti J Niemi. Discrete frenet frame, inflection point solitons,
556 and curve visualization with applications to folded proteins. *Physical Review E—Statistical,*
557 *Nonlinear, and Soft Matter Physics*, 83(6):061908, 2011b.
- 558 John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger,
559 Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate
560 protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.
- 561
- 562 Leonid V Kantorovich. Mathematical methods of organizing and planning production. *Management*
563 *science*, 6(4):366–422, 1960.
- 564 Adam Leach, Sebastian M Schmon, Matteo T. Degiacomi, and Chris G. Willcocks. Denoising
565 diffusion probabilistic models on SO(3) for rotational alignment. In *ICLR 2022 Workshop on*
566 *Geometrical and Topological Representation Learning*, 2022. URL <https://openreview.net/forum?id=BY88eBbkpe5>.
- 567
- 568 John M Lee. *Introduction to Riemannian manifolds*, volume 2. Springer, 2018.
- 569
- 570 Yeqing Lin and Mohammed AlQuraishi. Generating novel, designable, and diverse protein structures
571 by equivariantly diffusing oriented residue clouds. In *Proceedings of the 40th International*
572 *Conference on Machine Learning*, pp. 20978–21002, 2023.
- 573
- 574 Yeqing Lin, Minji Lee, Zhao Zhang, and Mohammed AlQuraishi. Out of many, one: Designing
575 and scaffolding proteins at the scale of the structural universe with genie 2. *arXiv preprint*
576 *arXiv:2405.15489*, 2024.
- 577 Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin,
578 Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level
579 protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- 580
- 581 Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings*
582 *of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2837–2845, 2021.
- 583 Dmitry I Nikolayev and Tatjana I Savyolov. Normal distribution on the rotation group so(3). *Texture,*
584 *Stress, and Microstructure*, 29(3-4):201–233, 1997.
- 585
- 586 John T Pelton and Larry R McLean. Spectroscopic methods for analysis of protein secondary
587 structure. *Analytical biochemistry*, 277(2):167–176, 2000.
- 588
- 589 Alfredo Quijano-Rubio, Umut Y Ulge, Carl D Walkey, and Daniel-Adriano Silva. The advent of
590 de novo proteins for cancer immunotherapy. *Current Opinion in Chemical Biology*, 56:119–128,
591 2020.
- 592 Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben
593 Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint*
arXiv:2011.13456, 2020.

