Amalga: Designable Protein Backbone Generation with Folding and Inverse Folding Guidance

Shugao Chen[1,2,*], **Ziyao Li**[1,3,*,†], **Xiangxiang Zeng**[2], **Guolin Ke**[1]
[1] DP Technology, Beijing, China
[2] College of Information Science and Engineering, Hunan University, China
[3] Center for Data Science, Peking University, China

Abstract

Recent advances in deep learning enable new approaches to protein design through inverse folding and backbone generation. However, backbone generators may produce structures that inverse folding struggles to identify sequences for, indicating designability issues. We propose Amalga, an inference-time technique that enhances designability of backbone generators. Amalga leverages folding and inverse folding models to guide backbone generation towards more designable conformations by incorporating "folded-from-inverse-folded" (FIF) structures. To generate FIF structures, possible sequences are predicted from step-wise predictions in the reverse diffusion and further folded into new backbones. Being intrinsically designable, the FIF structures guide the generated backbones to a more designable distribution. Experiments on both de novo design and motif-scaffolding demonstrate improved designability and diversity with Amalga on RFdiffusion.

1 Introduction

Rational protein design aims to create novel proteins or modify existing ones to obtain desired structures and functions. Accurate protein design methods enable direct applications such as enzyme engineering [10] and antibody-based drug design [15]. However, the vast combinatorial spaces of protein sequence and structure, along with their intricate interdependence, render this problem a longstanding challenge in biotechnology.

Fortunately, recent advances in deep learning illuminate new approaches to design proteins *de novo*. Capitalizing on abundant sequence and structure data, *inverse protein folding models* [7, 4] have succeeded in designing protein sequences that fold into specified target structures. Meanwhile, inspired by the formidable successes of diffusion models in image generation [6, 11], *diffusion-based backbone generators* [2, 13, 16, 17, 14] explore the prospects of generating novel protein backbone structures. The integration of these two methods outlines a pipeline to design proteins: 1) sample protein structures using backbone generators; 2) determine corresponding sequences with inverse folding models; 3) screen the generated proteins based on *designability* - how well the generated sequence folds into the accompanying structure; and 4) further screen the *designable* structures for desired applications, based on both sequences and structures.

While existing backbone generation models, as exemplified by RFdiffusion [14], produce backbones with sensible local structures and appropriate proportions of stable secondary structures (helices and sheets), inverse folding models struggle to identify sequences for a sizable proportion of the generated backbones, even when human evaluation deeming them reasonable. Quantitatively, RFdiffusion benchmarking indicates approximately 30% of samples did not satisfy the designability criterion. We reckon this issue arises due to two possible factors: 1) current protein folding and inverse folding

[*]This authors contributed equally to this paper. Shugao Chen contributed as an intern at DP Technology.
[†]Corresponding author: Ziyao Li <lizy01@dp.tech>

NeurIPS 2023 Generative AI and Biology Workshop.

models lack sufficient accuracy; 2) most existing backbone generation models are explicitly trained to reproduce structures alone, without capturing the intricate sequence-structure relationship which essentially depicts designability. We aim to address the second factor in this paper.

Here we propose Amalga, a simple yet effective inference-time technique to enhance the designability of diffusion-based backbone generators. By harnessing off-the-shelf folding and inverse folding models, Amalga guides backbone generation towards more designable conformations. Specifically, Amalga generates a set of "folded-from-inverse-folded" (FIF) structures by folding the sequences which are inverse folded from step-wise predicted backbones. These FIF structures, being inherently designable, are aligned to the predicted backbone and input into RFdiffusion's self-conditioning channel. Intuitively, this encourages RFdiffusion to match the distribution of designable structures. While retraining or finetuning RFdiffusion with FIF inputs may further improve performance, we demonstrate that Amalga significantly boosts designability when applied solely during inference.

