Enhancing Protein Design Robustness through Noise-Informed Sequence Design

Anonymous Authors[1]

Abstract

Recent advancements in protein design involve generating backbone structures first, followed by sequence design. Among these methods, one of the most popular is ProteinMPNN. A key limitation of ProteinMPNN is its inability to account for varying backbone quality at a per-position level, which is often problematic with structures containing both highly certain and relatively low-certainty regions. To address this, we propose introducing (1) a larger amount of Gaussian noise at a per-residue level and (2) labeling the amount of noise added to each residue as a new feature called a "noise label" to inform the model about backbone uncertainty or resolution. This enhancement improves sequence design success rates, as measured by the TM-score between the desired and predicted structures from the sequence. For partially redesigned scaffolds (i.e., motif scaffolding for enzymes or functional proteins), we introduce noise labels to the redesigned scaffolds while maintaining a fixed noise label of 0 for motif residues. This results in higher success rates for motif scaffolding structures, with reduced motif RMSD and overall structure RMSD. Incorporating residue-level noise label improves the designability of input protein backbones, as measured by correct prediction of the desired structure by AlphaFold2, from single sequence input.

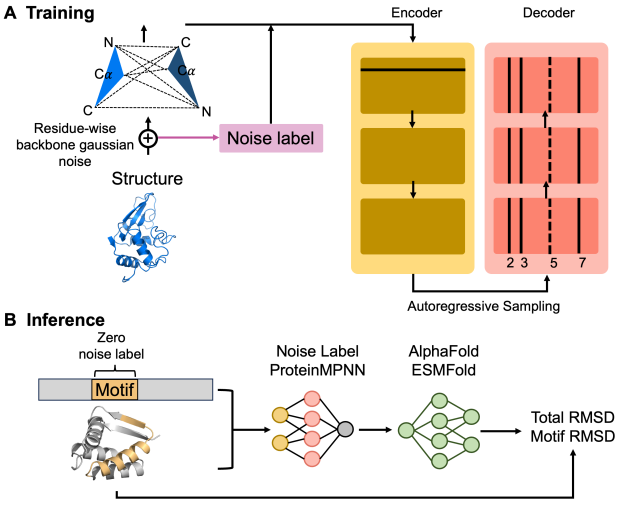

Figure 1. Overview of the protocol: (A) The model is trained to predict wild-type sequences given PDB structures with residue-wise Gaussian noise and noise labels as inputs. (B) For motif scaffolding, to maximize designiability of sequences, we set 0 Å noise labels for motif residues and set non-zero noise labels for scaffolding residues to reflect lower confidence of backbone.

1. Introduction

Recent advancements in de novo protein generation using deep learning have led to notable progress in addressing various challenges associated with generating diverse and functional proteins (Notin et al., 2024; Pan & Kortemme, 2021). One exemplary approach is diffusion models for protein design, exemplified by RFdiffusion and Chroma.(Watson et al., 2023; Ingraham et al., 2023). These models randomly initialize residue frames and refine the structure through iterative structure predictions, gradually reducing noise to achieve biophysically feasible structures. The resulting protein structures demonstrate highly diverse structures, showcasing substantial generalization capabilities beyond the domain of natural proteins.

Generative models have been widely applied to two protein design tasks: (i) unconditional, and (ii) conditional backbone generation. The unconditional regime generates novel and biologically feasible proteins (Yim et al., 2024). The conditional regime produces proteins conditioned on partial structural information or other functional properties. One common task, known as motif scaffolding (Wang et al., 2022), designs a backbone around a desired motif by fixing the coordinates of motif residues. This strategy is widely used in various fields, such as vaccine and enzyme design (Walls et al., 2020; Procko et al., 2014; Correia et al., 2015). Experimental validation has shown that motif scaffolding using diffusion models can generate high-quality structures

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

(Watson et al., 2023; Ingraham et al., 2023).