- 594 Michel Van Kempen, Stephanie S Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee,
595 Cameron LM Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein
596 structure search with foldseek. *Nature biotechnology*, 42(2):243–246, 2024.
- 597 Sergei Yu Venyaminov and Jen Tsi Yang. Determination of protein secondary structure. In *Circular*
598 *dichroism and the conformational analysis of biomolecules*, pp. 69–107. Springer, 1996.
- 600 Cédric Villani and Cédric Villani. The wasserstein distances. *Optimal transport: old and new*, pp.
601 93–111, 2009.
- 602 Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Dplm-2: A
603 multimodal diffusion protein language model. *arXiv preprint arXiv:2410.13782*, 2024.
- 605 Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach,
606 Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein
607 structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.
- 608 Wei Yang, Derrick R Hicks, Agnidipta Ghosh, Tristin A Schwartz, Brian Coventry, Inna Goreschnik,
609 Aza Allen, Samer F Halabiya, Chan Johng Kim, Cynthia S Hinck, et al. Design of high-affinity
610 binders to immune modulating receptors for cancer immunotherapy. *Nature Communications*, 16
611 (1):2001, 2025.
- 613 Fei Ye, Zaixiang Zheng, Dongyu Xue, Yuning Shen, Lihao Wang, Yiming Ma, Yan Wang, Xinyou
614 Wang, Xiangxin Zhou, and Quanquan Gu. Proteinbench: A holistic evaluation of protein foundation
615 models. *arXiv preprint arXiv:2409.06744*, 2024.
- 616 Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay,
617 and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. In
618 *Proceedings of the 40th International Conference on Machine Learning*, pp. 40001–40039, 2023.
- 620 Jason Yim, Andrew Campbell, Emile Mathieu, Andrew Y. K. Foong, Michael Gastegger, Jose
621 Jimenez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S. Veeling, Frank Noe, Regina
622 Barzilay, and Tommi Jaakkola. Improved motif-scaffolding with SE(3) flow matching. *Transactions*
623 *on Machine Learning Research*, 2024. ISSN 2835-8856. URL [https://openreview.net/](https://openreview.net/forum?id=falne8xDGn)
624 [forum?id=falne8xDGn](https://openreview.net/forum?id=falne8xDGn).
- 625 Angxiao Yue, Zichong Wang, and Hongteng Xu. Reqflow: Rectified quaternion flow for efficient and
626 high-quality protein backbone generation. *arXiv preprint arXiv:2502.14637*, 2025.
- 627 Yang Zhang and Jeffrey Skolnick. Tm-align: a protein structure alignment algorithm based on the
628 tm-score. *Nucleic acids research*, 33(7):2302–2309, 2005.
- 630 Zhuoqi Zheng, Bo Zhang, Bozitao Zhong, Kexin Liu, Zhengxin Li, Junjie Zhu, Jinyu Yu, Ting Wei,
631 and Hai-Feng Chen. Scaffold-lab: Critical evaluation and ranking of protein backbone generation
632 methods in a unified framework. *bioRxiv*, pp. 2024–02, 2024.
- 633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

APPENDIX

A USE OF LARGE LANGUAGE MODELS

We used an LLM solely to aid and polish writing. Specifically, OpenAI’s ChatGPT was employed for copy-editing (grammar, fluency, and tone), minor phrasing suggestions, and occasional condensation or expansion of author-written paragraphs without altering their scientific meaning.

B PRELIMINARY RESULTS ON RECENT METHODS

Table 3: Preliminary Evaluation Results of Recent Methods beyond SE(3) Paradigm.

Method	scTM \uparrow	scRMSD \downarrow	MaxTM \downarrow	PairwiseTM \downarrow	Information of Protein	Architecture
Length 100						
Proteina	0.95 \pm 0.03	0.75 \pm 0.23	0.48 \pm 0.19	0.36 \pm 0.07	C_α Coordinates Only	Flow Matching
Proteina-Long	0.88 \pm 0.06	1.01 \pm 0.42	0.40 \pm 0.24	0.32 \pm 0.05	C_α Coordinates Only	Flow Matching
MultiFlow	0.98 \pm 0.01	0.51 \pm 0.09	0.44 \pm 0.12	0.39 \pm 0.07	SE(3) Frames + Sequences	Flow Matching
DPLM-2	0.88 \pm 0.07	1.31 \pm 0.65	0.84 \pm 0.16	0.31 \pm 0.15	C_α Coordinates + Sequences	Language Model
Length 300						
Proteina	0.87 \pm 0.06	4.39 \pm 2.38	0.21 \pm 0.14	0.31 \pm 0.04	C_α Coordinates Only	Flow Matching
Proteina-Long	0.95 \pm 0.03	1.01 \pm 0.42	0.40 \pm 0.24	0.32 \pm 0.05	C_α Coordinates Only	Flow Matching
MultiFlow	0.98 \pm 0.04	0.64 \pm 0.11	0.24 \pm 0.12	0.33 \pm 0.05	SE(3) Frames + Sequences	Flow Matching
DPLM-2	0.94 \pm 0.05	1.76 \pm 1.63	0.83 \pm 0.14	0.27 \pm 0.03	C_α Coordinates + Sequences	Language Model
Length 500						
Proteina	0.41 \pm 0.08	44.1 \pm 5.26	0.20 \pm 0.10	0.23 \pm 0.03	C_α Coordinates Only	Flow Matching
Proteina-Long	0.93 \pm 0.10	2.22 \pm 2.13	0.32 \pm 0.18	0.31 \pm 0.06	C_α Coordinates Only	Flow Matching
MultiFlow	0.97 \pm 0.01	1.01 \pm 0.26	0.33 \pm 0.06	0.17 \pm 0.05	SE(3) Frames + Sequences	Flow Matching
DPLM-2	0.84 \pm 0.12	3.90 \pm 2.82	0.69 \pm 0.21	0.28 \pm 0.13	C_α Coordinates + Sequences	Language Model