2 Preliminaries

Diffusion-based Protein Backbone Generation. Recent works [1, 13, 17] have explored generating protein backbones using diffusion models such as denoising diffusion probabilistic model (DDPM) [6] and generative stochastic differential equations [12]. These generative models leverage forward and reverse diffusion processes to gradually transform samples from a simple prior distribution (often Gaussian) into complex backbone structures. The forward process perturbs the coordinates and orientations of each residue by adding noise with different scales on the timestep t. The reverse process then recovers high-quality backbones by iteratively predicting less noisy versions from the prior. Taking DDPM as an example, the forward and backward diffusion processes are formulated as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{2}$$

where $q(x_t|x_{t-1})$ perturbs x_{t-1} with Gaussian noise to obtain x_t, and p_θ predicts the reverse step result with neural networks based on the diffused sample x_t.

RFdiffusion. RFdiffusion [14] is a recent example of diffusion-based backbone generators. It finetunes RosseTTAFold [3], a multiple sequence alignment (MSA) based protein structure prediction model, with noised samples generated from the forward diffusion. Specifically, the structures of proteins to be designed are perturbed, and their corresponding sequences are masked. RFdiffusion also utilizes the original template channel in RosseTTAFold to input previously generated backbones into the model for self-conditioning. In this work, we additionally send the generated FIF samples through this channel to encourage the model to match the distribution of designable structures.

Designability Formulation. Given a folding model f and an inverse folding model f^{-1}, the general designability metric $\mathcal{D}(\mathbf{x})$ is defined as:

$$\mathcal{D}(\mathbf{x}) := \min_{\boldsymbol{s}\in\mathcal{S}} \|\mathbf{x} - f(\boldsymbol{s})\| \approx \|\mathbf{x} - f(f^{-1}(\mathbf{x}))\| \tag{3}$$

where $\boldsymbol{s}$ and $\mathbf{x}$ denote protein sequences and structures respectively, $\|\cdot\|$ quantifies the structural differences between two conformations (e.g. RMSD), and $\mathcal{S}$ represents the set of all feasible protein sequences. Conceptually, designability measures how accurately a structure $\mathbf{x}$ can be reproduced by its predicted sequence $f^{-1}(\mathbf{x})$ after folding, with lower values indicating higher designability. Notably, this metric depends on the accuracy of the folding and inverse folding models.

3 Method

Figure 1 illustrates the workflow of Amalga at each timestep t of the reverse diffusion process. We demonstrate Amalga on RFdiffusion, while the idea is broadly applicable to other baselines [17, 13]. The model takes as input the noised backbone $\mathbf{x}_{t+1}$ from the forward process, the predicted backbone $\hat{\mathbf{x}}_{t+1}$, and the FIF samples $\{\tilde{\mathbf{x}}^i_{t+1}\}^{N_{\text{FIF}}}_{i=1}$ generated in the previous step to make new backbone $\hat{\mathbf{x}}_t$. This predicted backbone, together with $\mathbf{x}_{t+1}$, is used to compute the noised input for the next timestep via the reverse diffusion formula. Amalga then inverse folds the prediction $\hat{\mathbf{x}}_t$ using ESM-IF [7] to generate possible sequences $\{\boldsymbol{s}^i_t\}^{N_{\text{FIF}}}_{i=1}$. These sequences are then folded using ESMFold [9] to obtain new FIF backbones $\{\tilde{\mathbf{x}}^i_t\}^{N_{\text{FIF}}}_{i=1}$ to guide the next step. Note that $\hat{\mathbf{x}}_t$ is directly produced by the backbone

