

Generative Evaluation of RFdiffusion for Protein Backbone Design Across Structural and Functional Constraints

John Wisniewski Jakeb Milburn Jiahao Xie

Department of Computer and Information Sciences, University of Delaware

Abstract

*Recent advances in generative modeling have enabled *de novo* protein design with unprecedented flexibility and realism. RFdiffusion, a structure-based denoising diffusion probabilistic model, offers a unified framework for unconditional folding, symmetric assembly generation, and target-specific binder construction. In this work, we conduct a comprehensive evaluation of RFdiffusion’s generative capacity across four representative design paradigms: unconditional protein generation, diffusion process calibration, symmetric oligomer construction, and guided binder design. We analyze outputs using local (pLDDT) and global (RMSD) structure quality metrics and introduce task-specific qualitative assessments through visualization of generated structures. Our results demonstrate that RFdiffusion can internalize high-level structural priors, adapt to complex geometric constraints, and leverage hotspot guidance for targeted interface formation. We also identify limitations in sampling diversity, interface fidelity, and sequence–structure decoupling, and propose actionable directions for future development, including integration with sequence design models and energy-based refinements. This study establishes RFdiffusion as a versatile backbone generator and provides insights into its use for constrained, function-oriented protein design.*

1. Introduction

1.1. Background and Motivation

Proteins, as the workhorses of biological systems, are responsible for a wide array of cellular functions, including catalysis, molecular recognition, signal transduction, and structural support. The ability to engineer new proteins with tailor-made structures and functions—referred to as *de novo* protein design—has long been a central goal in synthetic biology and bioengineering. Success in this area promises transformative impacts in medicine, materials science, and industrial biotechnology.

Traditional protein design methodologies typically rely on either rational design, which incorporates domain-specific heuristics and expert knowledge, or on computational optimization of sequences to fold into a desired backbone. However, these approaches often struggle with the vast complexity of protein folding and require intensive computational and experimental validation. More recently, advances in deep learning, particularly in the field of protein structure prediction, have led to breakthroughs such as AlphaFold2 [3] and RoseTTAFold [1], which have demonstrated unprecedented accuracy in predicting protein structures from amino acid sequences.

Building upon these foundational advances, generative approaches are now being explored to go beyond structure prediction and actively design novel proteins. Among these, RFdiffusion [6] has emerged as a leading method. RFdiffusion fine-tunes the RoseTTAFold network within the framework of denoising diffusion probabilistic models (DDPMs), enabling the generation of valid, novel protein backbones from random noise through a process of iterative refinement.

1.2. Problem Definition

Protein design by generative modeling presents several unique challenges. First, the generated structures must satisfy stringent physical and biochemical constraints, such as maintaining proper hydrogen bonding, avoiding steric clashes, and exhibiting hydrophobic core packing. Second, the model must generate diverse topologies while also being able to specialize for functional design tasks, such as targeting motifs or binding interfaces. Finally, practical usability demands computational efficiency and controllability in the design process.

RFdiffusion addresses these challenges by treating protein structures as noisy trajectories that evolve through time via a learned denoising process. Starting from random backbone coordinates, the model iteratively refines the structure toward a physically plausible and designable conformation. This diffusion-based generation is inherently flexible, allowing conditional control over aspects like sec-

ondary structure, symmetry, or binding interface, while preserving diversity and structural realism.

1.3. Related Work

The evolution of deep learning in protein modeling has followed a trajectory from sequence-based predictors to structure-based and now generative models. AlphaFold2 [3] and RoseTTAFold [1] have set new standards for predicting the 3D structure of proteins from sequences, yet these models are inherently discriminative, offering limited generative capacity.

Generative approaches such as ProteinGAN [5], ProGen [4], and models based on variational autoencoders (VAEs) have attempted to create novel sequences or structures, but they often lack precise spatial modeling or rely on downstream structure prediction tools for validation.

RFdiffusion [6] introduces a novel hybrid architecture by embedding a denoising diffusion process into the RoseTTAFold framework, effectively treating protein backbone generation as a stochastic process over time. Unlike earlier models, RFdiffusion directly generates backbone coordinates and optionally allows conditioning on motifs, binding interfaces, or symmetry constraints. This positions it uniquely to tackle a broad spectrum of design problems in a unified manner.