In addition to the methods integrated in our proposed Protein-SE(3), we also have evaluate several recent approaches beyond SE(3) paradigm, including Proteina (Geffner et al., 2025), MultiFlow (Campbell et al., 2024), and DPLM-2 (Wang et al., 2024). For these methods, we use the official checkpoints released by the authors, while noting that retraining under our standardized dataset and framework is left for future work. The corresponding results are reported in Table 3, which further indicate that the method MultiFlow incorporating SE(3) information achieves superior performance, particularly on Quality-related metrics such as scTM and scRMSD, suggesting that SE(3)-based approaches merit greater attention and exploration.

C DEFINITIONS

Definition 1. IGSO(3): The Isotropic Gaussian Distribution on SO(3) is a probability distribution over the 3D rotation group SO(3), which generalizes the idea of a Gaussian (normal) distribution to the non-Euclidean manifold of 3D rotations (Nikolayev & Savyolov, 1997). Detailed sampling process is described in the [Appendix E.1](#).

Definition 2. Wasserstein Distance: The Wasserstein distance (Kantorovich, 1960; Villani & Villani, 2009) measures the minimum cost required to transform one probability distribution into another. For probability distributions μ and ν over a metric space χ (e.g. \mathbb{R}^3 and SO(3)), let $\Gamma(\mu, \nu)$ is the set of all joint distributions on $\chi \times \chi$, the 1st-order Wasserstein distance is defined as:

$$W_1(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x,y) \sim \gamma} [\|x - y\|] \quad (3)$$

where $\gamma(x, y)$ represents a transport plan and $\|x - y\|$ denotes the transport cost from point x to y .

D THEORETICAL FORMULATIONS OF \mathbb{R}^3 ALIGNMENT

The translations (C_α positions) of proteins are defined in the standard \mathbb{R}^3 Euclidean space, where the probability path can be constructed easily through the previously derived closed form equations. From different perspectives (DDPM (Lin & AlQuraishi, 2023), Score Matching (Yim et al., 2023) and Flow Matching (Bose et al., 2024)), we summarize core formulations for \mathbb{R}^3 distribution alignment implemented in Section 7.

D.1 TRANSLATION ALIGNMENT BASED ON DDPM

Here, we present the formulations of DDPM-based translation alignment following the description of Genie1 (Lin & AlQuraishi, 2023). Let $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ denote a sequence of C_α coordinates of length N . Given a sample \mathbf{x}_0 from the target distribution over the synthetic data, the forward process iteratively adds Gaussian noise to the sample with a cosine variance schedule $\beta = [\beta_1, \beta_2, \dots, \beta_T]$, where the diffusion steps T is set to 1,000:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t | \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (4)$$

By applying the reparameterization trick to the forward process, we have

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t | \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}),$$

$$\bar{\alpha}_t = \prod_{i=1}^t \alpha_i, \quad \alpha_t = 1 - \beta_t \quad (5)$$

According to the derivation of DDPM, the reverse process is modeled with a Gaussian distribution:

$$p(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1} | \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t) \mathbf{I}),$$

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right), \quad \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t) = \beta_t \quad (6)$$