Once backbone structures are designed, sequence design models can generate amino acid sequences likely to fold into the provided structure. ProteinMPNN has been experimentally validated as a robust sequence design model that can be extended to de novo structures (Dauparas et al., 2022; Sumida et al., 2024). In sequence design models, "robustness" often refers to the ability to tolerate minor perturbations in atomic coordinates. This consideration is important when the protein backbone geometry is uncertain at atomic resolution (Dauparas et al., 2022). This robustness is achieved by adding Gaussian noise to backbone coordinates during training, which prevents the model from memorizing the input structures and designs sequences less sensitive to structural artifacts. In the training process of ProteinMPNN, backbone atoms are subjected to minor perturbations using Gaussian noise with a standard deviation of 0.1 Å to 0.3 Å. The model trained using 0.3 Å noise levels achieves the highest success rates, measured by RMSD between AlphaFold2 predicted structure and the desired structure.

However, challenges arise when utilizing ProteinMPNN for de novo designed proteins, particularly in large proteins and specific motif scaffolds. This becomes apparent as sequences designed by ProteinMPNN with RFdiffusion motif scaffolding structures can still result in low success rates for some designs (Watson et al., 2023). We hypothesize that this is due to the model being predominantly trained on experimentally determined crystal structures with a small amount of noise. This may not adequately capture the uncertainty arising from structures with a higher degree of noise. Thus, ProteinMPNN trained on uniform Gaussian noise may not be able to distinguish which regions of the structure are more uncertain than others. For example, in motif scaffolding, the coordinates of motifs from experimental results may be more accurate than the surrounding backbone structure generated by the model.

To address the challenge of sequence design robustness, we propose a novel strategy incorporating per-position labeling noise during training, with the amount of noise included as an additional feature. As illustrated in Figure 1, our approach integrates noise labels into the training process unlike the original ProteinMPNN, which is trained using uniformly distributed Gaussian noise on every atom without noise labels. During inference, instead of adding actual noise to the backbone, we introduce noise labels to modulate the resolution of the backbone at individual positions, thereby enhancing sequence designability and diversity. We can decrease or eliminate noise from specific motifs, ensuring the fidelity of crucial structural elements.

Our primary contributions are:

1. Enhancing the model's ability to sample better sequences for the query structure by incorporating a noise label for each residue, thereby increasing the success rate of designs.

2. Facilitating the generation of partially fixed structures, such as motif scaffolding, by assigning different noise labels to the fixed motif and the freely generated parts.

2. Related Work

2.1. Generative models of protein backbones

RFdiffusion is a protein backbone design model achieved through fine-tuning the RoseTTAFold structure prediction network (Baek et al., 2021) on protein structure denoising tasks (Watson et al., 2023). It demonstrates high performance across various protein design tasks, including monomer, binder, oligomer, and active site design, showcasing its versatility and efficacy through experimental characterization of numerous designed protein assemblies and functions. For each frame representing a residue, the model updates it by moving in the predicted direction, introducing some level of noise to create input for the subsequent step. The type and magnitude of this noise, along with the size of the reverse step, are selected to ensure that the process of removing noise matches the distribution of the original noise.

Chroma employs a method that combines non-equilibrium reverse diffusion with equilibrating Langevin dynamics to design backbone structures by numerically integrating a stochastic differential equation (Ingraham et al., 2023). The result indicates that near-exact design often significantly improves one-shot refolding across AlphaFold and ESMFold. A diffusion time parameter (t) of 0.5 for generated structures yields robust refolding results. However, a value of 0 for t may be more appropriate for experimentally precise structures. Intermediate values can offer a useful balance between robustness and precision.

AlphaFold2 Hallucination By inverting the AlphaFold2 (AF2) structure prediction model, sequence generation is guided towards adopting a desired fold, through a process termed "hallucination" (Anishchenko et al., 2021). This process optimizes a randomly initialized amino acid sequence using a specified loss function and the gradient descent algorithm to find the desired sequence. Then, the method backpropagates through AF2 to generate sequences that refold into a target protein structure with high AF2 confidence pLDDT (per-residue model confidence score) and low pAE (predicted aligned error). Experimental evidence has validated that numerous generated proteins fold into their intended structures (Goverde et al., 2023). Though AF2 hallucination itself generates sequences, we redesigned it with ProteinMPNN to compare success rates based on backbone structures with other models.