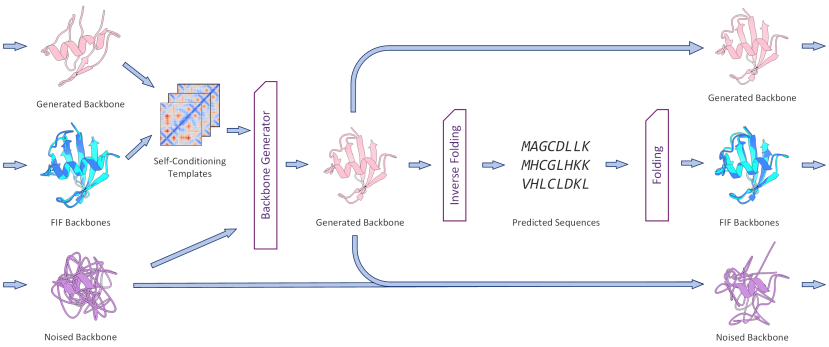

Figure 1: Amalga pipeline in each step. Amalga generates FIF samples from step-wise predicted backbones and inputs them to the model via the self-conditioning channel.

generator, thus assumed to have more reasonable structures and sensible inverse folding results, while $\mathbf{x}_t$ is the intermediate result in the diffusion process, thus with discontinuities.

In an attempt to further guide the model towards predicting foldable protein structures, we also explored providing the predicted sequences ($\{\boldsymbol{s}_t^i\}$) as inputs to the sequence channel of RFdiffusion. As the sequence channel in the original RosseTTAFold architecture leveraged multiple sequence alignments (MSAs) to inform structural predictions, we hypothesized that explicitly providing these inverse-folded sequences could similarly enhance folding precision. However, as is shown in the next section, experiments inserting $\{\boldsymbol{s}_t^i\}$ did not always result in improved performance. We believe the model likely requires a finetuning stage in order to integrate the sequences more properly.

In fact, the FIF samples directly inherit zero designability error, since their sequences are known to fold into the generated structures (up to the error of the folding model). As such, they are already successful design outcomes for unconstrained, *de novo* generation. However, these FIF structures may not conform to the motif constraints specified by the user. In contrast, the final output structure from the reverse diffusion process will explicitly satisfy the desired motifs, since they are fixed during this process. Therefore, Amalga balances global designability, provided by guiding the diffusion model with the FIF samples, and precise motif reconstruction in the final output.

4 Experiments

Settings. We conducted comparative experiments between the original RFdiffusion model and the RFdiffusion model augmented with Amalga. We utilized ProteinMPNN [4] following RFdiffusion for inverse folding, however, we replaced AlphaFold [8] with ESMFold to fold the final structures, as ESMFold achieves superior performance when multiple sequence alignments are unavailable. This replacement did not significantly altered the results, as we have analyzed in the appendix. For Amalga, we tested settings with $N_{\text{FIF}} = 1, 5$. We reported results for two Amalga variants: one where we input predicted sequences via the MSA channel (denoted "+seq"), and one where we did not.

We evaluated two backbone generation task schemes: 1) de novo design, in which backbones are generated without external constraints, and 2) motif-scaffolding, in which backbones should contain a predefined motif with known sequence and structure. In the former task, we generated 20 structures of lengths 100, 150, 200, 250 and 300, respectively. In the latter task, we generated 100 structures for each of the 25 benchmark tasks in RFdiffusion. We use the root-mean-square deviation of self-consistency (scRMSD) and the in silico success rate to depict designability. The scRMSD measures the error between a generated structure and its closest foldable structure:

$$\text{scRMSD}(\hat{\mathbf{x}}) = \min_{\boldsymbol{s} \in \mathcal{F}^{-1}(\hat{\mathbf{x}})} \text{RMSD}(\hat{\mathbf{x}}, f(\boldsymbol{s})) \tag{4}$$

where $\mathcal{F}^{-1}(\hat{\mathbf{x}})$ denotes the set of 8 inverse-folded sequences of $\hat{\mathbf{x}}$. The in silico success criteria were adopted from RFdiffusion: for de novo design, an scRMSD below 2Å was required to be considered successful; for motif-scaffolding, an additional requirement was that the RMSD between the motif in the best design and the target motif be less than 1Å. We also report a metric of design diversity: all success backbones were clustered using MaxCluster [5] with a TM-score threshold of 0.5. The diversity was quantified as the number of the unique clusters in the generated success samples.