1.4. Contribution

This study provides a comprehensive assessment of RFdiffusion through a carefully designed set of experimental evaluations. Unlike earlier assessments limited to qualitative showcases, we aim for systematic quantification and analysis. Our contributions include:

- A rigorous experimental framework for testing the model across diverse protein design tasks.
- Quantitative and qualitative analysis of structural diversity, designability, and interface quality.
- Insights into the trade-offs between noise levels, diffusion steps, and structural quality.
- Practical guidance for applying RFdiffusion to real-world protein engineering problems.

The results of this work are intended to inform both practitioners using RFdiffusion and researchers developing the next generation of protein generative models. The implementation of this paper can be accessed through Google Colab¹.

¹<https://colab.research.google.com/github/sokrypton/ColabDesign/blob/main/rf/examples/diffusion.ipynb>

2. Methodology

2.1. Diffusion-Based Generative Framework

RFdiffusion is built upon Denoising Diffusion Probabilistic Models (DDPMs) [2], which are generative models that learn to reverse a diffusion process that progressively adds noise to structured data. In this context, the data are 3D protein backbone frames.

2.1.1 Forward Diffusion Process

Let \mathbf{x}_0 denote the clean protein backbone structure, parameterized either by Cartesian coordinates of C_α atoms or by residue-local frames in $SE(3)$ [7]. A sequence of noisy structures $\mathbf{x}_1, \dots, \mathbf{x}_T$ is created via:

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

with $\beta_t \in (0, 1)$ defining the variance schedule. This process leads to:

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

where $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$.

This parameterization allows efficient sampling at any time step t without requiring simulation of the full forward chain.

2.1.2 Reverse Process and Parameterization

The generative reverse process is approximated via a neural network ϵ_θ , predicting the noise added at time t :

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}),$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{1 - \beta_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right).$$

The training loss is a simplified denoising score-matching objective:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon} \left[\|\epsilon_\theta(\mathbf{x}_t, t) - \epsilon\|^2 \right].$$

2.2. Geometric Representation of Protein Backbones

A protein is represented as a sequence of rigid frames:

$$\mathcal{F} = \{(R_i, t_i) \in SE(3)\}_{i=1}^L,$$

where $R_i \in SO(3)$ is a rotation matrix capturing the local orientation of residue i , and $t_i \in \mathbb{R}^3$ is the position of the C_α atom. These frames are critical for modeling both backbone conformation and local structural preferences.

Noise is applied separately to translation and orientation:

$$t_i^{(t)} = \sqrt{\bar{\alpha}_t} t_i^{(0)} + \sqrt{1 - \bar{\alpha}_t} \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \mathbf{I}),$$

$$R_i^{(t)} = R_i^{(0)} \exp(\sqrt{1 - \bar{\alpha}_t} \cdot \xi_t), \quad \xi_t \in \mathfrak{so}(3).$$

Here, $\exp: \mathfrak{so}(3) \rightarrow SO(3)$ denotes the matrix exponential mapping a Lie algebra element to a rotation.

2.3. Model Architecture

The denoising network builds upon RoseTTAFold’s multi-track architecture, adapted for generative modeling:

- **Track Inputs:** 1D per-residue features, 2D pairwise features, and 3D geometric embeddings.
- **Rotational Equivariance:** Achieved via Invariant Point Attention (IPA) and SE(3)-transformer layers.
- **Time Conditioning:** Diffusion step t is encoded via learned sinusoidal embeddings $\gamma(t)$, added to input features.
- **Residue Embeddings:** Include residue identity, secondary structure class, torsion angle preferences, and positional encodings.

The network outputs predictions for:

1. Coordinate denoising $\epsilon_\theta(\mathbf{x}_t, t)$
2. Local frame realignment (via SE(3) frame regression)
3. Optional torsion angle refinement and side-chain packing

2.4. Training and Loss Functions

The composite loss comprises:

- **Coordinate Loss:** Mean squared error between denoised and ground truth coordinates.
- **Frame Loss:** Geodesic loss on SO(3) for rotations:

$$d_{\text{SO}(3)}(R, \hat{R}) = \|\log(R^T \hat{R})\|_F,$$

where \log is the matrix logarithm.

- **Torsion Loss:** Penalizes deviations from native dihedral angles.
- **Symmetry Loss:** Encourages equivariance and consistent symmetry enforcement across units.