2.2. Sequence designs Models

ProteinMPNN is a message-passing neural network designed for protein sequence design based on a given protein structure (Dauparas et al., 2022). Given a backbone protein structure, it utilizes an encoder that takes in node features representing residues and edge features, encompassing Euclidean distances and primary sequence space distances between residues. This encoder updates nodes and edges by gathering information from neighboring residues through message passing, with edges also undergoing updates. The decoder then decodes per-position amino acid residues in an autoregressive manner. Influenced by existing sequence and structural data, ProteinMPNN has been shown to enhance natural proteins' expression, stability, and functionality (Sumida et al., 2024).

3. Method

3.1. Incorporate per residue Gaussian noise labels

ProteinMPNN initializes node embeddings for every residue with a 128-dimensional zero vector. For the edge embedding, it computes the distance between backbone atoms, applying randomly sampled noise from a standard normal distribution to these atoms. Here, we re-trained ProteinMPNN with per-residue Gaussian noise, sampled from a normal distribution multiplied by a maximum noise label constant. The 128th dimension of each vector was replaced with the applied noise level. As shown in Algorithm 1, we apply the same noise to five atoms within a single residue rather than assigning different noise levels to each atom. We train the ProteinMPNN model with PDB multimer datasets with 48 nearest neighbors. For the baseline model, we train using the same noise-adding method but without noise labels. During validation, the label noise remains consistent with the training set, and we refrain from adding any backbone noise to the structure.

Algorithm 1 Modified Sequence Design with Label Noise

Input: X (backbone coordinates), η (noise label), $\epsilon_{\max}$ (maximum noise level)
if training **then**
  $r \sim \mathcal{N}(0,1)$
  $\mathbf{n} = \frac{r}{\|r\|}$
  $s = (\mathcal{N}(0,1) \cdot \epsilon_{\max})$
  $\mathcal{N}_{\text{scaled}} = \mathbf{n} \cdot s$
  $\eta = \text{reshape}(s)$
  $X = X + \mathcal{N}_{\text{scaled}}$
else if inference **then**
  $s = (\mathcal{N}(0,1) \cdot \epsilon_{\max})$
  $\eta = \text{reshape}(s)$
end if
Return X, η

4. Experiments

We analyze the effectiveness of our approach in the context of two sequence design tasks: 1) unconditional protein design and 2) motif scaffolding. For the baseline experiments, we compare with the v_48_030 model, which is the original ProteinMPNN model trained with 0.3 Å and 48 nearest neighbor residues.

4.1. Sequence Design on Unconditional Protein Design

To evaluate on a wide range of length scales, we sampled 20 unconditional structures from each length scale: 100, 200, 300, 400, 500, 600, 700, 800, and 1000, for each model - AlphaFold2 Hallucination, RFDiffusion, and Chroma. Then for each unconditional structure, we generate 8 sequences and refold them using AlphaFold2 and ESMFold based on a single-sequence prediction of the structure, allowing for three recycling steps. We obtained the TM score (Zhang & Skolnick, 2005) between the original design structure and the predicted structure, then plotted the average TM score. The backbone scaffolds generated from different models may embed varying amounts of noise or uncertainty. To calibrate the optimal range of noise for inference, we compute unconditional logits from the ProteinMPNN using different levels of constant noise labels, and calculate the cross-entropy score between the unconditional logits and sampled sequences 4, measures the dissimilarity between the model's predicted logits and the sampled sequences. By comparing the scores between a model trained without noise labels and one trained with noise labels, we can approximate the optimal range of noise labels that may reflect the model's certainty with the sampled sequences when sampled with a noise label. A noise level showing a lower cross-entropy score with designs may inherently indicate quality or resolution of the input backbone.