Table 1: Results of Amalga and the original RFdiffusion, averaged over all cases.

		De Novo			Motif-Scaffolding		
		% success	scRMSD	Diversity	% success	scRMSD	Diversity
RFdiffusion		86.00	**1.29**	15.60	69.28	1.24	14.24
Amalga	($N_{\text{FIF}} = 1$)	86.00	1.58	**16.00**	73.20	1.05	14.88
	($N_{\text{FIF}} = 1$, +seq)	**87.00**	1.30	15.40	74.24	**0.98**	14.24
	($N_{\text{FIF}} = 5$)	83.00	1.65	15.40	**75.64**	**0.98**	**16.56**
	($N_{\text{FIF}} = 5$, +seq)	83.00	1.43	15.80	70.84	1.03	16.44

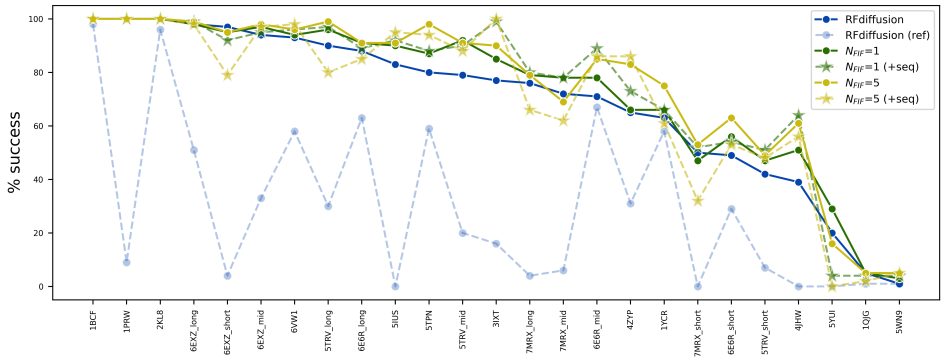

Figure 2: Success rate of 25 benchmark tasks in RFdiffusion. RFdiffusion (ref) in light blue displays statistics directly taken from [14], while RFdiffusion in dark blue shows reproduced statistics under our settings, comparable to Amalga. The columns are ranked by RFdiffusion performance.

Results. Table 1 shows metrics averaged on all 100 de novo generation samples and 2500 motif-scaffolding samples. Notably, results for motif-scaffolding are more informative, as in de novo task the FIF samples are already samples with zero scRMSD. Overall, Amalga obtained superior designability and diversity over RFdiffusion. Adding the sequence into the MSA channel (+seq) with $N_{\text{FIF}} = 1$ improves the performance, while the contradictory result holds with $N_{\text{FIF}} = 5$. We posit that the model needs further training to adapt to more inverse-folded sequences. We examine the specific success rate of 25 motif-scaffolding tasks in Figure 2. Amalga consistently outperforms RFdiffusion on the 25 benchmark cases with few exceptions. Notably, we observed significant improvements in the RFdiffusion performance over the originally reported. We posit the current release of RFdiffusion parameters have been refined since its publication. Results with regard to motif RMSDs, etc. are available in the appendix.

Efficiency. We analyzed the running time of Amalga to quantify the introduced complexity of FIF computation. To generate one 100 amino acid protein, the running time on a 32GB NVIDIA V100 GPU increased from $1'00''$ to $3'04''$ with $N_{\text{FIF}} = 1$ and $5'12''$ with $N_{\text{FIF}} = 5$. Overall, the introduced complexity is comparable to the original model's complexity.

5 Conclusion & Future Work

In this work, we have proposed Amalga as a broadly applicable inference-time technique to enhance the designability of diffusion-based backbone generators exemplified by RFdiffusion. Our experiments demonstrate that Amalga successfully improves the designability and diversity of generated structures from RFdiffusion, at the cost of additional inverse folding and folding computations. As a direct path for improvement, an obvious next step is to fine-tune RFdiffusion to better adapt it to Amalga inputs. Furthermore, inference speed could be enhanced by optimizing Amalga implementation, such as enabling batched ESMFold inference. Since predicted backbones may not vary drastically step-by-step, utilizing longer intervals between FIF evaluations leads to another gain of efficiency. For more rigorous validation, pending experimental conditions, we hope to perform wet lab experiments to further prove the effectiveness of Amalga designs. As ongoing work, we are also actively exploring adaptations of this approach to other existing protein design baselines.