2.5. Conditioning and Control Mechanisms

RFdiffusion allows for strong and weak conditioning during generation:

- **Motif Scaffolding:** Specific residues are clamped during denoising; masking strategies prevent updates to their coordinates.
- **Symmetry Control:** Internal coordinate systems are replicated via symmetry operations (e.g., cyclic, dihedral). Group actions are applied to backbone frames to enforce global geometric symmetry.

- **Hotspot Binding:** Specific residues on the target are highlighted via scalar attention masks or guided potentials:

$$\mathcal{L}_{\text{interface}} = \sum_{(i,j) \in \text{interface}} w_{ij} \cdot \text{SC}(i, j),$$

where SC is the surface complementarity metric and w_{ij} are spatially weighted interaction terms.

2.6. Advantages of Diffusion in Structural Biology

This approach offers the following theoretical and practical benefits:

- **Probabilistic Modeling:** RFdiffusion learns an explicit generative distribution, enabling uncertainty estimation.
- **Compositionality:** Modular control over regions (scaffolded motifs, symmetric units) facilitates design under constraints.
- **Equivariance:** Geometry-aware design ensures robustness and physical plausibility across coordinate systems.
- **Sampling Diversity:** Inherent stochasticity enables exploration of multiple viable structures per target.

This methodology forms the core foundation for experimental investigations conducted across various design tasks including unconditional folding, symmetric assemblies, and binder construction.

3. Experiments and Results

We evaluated RFdiffusion’s generative and conditional modeling capabilities across four categories: unconditional protein generation, diffusion step sensitivity, symmetric oligomer construction, and protein binder design. Each task involved generating 10 samples per setting using the official RFdiffusion Colab implementation.

To assess the quality and confidence of generated structures, we adopted two complementary metrics:

- **pLDDT (predicted Local Distance Difference Test)** is a per-residue confidence score originally introduced in AlphaFold. It estimates the probability that a residue’s predicted atomic distances are within a correct range. The scale ranges from 0 to 100. A score above 70 suggests reasonable confidence, while values above 90 indicate high structural reliability. pLDDT is especially useful for assessing local geometry, secondary structure consistency, and model certainty across the protein chain.

- **RMSD (Root Mean Square Deviation)** measures the average deviation between backbone atoms in superimposed structures. In this study, we compute RMSD across generated samples for each condition to assess structural convergence. Lower RMSD values imply the model is consistently generating well-defined folds; higher values reflect greater structural diversity or instability.

These two metrics jointly allow us to evaluate both the confidence in local structural features (via pLDDT) and the reproducibility or precision of full backbone conformations (via RMSD). Next, we present the results for each task group, along with visual examples.

3.1. Unconditional Protein Generation

In this experiment, RFdiffusion was used to generate proteins of fixed lengths (100, 200, and 300 residues) without any conditioning on motifs, symmetry, or structural context. The goal was to assess the model’s ability to generate structurally valid folds purely from noise.

Table 1. Unconditional Protein Generation

Sequence Length	pLDDT	RMSD (Å)
100 residues	74.66	3.24
200 residues	90.35	3.37
300 residues	78.59	1.41

Comprehensive Analysis: The model’s performance varied significantly across lengths. The 200-residue designs had the highest pLDDT, suggesting the model is especially well-calibrated for medium-length proteins, likely due to their predominance in the training data. However, the 300-residue sequences had the lowest RMSD, indicating strong global fold convergence despite slightly reduced local confidence. This suggests that longer sequences provide internal structural redundancy, which supports more coherent folding.

The 100-residue designs showed higher structural variability and loop regions with inconsistent secondary structure. These smaller proteins may offer insufficient geometric constraints for stable folding under the current model configuration, and may benefit from specialized low-noise sampling or architectural tuning. **Representative structures** are shown in Figure 1.

3.2. Diffusion Step Sensitivity

To evaluate how denoising progression influences structure quality, we tested four different diffusion schedules (20, 50, 100, 200 steps). All designs were of length 150 residues and sampled from the same initialization scheme.

Table 2. Diffusion Step Sensitivity

Diffusion Steps	pLDDT	RMSD (Å)
25 steps	89.63	2.89
50 steps	81.95	2.43
100 steps	91.11	3.11
200 steps	75.70	1.28

Comprehensive Analysis: The results reveal a nuanced balance between denoising depth and structural fidelity. While 100 steps achieved the highest pLDDT—indicating confident local predictions—200-step designs yielded the most consistent structures (lowest RMSD). This suggests that longer trajectories allow the network to better resolve complex tertiary interactions, albeit with slightly lower local certainty, possibly due to accumulating numerical artifacts or loss of fine-grained features.