This reverse process requires evaluating $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ based on MLP layers, which predict the noise added at time step t . The loss function is defined as:

$$L = \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}} \left[\sum_{i=1}^N \|\boldsymbol{\epsilon}_t^i - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|^2 \right]$$

$$= \mathbb{E}_{t, \mathbf{x}_0, \boldsymbol{\epsilon}} \left[\sum_{i=1}^N \|\boldsymbol{\epsilon}_t^i - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}_t, t)\|^2 \right] \quad (7)$$

where $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}^1, \boldsymbol{\epsilon}^2, \dots, \boldsymbol{\epsilon}^N]$ and each $\boldsymbol{\epsilon}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

D.2 TRANSLATION ALIGNMENT BASED ON SCORE MATCHING

Following FrameDiff (Yim et al., 2023), the process of \mathbb{R}^3 alignment is modeled as an Ornstein–Uhlenbeck process (also called VP-SDE (Song et al., 2020)). Still, Let $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ denote a sequence of C_α coordinates of length N . Converging geometrically towards Gaussian, the VP-SDE of \mathbf{x} in the \mathbb{R}^3 space is:

$$d\mathbf{x} = -\frac{1}{2} \beta(t) \mathbf{x} dt + \sqrt{\beta(t)} d\mathbf{w} \quad (8)$$

where $\beta(t)$ is a non-negative function of t to describe the time-dependent noise schedule, and \mathbf{w} is the standard Wiener process. Its analytical solution is:

$$\mathbf{x}_t = \alpha(t) \mathbf{x}_0 + \sigma(t) \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\alpha(t) = \exp\left(-\frac{1}{2} \int_0^t \beta(s) ds\right), \quad \sigma^2(t) = 1 - \alpha^2(t) \quad (9)$$

To train a score-based model, we want the network $s_\theta(x_t, t)$ to approximate the score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0)$. Note that the conditional distribution is Gaussian:

$$p(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\alpha(t)\mathbf{x}_0, \sigma^2(t)\mathbf{I}), \quad (10)$$

so the formulation of the true score becomes:

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0) = -\frac{1}{\sigma^2(t)}(\mathbf{x}_t - \alpha(t)\mathbf{x}_0). \quad (11)$$

With the simple MSE loss function to train the network $s_\theta(x_t, t)$:

$$L = \|s_\theta(x_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t | \mathbf{x}_0)\|^2, \quad (12)$$

the reverse process from \mathbf{x}_t to \mathbf{x}_{t-1} can be formulated as:

$$d\mathbf{x} = \left[-\frac{1}{2}\beta(t)\mathbf{x} + \beta(t)s_\theta(x, t) \right] dt + \sqrt{\beta(t)}d\bar{\mathbf{w}}, \quad (13)$$

where $\bar{\mathbf{w}}$ is denotes Wiener process run backward in time.

D.3 TRANSLATION ALIGNMENT BASED ON FLOW MATCHING

Flow Matching in the \mathbb{R}^3 space aims to learn a time-dependent velocity field $u_\theta(\mathbf{x}, t)$ such that the solution to the following ordinary differential equation (ODE):

$$\frac{d\mathbf{x}}{dt} = u_\theta(\mathbf{x}, t), \quad (14)$$

which maps samples from a known base distribution $p_0(\mathbf{x}_0)$ to samples from the target distribution $p_1(\mathbf{x}_1)$ over time $t \in [0, 1]$. In the \mathbb{R}^3 space, the target velocity at time t is simply:

$$v_t = \mathbf{x}_1 - \mathbf{x}_0 \quad (15)$$

The model is trained to predict this velocity via a regression loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0, \mathbf{x}_1, t} \left[\|u_\theta(\mathbf{x}_t, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|^2 \right] \quad (16)$$

Once the model $u_\theta(\mathbf{x}, t)$ is trained, it defines a velocity field over space and time. The sampling process amounts to integrating the learned ODE from a noise sample to a data sample:

$$\mathbf{x}_{t-1} = \mathbf{x}_t + u_\theta(\mathbf{x}_t, t) dt \quad (17)$$

Solvers like Euler method and Runge-Kutta can be further used to numerically integrate the ODE for higher speed and accuracy.