References

[1] Namrata Anand and Tudor Achim. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv*, 2022.

[2] Namrata Anand and Possu Huang. Generative modeling for protein structures. *Advances in neural information processing systems*, 31, 2018.

[3] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

[4] Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022.

[5] Alex Herbert. MaxCluster - a tool for protein structure comparison and clustering. http://www.sbg.bio.ic.ac.uk/~maxcluster.

[6] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

[7] Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. *Proceedings of the 39th International Conference on Machine Learning*, 2022.

[8] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021.

[9] Yeqing Lin and Mohammed AlQuraishi. Generating Novel, Designable, and Diverse Protein Structures by Equivariantly Diffusing Oriented Residue Clouds. *Proceedings of the 40th International Conference on Machine Learning*, 2023.

[10] Stanislav Mazurenko, Zbynek Prokop, and Jiri Damborsky. Machine Learning in Enzyme Engineering. *ACS Catalysis*, 10(2):1210–1223, January 2020.

[11] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with CLIP latents. *arXiv*, 2204.06125, April 2022.

[12] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.

[13] Brian L. Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. *International Conference on Learning Representations*, 2206.04119, March 2023.

[14] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. De novo design of protein structure and function with RFdiffusion. *Nature*, July 2023.

[15] Wiktoria Wilman, Sonia Wróbel, Weronika Bielska, Piotr Deszynski, Paweł Dudzic, Igor Jaszczyszyn, Jędrzej Kaniewski, Jakub Młokosiewicz, Anahita Rouyan, Tadeusz Satława, Sandeep Kumar, Victor Greiff, and Konrad Krawczyk. Machine-designed biotherapeutics: Opportunities, feasibility and advantages of deep learning in computational antibody discovery. *Briefings in Bioinformatics*, 23(4):bbac267, July 2022.

[16] Kevin E. Wu, Kevin K. Yang, Rianne van den Berg, James Y. Zou, Alex X. Lu, and Ava P. Amini. Protein structure generation via folding diffusion. *arXiv*, 2209.15611, November 2022.

[17] Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. *arXiv*, 2302.02277, February 2023.

A Implementation Details

Our work implements the open-source code of RFdiffusion [3]. We use the default sampling settings from RFdiffusion, except where noted. The number of sampling steps is set to 50, and the C_α translation noise scalar is 1. For generating FIF samples, we leverage ESM-IF and ESMFold models [4] due to their state-of-the-art performance and efficiency. To evaluate generated sequences orthogonally, we follow the original RFdiffusion and use ProteinMPNN [5], replacing the protein folding model from AlphaFold with ESMFold for its superior single-sequence structure prediction.

B Effect of Folding Models

To examine whether the folding model used for evaluation impacts the final results, we generate 10 samples for each case and fold the same ProteinMPNN sequences with both ESMFold and Alphafold2. Using the same designability criteria as described in Section 4, Fig. 3 shows that the choice of folding model does not substantially influence the evaluation of designability.

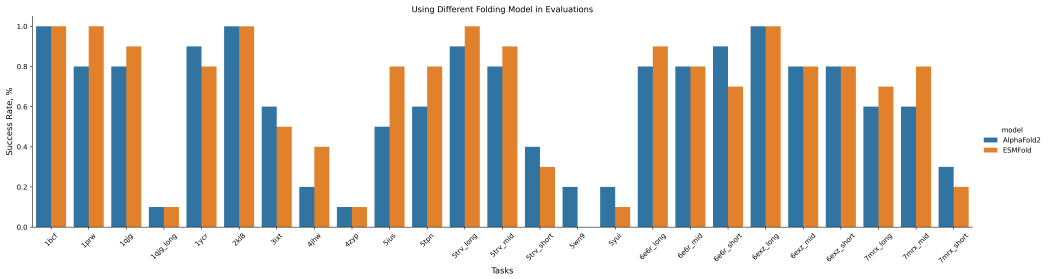

Figure 3: Designability metrics between AlphaFold and ESMFold.