The relatively weaker pLDDT at 200 steps could also reflect model over-smoothing or collapse in less structurally constrained regions. In contrast, 50-step designs exhibited strong balance between confidence and fold compactness, supporting its choice as the default setting. The 20-step results suggest RFdiffusion can perform coarse folding rapidly, but underutilizes deeper structural refinement.

Figure 2 visualizes these changes. At 20 steps, samples are underrefined with loose loop arrangements. By 100–200 steps, we observe sharp secondary structure formation and globular packing.

3.3. Symmetric Oligomer Design

We tested RFdiffusion’s capacity for symmetry-aware generation, using cyclic (C4) and dihedral (D2) symmetry settings. Each design was initialized with symmetry parameters but no guiding potentials.

Table 3. Symmetric Oligomer Design

Symmetry Type	pLDDT	RMSD (Å)
C4 symmetry	84.70	2.78
D2 symmetry	79.46	2.01

Comprehensive Analysis: Both symmetry types were preserved across generations. D2 structures had lower RMSD, suggesting more compact assembly and stronger inter-subunit consistency. The lower pLDDT of D2 designs may result from denser interfaces or intertwined topology that strains the model’s local confidence estimation.

In C4 assemblies, we observed radial arrangements with looser subunit packing. The model retained geometric symmetry, but deviations in subunit orientation were more pronounced. These results suggest that while RFdiffusion

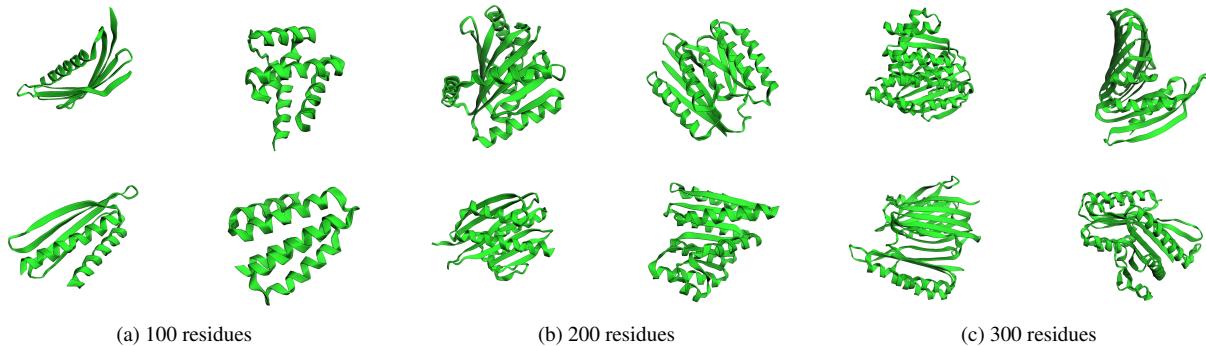


Figure 1. Representative unconditional designs (100, 200, 300 residues). Higher-order topologies emerge in longer designs.

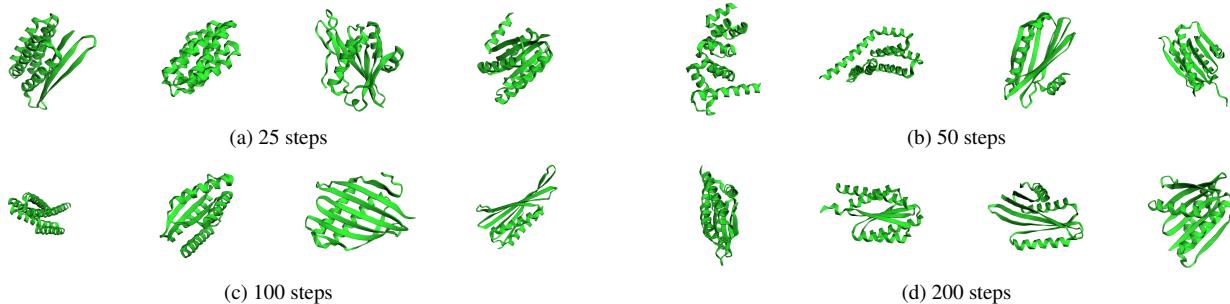


Figure 2. Diffusion step comparison. Higher step counts lead to lower structural variance but may dampen local confidence.