Table 1. Comparing sequence diversity across noise labels, no noise labels, and v_48_030 model for designs from AlphaFold2 Hallucination (AD), RFdiffusion (RFD), and Chroma.

	DIVERSITY (↑)		
METHOD	AD	RFD	CHROMA
TRAINED W/ NOISE LABEL	0.3975	0.4165	0.4183
TRAINED W/O NOISE LABEL	0.3474	0.4006	0.4096
V_48_030	0.3355	0.3853	0.3973

4.2. Sequence Design on Motif Scaffolding

We collect motif scaffold designs from the RFdiffusion paper (Watson et al., 2023). These designs maintain the fixed 3D structure of the motif while redesigning the scaffold with

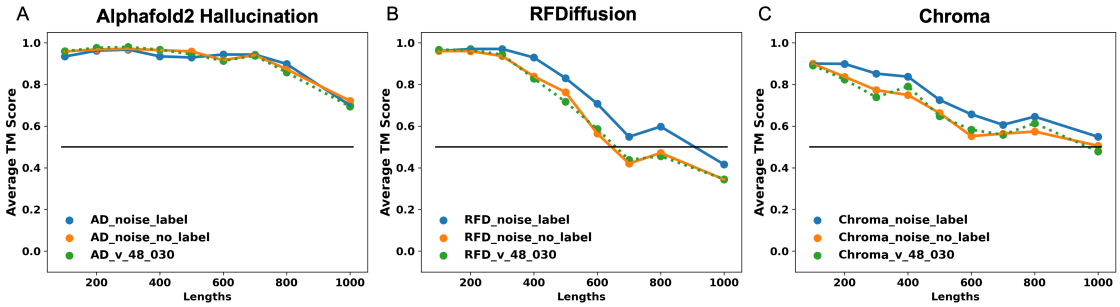

Figure 2. Average TM score of designs between the AlphaFold2-predicted structure and the input structure: (A) AlphaFold2 Hallucination, (B) RFDiffusion, (C) Chroma.

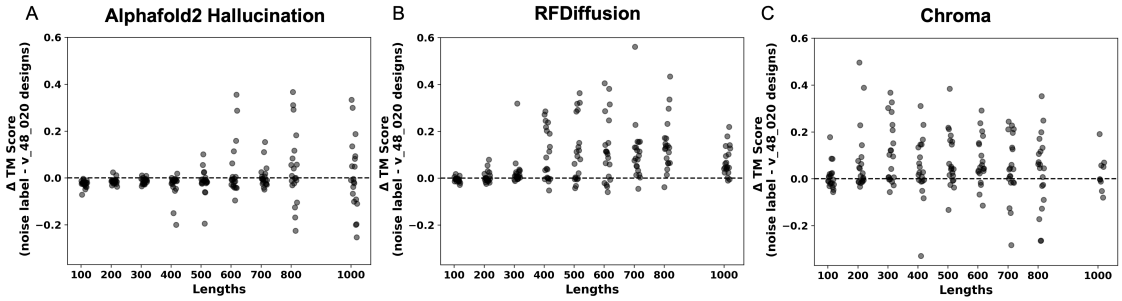

Figure 3. ΔTM score between sampled sequences from the noise label model and the v_48_030 model: (A) AlphaFold2 hallucination, (B) RFDiffusion, and (C) Chroma designs. For each backbone, 8 sequences are generated, and the maximum TM score is computed.

varied lengths, without fixing the motif's position within the proteins. Proteins are generated either by adding noise or without noise using reverse diffusion. In sequence design, we compare two scenarios: one where we add the same noise label to all positions, and another where we add the same noise label to all positions except the motif position, where we set the noise label to 0 to indicate that the region is highly certain. We refold the generated sequences using ESMFold to evaluate designability by metrics pLDDT and root mean squared distance (RMSD). The RMSD quantifies the disparity between the structure predicted by ESMFold and the original structure from which the sequence originated. Here, we evaluate either motif RMSD, total RMSD, or both between the predicted and input structures.