C RMSD Variation in Sampling

We plot the step-wise RMSD and motif RMSD between the backbone generator output $\hat{\mathbf{x}}_t$ and the FIF samples $\{\tilde{\mathbf{x}}_t^i\}$ during the denoising process on two motif-scaffolding cases. Notably, as $\hat{\mathbf{x}}_t$ maintains the target motif almost identically, the reported motif RMSD reflects the deviation between FIF samples and the target motif. As shown in Figure 4, both RMSDs decrease consistently following the reverse process.

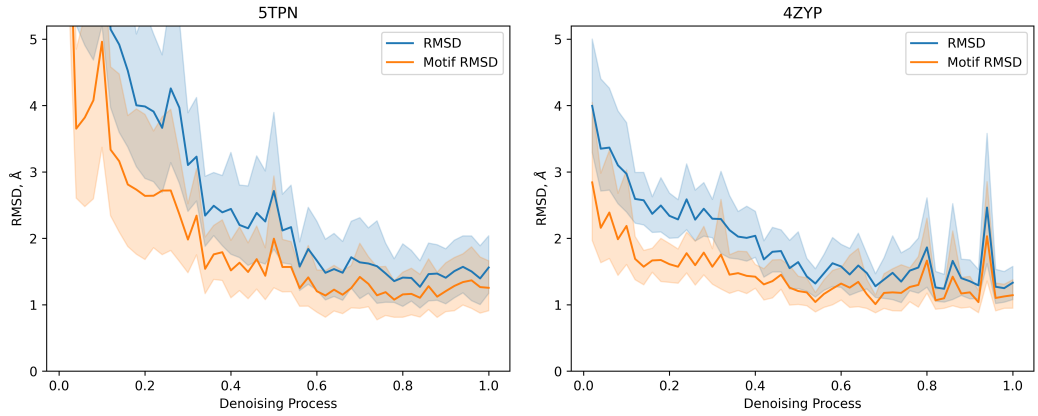

Figure 4: Step-wise RMSDs between FIF samples and generated backbones.

[3] https://github.com/RosettaCommons/RFdiffusion
[4] https://github.com/facebookresearch/esm
[5] https://github.com/dauparas/ProteinMPNN

Table 2: Detailed results of 25 benchmark tasks without additional sequences. MRMSD refers to the motif RMSD and SR refers to the success rate.