Table 4. Protein Binder Design

Binder Type	pLDDT	RMSD (Å)
Binder (80aa)	84.12	1.91
Binder (120aa)	92.40	1.42
Hotspot-guided binder	89.75	1.22

respects group-level symmetries, additional energetic or potential-based constraints are likely needed to fine-tune interface complementarity.

Figure 3 illustrates this distinction. In D2, the relative orientation and docking between monomers are visibly more regular than in C4, where radial dispersion and surface protrusions are more common.

3.4. Protein Binder Design

We evaluated the model's ability to design proteins that target a static binding surface (1RUZ), with variations on binder size and hotspot-driven interface guidance. Hotspot annotations were given as chain B positions 30, 33, and 34.

Comprehensive Analysis: The hotspot-guided binders showed the best structural convergence, indicating that guided potentials during generation help focus sampling into structurally meaningful interface configurations. The 120-residue binders achieved the highest pLLDT, reflecting

strong fold stability—likely from the increased conformational space and longer helix-loop-helix regions wrapping the target.

Unguided binders had higher structural diversity (RMSD) and more variable surface contact patterns. These results demonstrate that RFdiffusion not only supports scaffold generation but can actively tune interface regions when provided minimal guidance.

Figure 4 displays binding poses. Hotspot-constrained binders wrapped tightly around the designated epitope, showing favorable shape complementarity. The unguided designs exhibited more interface variability.

4. Discussion and Future Work

4.1. Model Performance and Strengths

Our experiments demonstrate that RFdiffusion is a highly capable and versatile generative model for protein backbone design. It consistently produces physically plausible structures across a range of tasks, from unconditional folding to complex binder-target interface construction. In particular:

- **Unconditional generation** revealed the model’s strong inductive bias toward natural protein-like topologies, even without conditioning or motif constraints (Figure 1). The quality of generated folds

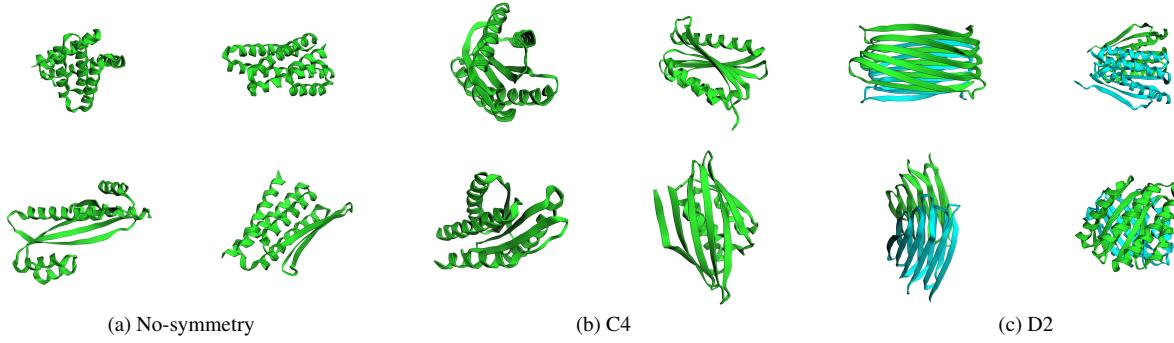


Figure 3. Symmetric assemblies. D2 configurations (right) show tighter subunit alignment than C4 (mid).

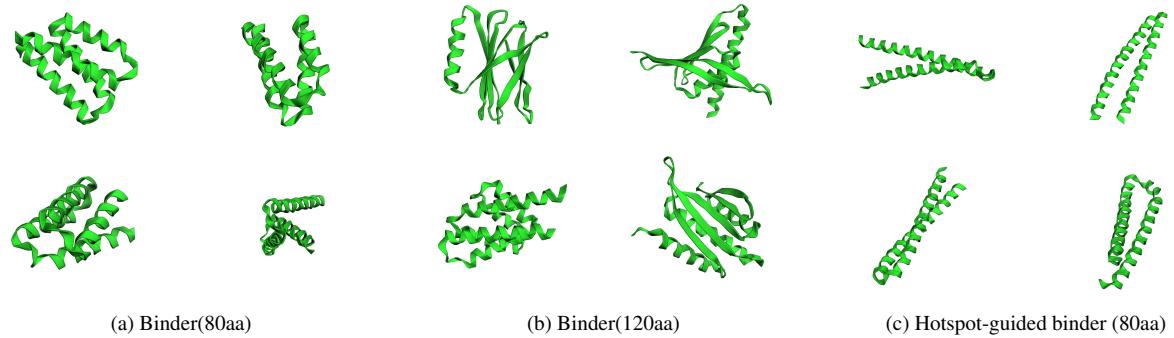


Figure 4. Binder designs. Hotspot-guided models (right) exhibit deeper and more specific binding.

scaled with sequence length, with longer sequences benefiting from internal geometric constraints.