E THEORETICAL FORMULATIONS OF SO(3) ALIGNMENT

Different from Euclidean space, SO(3) is a Riemannian manifold. Therefore, the diffusion process (or probability flow) on SO(3) has been the focus of previous studies, requiring thoughtful design. Here we provide the formulations of rotation alignment in the SO(3) space, which is implemented from different perspectives(DDPM, Score Matching and Flow Matching) in Section 7.

E.1 PRELIMINARIES

The exponential and logarithmic maps. Generally speaking, the exponential and logarithmic relate the elements in Lie Group (rotation matrices) to the ones in Lie Algebra (skew-symmetric matrices). The skew-symmetric matrices in Lie Algebra can be specified with a rotation vector Φ :

$$\hat{\Phi} = \begin{bmatrix} 0 & z & -y \\ -z & 0 & x \\ y & -x & 0 \end{bmatrix}, \text{ where } \Phi = (x, y, z). \quad (18)$$

The magnitude of this vector $\omega = \|\Phi\|$ represents the angle of rotation, and its direction $\mathbf{n} = \Phi/\omega$ denotes the axis of rotation. Following the Rodrigues formula, the exponential map (rotation vector Φ to rotation matrix r) can be simplified to a closed form:

$$r = \exp(\hat{\Phi}) = \cos(\omega)I + \sin(\omega)\hat{\mathbf{n}} + (1 - \cos(\omega))\mathbf{nn}^T \quad (19)$$

Similarly, the matrix logarithm can be expressed using the rotation angle:

$$\hat{\Phi} = \log(r) = \frac{\omega}{2\sin\omega}(r - r^T), \quad \omega = \arccos[(\text{Trace}(r) - 1)/2] \quad (20)$$

810 **Sampling from IGSO3**(μ, ϵ^2) To form a rotation matrix $r \sim \text{IGSO3}(\mu, \epsilon^2)$, we first perform a
 811 random sampling from $\text{IGSO3}(\mathbf{I}, \epsilon^2)$. The axis of rotation \mathbf{n} is uniformly sampled, and the rotation
 812 angle θ is given by the following CDF (Nikolayev & Savyolov, 1997):
 813

$$814 f(\omega) = \sum_{\ell=0}^{\infty} (2\ell + 1) \exp(-\ell(\ell + 1)\epsilon^2) \frac{\sin((\ell + 1/2)\omega)}{\sin(\omega/2)}, \quad (21)$$

815 which together yield a rotation vector $\hat{\Phi} = \omega \mathbf{n}$. Then the rotation vector is shifted by the mean of the
 816 distribution to obtain $R = \mu \exp(\hat{\Phi})$ as the sampled rotation matrix.
 817

820 E.2 ROTATION ALIGNMENT BASED ON DDPM

821 Following the previous work (Leach et al., 2022), we implement a $\text{SO}(3)$ DDPM model based on
 822 several MLP layers for the rotation alignment in Section 7. With definitions described in Section
 823 E.1, the rotation matrix can be scaled by converting them into the Lie algebra (rotation vector),
 824 element-wise multiplying by scalar value, and converting back to rotation matrix through exponential
 825 map. The matrix scaling operation is defined as:
 826

$$827 \lambda(c, r) = \exp(c \log(r)), \quad (22)$$

828 where $\lambda(\dots)$ is the geodesic flow from \mathbf{I} to R by the amount c . Applying these to equations from the
 829 original DDPM model we arrive at the following definitions:
 830

$$831 q(r_t | r_0) = \text{IGSO3}(\lambda(\sqrt{\bar{\alpha}_t}, r_0), (1 - \bar{\alpha}_t));$$

$$832 p(r_{t-1} | r_t, r_0) = \text{IGSO3}(\tilde{\mu}(r_t r_0), \tilde{\beta}_t) \quad (23)$$