5. Results

5.1. Noise label model enhances refolding success

Figure 2 illustrates that the inclusion of noise labels during sequence sampling leads to a higher average TM score compared to the v_48_030 model. For each individual backbone, Δ TM score between sequences sampled by the noise label model and v_48_030 were frequently positive, implying that the sequences from the noise label model often have higher TM scores than those from v_48_030 (Figure 3). In the case of AF2 hallucination designs, the efficacy of noise labels is limited, possibly due to the high certainty of input structures; informing the model about noisy positions does not significantly enhance sequence sampling for better-designed sequences. However, for RFdiffusion and Chroma designs, moderate noise labels improve the average TM score of generated sequences, surpassing models without noise labels and the v_48_030 model. Furthermore, sequences generated from the noise label model show higher sequence diversity than the v_48_030 model. Sequence diversity is measured by the Levenshtein distance.(Berger et al., 2020) between sequences of the same length (Table 1).

To find the optimal noise label for each design, considering that designs sampled from different generating methods with varying quality, one strategy is to evaluate the cross-entropy between sequences sampled with varying noise labels, ranging from 0 to 1. Based on the ProteinMPNN cross-entropy between output logits and sampled sequences, this evaluation approximates the model's confidence for a sampled sequence. It helps determine the optimal noise label for specific design types. As shown in Figure 4, we evaluate the cross-entropy of all designs in Figure 2 with varied noise labels. AF2 hallucination designs achieve the lowest cross-entropy, indicating the highest confidence at a zero noise label. However, RFDiffusion and Chroma designs achieve the lowest cross-entropy between noise levels of 0 to 0.5, indicating that these generated backbones

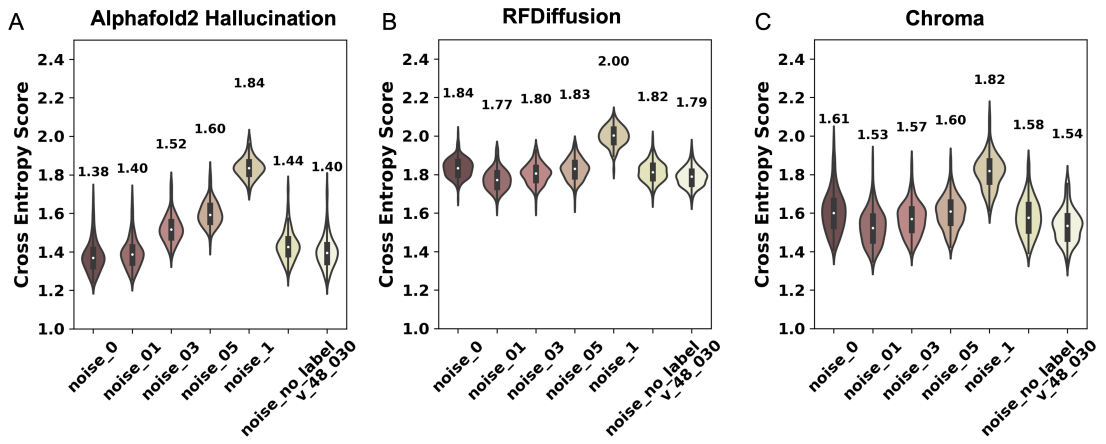

Figure 4. Cross-entropy scores between model-predicted logits and sampled sequences at varied noise inputs (A) AlphaFold2 hallucination, (B) RFDiffusion, and (C) Chroma designs.

may be of lower quality compared to those generated by AlphaFold2 hallucination. This may explain why we do not see a significant improvement by increasing the noise label for AF hallucinated backbones, compared to RFDiffusion and Chroma (Figure 2, 3).