	RFdiffusion				$N_{\text{FIF}} = 1$				$N_{\text{FIF}} = 5$			
	RMSD	MRMSD	SR, %	Diversity	RMSD	MRMSD	SR, %	Diversity	RMSD	MRMSD	SR, %	Diversity
1BCF	0.44±0.07	0.36±0.07	100	1	0.42±0.06	0.34±0.07	100	1	0.41±0.05	0.34±0.06	100	1
1PRW	0.48±0.10	0.36±0.05	100	1	0.46±0.10	0.34±0.05	100	1	0.46±0.12	0.34±0.05	100	1
1QJG	1.08±1.53	1.03±1.66	5	47	0.82±0.87	0.77±0.70	5	48	0.79±1.10	0.81±1.04	5	45
1YCR	2.07±2.30	1.10±0.87	63	14	1.83±2.01	0.93±0.84	66	11	1.61±1.92	0.84±0.84	75	10
2KL8	0.45±0.05	0.48±0.05	100	1	0.40±0.04	0.41±0.03	100	1	0.40±0.03	0.41±0.03	100	1
3IXT	1.16±1.50	0.99±1.44	77	2	0.75±0.46	0.70±0.39	85	3	0.64±0.47	0.64±0.35	90	1
4JHW	1.49±1.23	1.33±0.80	39	1	1.24±0.86	1.13±0.54	51	1	1.17±0.80	1.11±0.65	61	2
4ZYP	1.22±0.62	0.97±0.45	65	4	1.02±0.48	0.91±0.43	66	3	0.85±0.39	0.77±0.37	83	2
5IUS	0.87±0.29	0.86±0.29	83	1	0.83±0.32	0.78±0.23	90	1	0.82±0.39	0.78±0.32	91	1
5TPN	0.85±0.36	0.78±0.32	80	2	0.69±0.27	0.62±0.24	87	2	0.62±0.22	0.55±0.19	98	2
5TRV_long	0.90±1.31	0.72±0.65	90	20	0.65±0.26	0.60±0.31	96	23	0.60±0.22	0.50±0.19	99	33
5TRV_mid	1.09±1.35	1.02±1.49	79	17	0.87±1.00	0.83±1.16	92	15	0.75±0.86	0.70±0.52	91	20
5TRV_short	2.31±2.12	1.92±1.77	42	4	2.03±1.91	1.69±1.49	47	4	1.91±1.73	1.69±1.44	49	5
5WN9	5.37±1.94	5.29±2.12	1	1	4.33±1.80	4.20±2.12	3	1	4.16±1.89	3.75±2.01	5	2
5YUI	1.99±1.30	1.88±1.02	20	12	1.81±1.55	1.70±1.41	29	16	1.61±1.39	1.50±1.08	16	16
6E6R_long	0.75±0.75	0.84±0.95	88	64	0.64±0.55	0.72±0.78	91	68	0.59±0.18	0.65±0.35	91	72
6E6R_mid	0.99±1.11	0.96±0.96	71	34	0.76±0.63	0.96±1.22	78	38	0.84±1.10	0.90±1.19	85	49
6E6R_short	1.92±2.00	1.87±1.87	49	16	1.52±1.61	1.43±1.42	56	15	1.56±1.90	1.58±1.93	63	23
6EXZ_long	0.54±0.19	0.45±0.24	98	33	0.51±0.20	0.44±0.17	98	34	0.47±0.11	0.45±0.18	99	37
6EXZ_mid	0.56±0.22	0.49±0.24	94	20	0.52±0.19	0.47±0.22	97	21	0.51±0.18	0.45±0.16	98	17
6EXZ_short	0.56±0.22	0.49±0.22	97	4	0.71±1.42	0.62±1.31	95	4	0.58±0.39	0.51±0.23	95	3
6VW1	0.67±0.30	0.59±0.29	93	1	0.60±0.32	0.53±0.32	94	1	0.56±0.23	0.50±0.22	96	1
7MRX_long	0.76±0.52	0.95±0.61	76	42	0.77±0.75	1.03±1.26	79	42	0.67±0.40	0.85±0.52	79	52
7MRX_mid	0.96±1.27	1.19±1.50	72	12	0.82±1.01	0.96±0.91	78	15	0.74±0.35	0.90±0.48	69	15
7MRX_short	1.44±1.34	1.58±1.50	50	2	1.25±1.17	1.43±1.21	47	3	1.28±1.40	1.49±1.38	53	3

Table 3: Detailed results of 25 benchmark tasks with additional sequences. MRMSD refers to the motif RMSD and SR refers to the success rate.