833 and

$$834 \tilde{\mu}(r_t, r_0) = \lambda\left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t}, r_0\right) \lambda\left(\frac{\sqrt{\bar{\alpha}_{t-1}}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}, r_t\right) \quad (24)$$

835 where β and α are schedule values in line with the description in Section D.1. To train the DDPM
 836 model $\epsilon_\theta(R_t, t)$ that predicts R_0 , the loss function is formulated as follows:
 837

$$838 L = \mathbb{E} \|\epsilon_\theta(r_t, t) r_0^T - I\|_F^2, \quad (25)$$

839 where $\|\cdot\|_F$ represents Frobenius norm.
 840

846 E.3 ROTATION ALIGNMENT BASED ON SCORE MATCHING

847 We abstract the mathematical principles behind score matching for $\text{SO}(3)$ from the previous work
 848 FrameDiff (Yim et al., 2023). For any $t \in [0, 1]$ and $r_0 \in \text{SO}(3)$, it assumes that $r_t \sim \text{IGSO3}(r_0, t)$,
 849 which is stated as the Brownian motion on $\text{SO}(3)$. To train a score-based model on $\text{SO}(3)$, we want
 850 the network $s_\theta(r_t, t)$ to approximate the score function $\nabla_{r_t} \log p(r_t | r_0)$:
 851

$$852 \nabla_{r_t} \log p(r_t | r_0) = \frac{r_t}{r_0^\top r_t} \log \{r_0^\top r_t\} \frac{\partial_\omega f(\omega(r_0^\top r_t), t)}{f(\omega(r_0^\top r_t), t)} \quad (26)$$

853 where $\omega(r)$ is the rotation angle in radians for any $r \in \text{SO}(3)$. In our toolkit, we use MSE loss to
 854 train the MLP network $s_\theta(r_t, t)$:
 855

$$856 L = \|s_\theta(r_t, t) - \nabla_{r_t} \log p(r_t | r_0)\|^2. \quad (27)$$

857 The reverse process from r_t to r_{t-1} can be formulated as:
 858

$$859 r_{t-1} = r_t [g(t)^2 s_\theta(r_t, t) dt + g(t) d\bar{\mathbf{w}}], \quad (28)$$

860 where $g(t)$ is the diffusion coefficient and $\bar{\mathbf{w}}$ is the time-reversal Wiener process.
 861

864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917

E.4 ROTATION ALIGNMENT BASED ON FLOW MATCHING

According to the method described in FoldFlow (Bose et al., 2024), we implemented the SO(3) alignment process based on Flow Matching in our toolkit. Given two rotation matrices $r_0, r_1 \in \text{SO}(3)$, the geodesic interpolation index by t has the following form:

$$r_t = r_0 \cdot \exp[t \cdot \log(r_0^\top r_1)] \quad (29)$$

Flow Matching over SO(3) aims to learn a time-dependent velocity field $u_\theta(\mathbf{r}, t)$ such that the solution to the following ordinary differential equation (ODE):

$$\frac{d\mathbf{r}}{dt} = u_\theta(\mathbf{r}, t) \quad (30)$$

FoldFlow takes $\log(r_t^\top r_0)$ divided by t as the target velocity at r_t . Thus we train the velocity field (MLP layers) $u_\theta(\mathbf{r}, t)$ with the following loss function:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{r}_0, \mathbf{r}_1, t} \|u_\theta(\mathbf{r}_t, t) - \log(r_t^\top r_0)/t\|_{\text{SO}(3)}^2 \quad (31)$$

where the distance induced by the $\|\cdot\|_{\text{SO}(3)}$ metric is given by:

$$d_{\text{SO}(3)}(r_0, r_1) = \|\log(r_0^\top r_1)\|_F, \quad (32)$$

With learned velocity field $u_\theta(\mathbf{r}, t)$, the reverse process can be written as:

$$r_{t-1} = r_t \cdot \exp[r_0^\top \cdot u_\theta(r_t, t)dt] \quad (33)$$