5.2. Noise label model helps specify motifs in motif scaffolding designs

In Figure 5A, we plot the success rate of motif scaffolding designs. Compared to the success rate of the v_48_030 model (marked by an 'x'), sequences generated from the noise-labeled model achieved an overall higher success rate for the motif scaffolds. As shown in the right plot of Figure 5B, sequences generated by the noise label model with an input of motif noise label 0 achieved a higher success rate for the motif compared to the sequences generated by the model with a constant sequence recovery rate for all residues. For an overall scaffold RMSD < 2, the success rate of the noise model with zero motif noise shows that 60% of data points have a higher success rate than noise applied to all residues. Notably, for motif RMSD < 1, 85.0% of data points have higher success rates. It supports our hypothesis that assigning zero noise labels to the motif residues helps to increase precise sampling around the motif while maintaining overall robustness. Adding label noise to the entire structure still helps in sampling better sequences compared to the v_48_030 model and zero noise label models.

6. Conclusion

In summary, our study presents a robust sequence design method that incorporates a noise label during training to inform the model about backbone certainty. This integration significantly improves the success rate of sequence design, as evidenced by the TM score between input and refolded structures. Our experiments demonstrate that including noise labels during inference yields the highest success rates, particularly in backbone generation using hallucination and diffusion models. Additionally, in motif scaffolding designs, we introduce noise selectively to the backbone while maintaining a fixed noise label of 0 for motif regions, ensuring motif certainty. This approach increases the success rate by reducing both the total backbone RMSD and the motif RMSD. We anticipate that our model will facilitate the generation of robust sequences for de novo protein backbones across various protein generation methods.

References

Anishchenko, I., Pellock, S. J., Chidyausiku, T. M., Ramelot, T. A., Ovchinnikov, S., Hao, J., Bafna, K., Norn, C., Kang, A., Bera, A. K., et al. De novo protein design by deep network hallucination. *Nature*, 600(7889):547–552, 2021.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.

Berger, B., Waterman, M. S., and Yu, Y. W. Levenshtein distance, sequence comparison and biological database search. *IEEE transactions on information theory*, 67(6): 3287–3294, 2020.

Correia, B., Bates, J., Loomis, R., Carrico, C., Jardine, J., Baker, D., Strong, R., Crowe, J., Johnson, P., and Schief, W. Proof of principle for epitope-focused vaccine design. *Protein Science*, 24:181–184, 2015.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I., Courbet, A., de Haas,