	RFdiffusion				$N_{\text{FIF}} = 1$ + seq				$N_{\text{FIF}} = 5$ + seq			
	RMSD	MRMSD	SR, %	Diversity	RMSD	MRMSD	SR, %	Diversity	RMSD	MRMSD	SR, %	Diversity
1BCF	0.44±0.07	0.36±0.07	100	1	0.43±0.07	0.35±0.08	100	1	0.42±0.07	0.35±0.07	100	1
1PRW	0.48±0.10	0.36±0.05	100	1	0.43±0.10	0.33±0.04	100	1	0.45±0.09	0.35±0.05	100	1
1QJG	1.08±1.53	1.03±1.66	5	47	0.83±0.78	0.84±0.84	4	44	1.06±1.64	0.98±1.40	2	45
1YCR	2.07±2.30	1.10±0.87	63	14	1.97±2.17	0.97±0.82	66	11	2.06±2.29	1.04±0.99	61	11
2KL8	0.45±0.05	0.48±0.05	100	1	0.38±0.04	0.37±0.04	100	1	0.40±0.05	0.38±0.04	100	1
3IXT	1.16±1.50	0.99±1.44	77	2	0.56±0.22	0.53±0.19	99	1	0.50±0.13	0.47±0.12	100	1
4JHW	1.49±1.23	1.33±0.80	39	1	1.16±0.88	1.04±0.60	64	3	0.89±0.46	0.89±0.41	056	6
4ZYP	1.22±0.62	0.97±0.45	65	4	0.91±0.45	0.83±0.39	73	2	0.80±0.52	0.75±0.41	86	2
5IUS	0.87±0.29	0.86±0.29	83	1	0.73±0.30	0.70±0.26	92	1	0.70±0.34	0.65±0.25	95	1
5TPN	0.85±0.36	0.78±0.32	80	2	0.73±0.34	0.64±0.35	88	2	0.62±0.36	0.54±0.32	94	4
5TRV_long	0.90±1.31	0.72±0.65	90	20	0.61±0.27	0.51±0.22	97	24	0.66±0.61	0.56±0.42	80	36
5TRV_mid	1.09±1.35	1.02±1.49	79	17	0.80±1.01	0.75±0.94	90	17	0.64±0.53	0.63±0.77	88	19
5TRV_short	2.31±2.12	1.92±1.77	42	4	1.69±1.44	1.46±1.26	51	4	2.11±2.10	2.08±2.07	48	4
5WN9	5.37±1.94	5.29±2.12	1	1	4.21±1.83	3.92±1.84	4	2	3.80±2.12	3.17±1.64	5	3
5YUI	1.99±1.30	1.88±1.02	20	12	1.49±0.99	1.44±0.86	4	9	1.33±1.57	1.18±1.29	0	15
6E6R_long	0.75±0.75	0.84±0.95	88	64	0.83±1.95	0.79±1.28	89	66	0.76±0.79	0.75±0.69	85	71
6E6R_mid	0.99±1.11	0.96±0.96	71	34	0.62±0.43	0.67±0.46	89	37	0.68±0.48	0.69±0.56	86	42
6E6R_short	1.92±2.00	1.87±1.87	49	16	1.82±2.03	1.83±2.05	54	13	1.96±2.81	1.83±2.13	53	13
6EXZ_long	0.54±0.19	0.45±0.24	98	33	0.52±0.13	0.44±0.16	99	38	0.51±0.27	0.45±0.25	98	37
6EXZ_mid	0.56±0.22	0.49±0.24	94	20	0.54±0.19	0.49±0.21	95	17	0.56±0.47	0.49±0.49	97	19
6EXZ_short	0.56±0.22	0.49±0.22	97	4	0.70±0.59	0.58±0.33	92	5	1.03±1.15	0.78±0.60	79	2
6VW1	0.67±0.30	0.59±0.29	93	1	0.55±0.23	0.48±0.22	96	1	0.54±0.22	0.47±0.20	98	1
7MRX_long	0.76±0.52	0.95±0.61	76	42	0.62±0.18	0.81±0.40	80	42	0.97±1.29	1.05±0.77	66	50
7MRX_mid	0.96±1.27	1.19±1.50	72	12	0.69±0.26	0.87±0.48	78	11	0.94±1.04	1.07±0.76	62	22
7MRX_short	1.44±1.34	1.58±1.50	50	2	1.15±0.84	1.31±0.98	52	3	1.45±0.98	1.54±0.84	32	4