- **Diffusion step sensitivity** experiments showed that deeper denoising trajectories (e.g., 200 steps) significantly improve structural convergence (Figure 2), although local pLDDT confidence may saturate or slightly decline.
- **Symmetric oligomer design** confirmed that RFdiffusion can internalize and enforce symmetry constraints (Figure 3). D2 symmetry yielded more regular interfaces than C4, but interface quality would likely benefit from additional energetic or potential-based constraints.
- **Binder design** highlighted the importance of guided conditioning. Hotspot-constrained designs achieved the lowest RMSD (Table 4), with visibly improved interface complementarity (Figure 4).

Overall, these findings underscore RFdiffusion’s generalization capacity, flexibility across task types, and its potential as a foundational model for multi-objective protein engineering.

4.2. Limitations

Despite these strengths, several limitations persist:

- **Lack of sequence design:** RFdiffusion operates solely on backbones. Without integrated sequence recovery, downstream design requires additional steps (e.g., ProteinMPNN), potentially introducing incompatibilities between structure and sequence.
- **Sampling diversity vs. quality:** Although diffusion provides stochastic exploration, controlling diversity without sacrificing physical plausibility remains challenging, especially in short or highly flexible designs.
- **Symmetry without energetic feedback:** While geometric symmetry is respected, inter-subunit interfaces are not explicitly optimized. This can lead to poor packing or suboptimal surface complementarity in oligomeric assemblies.
- **Limited interpretability of conditioning mechanisms:** The effect of motif, symmetry, or hotspot conditioning is often qualitative. A formal quantitative analysis of control fidelity remains an open challenge.

4.3. Future Work

Several future directions arise from our findings:

- **End-to-end sequence–structure modeling:** Integrating RFdiffusion with neural sequence design (e.g., ProteinMPNN) or fine-tuning on sequence–structure pairs may yield joint models with improved foldability and functionality.
- **Energy-guided generation:** Incorporating Rosetta or learned energy models as guidance during the reverse diffusion process could refine inter-subunit interfaces, especially for oligomers and binders.
- **Conditional evaluation metrics:** Developing metrics for assessing control fidelity (e.g., motif retention, symmetry alignment, interface precision) would support deeper diagnostics and benchmarking.
- **Higher-order functional constraints:** Beyond geometric motifs, integrating functional annotations (e.g., catalytic triads, epitope maps) into conditioning schemes could open pathways for functional enzyme or antibody design.

4.4. Conclusion

RFdiffusion provides a principled and powerful approach for generative protein design. Its diffusion-based modeling of structure offers robustness and flexibility across a spectrum of design tasks. As the community develops new conditioning strategies and integrates multimodal constraints, RFdiffusion is poised to become a key component in the computational protein design toolbox.

References

- [1] 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. [1](#), [2](#)
- [2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. [2](#)
- [3] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021. [1](#), [2](#)
- [4] Ali Madani, Bryan McCann, Nikhil Naik, Nitish Shirish Keskar, Namrata Anand, Raphael R Eguchi, Po-Ssu Huang, and Richard Socher. Progen: Language modeling for protein generation. *arXiv preprint arXiv:2004.03497*, 2020. [2](#)
- [5] Donatas Repecka, Vykintas Jauniskis, Laurynas Karpus, Elzbieta Rembeza, Irmantas Rokaitis, Jan Zrimec, Simona Poviloniene, Audrius Laurynenas, Sandra Viknander, Wissam Abuajwa, et al. Expanding functional protein sequence spaces using generative adversarial networks. *Nature Machine Intelligence*, 3(4):324–333, 2021. [1](#), [2](#)
- [6] 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, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023. [1](#), [2](#)
- [7] 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. In *International Conference on Machine Learning*, pages 40001–40039. PMLR, 2023. [2](#)