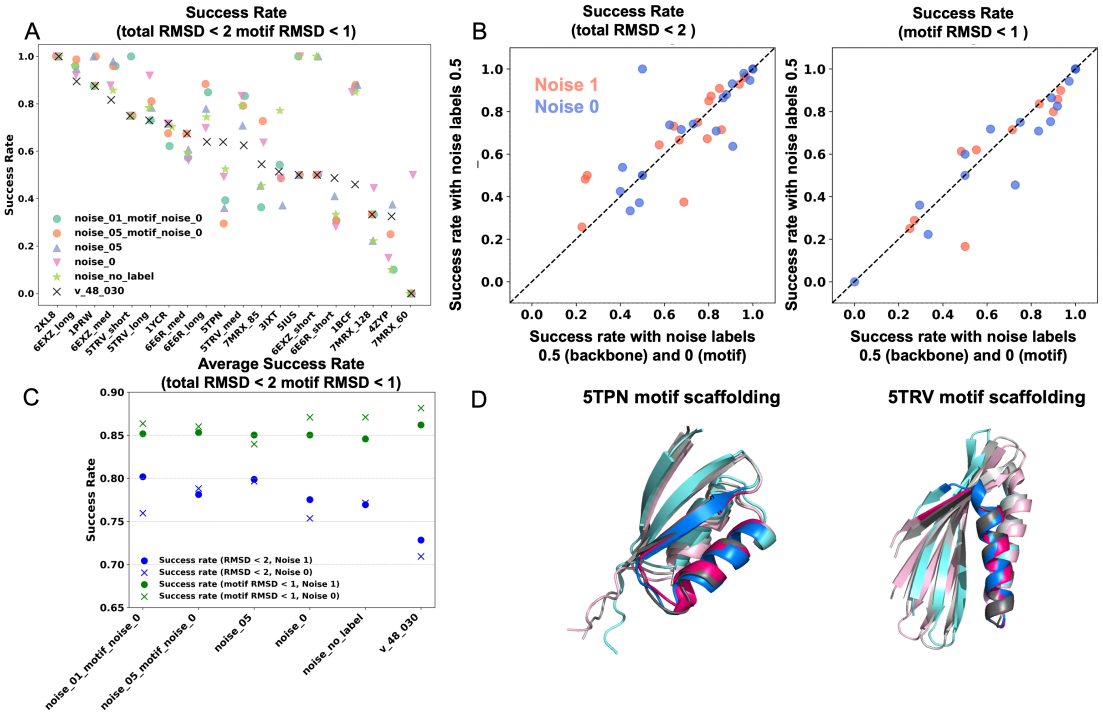

Figure 5. (A) Motif scaffolding success determined by RMSD thresholds: scaffold < 2, motif < 1. (B) Success rate comparison for two conditions: (1) 0 noise in motif, 0.5 in backbone; (2) 0.5 noise in all regions. Pink (Noise 1) includes noise in reverse diffusion, blue (Noise 0) has none. (C) Average success rates under different motif and overall noise levels. (D) Example designs: $5TPN\_noise\_1\_design\_18$ and $5TRV\_med\_noise\_0\_design\_8$, showcasing noise labels impact on refolded structures. Gray = RFdiffusion original design. Dark colors = motif. Pink = refolded sequence with 0 motif noise and 0.5 overall noise. Blue = refolded sequence from v_48_030.

R. J., Bethel, N., et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378 (6615):49–56, 2022.

Goverde, C. A., Wolf, B., Khakzad, H., Rosset, S., and Correia, B. E. De novo protein design by inversion of the alphafold structure prediction network. *Protein Science*, 32(6):e4653, 2023.

Ingraham, J. B., Baranov, M., Costello, Z., Barber, K. W., Wang, W., Ismail, A., Frappier, V., Lord, D. M., Ng-Thow-Hing, C., Van Vlack, E. R., et al. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, 2023.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Notin, P., Rollins, N., Gal, Y., Sander, C., and Marks, D. Machine learning for functional protein design. *Nature Biotechnology*, 42(2):216–228, 2024.

Pan, X. and Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *Journal of Biological Chemistry*, 296, 2021.

Procko, E., Berguig, G. Y., Shen, B. W., Song, Y., Frayo, S., Convertine, A. J., Margineantu, D., Booth, G., Correia, B. E., Cheng, Y., et al. A computationally designed inhibitor of an epstein-barr viral bcl-2 protein induces apoptosis in infected cells. *Cell*, 157(7):1644–1656, 2014.

Sumida, K. H., Núñez-Franco, R., Kalvet, I., Pellock, S. J., Wicky, B. I., Milles, L. F., Dauparas, J., Wang, J., Kipnis, Y., Jameson, N., et al. Improving protein expression, stability, and function with proteinmpnn. *Journal of the American Chemical Society*, 146(3):2054–2061, 2024.

Walls, A. C., Fiala, B., Schäfer, A., Wrenn, S., Pham, M. N., Murphy, M., Longping, V. T., Shehata, L., O'Connor, M. A., Chen, C., et al. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for sars-cov-2. *Cell*, 183(5):1367–1382, 2020.

Wang, J., Lisanza, S., Juergens, D., Tischer, D., Watson, J. L., Castro, K. M., Ragotte, R., Saragovi, A., Milles, L. F., Baek, M., et al. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein struc-

ture and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.

Yim, J., Stärk, H., Corso, G., Jing, B., Barzilay, R., and Jaakkola, T. S. Diffusion models in protein structure and docking. *WIREs Computational Molecular Science*, 14 (2):e1711, 2024. doi: https://doi.org/10.1002/wcms.1711. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1711.

Zhang, Y. and Skolnick, J. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic acids research*, 33(7):2302–2309, 2005.