

Controllable Protein Sequence Generation with LLM Preference Optimization

Xiangyu Liu¹, Yi Liu¹, Silei Chen^{2,3}, Wei Hu^{1,3,*}

¹ State Key Laboratory for Novel Software Technology, Nanjing University, China

² Medical School, Nanjing University, China

³ National Institute of Healthcare Data Science, Nanjing University, China

{xyl, yiliu07}.nju@gmail.com, {chensl, whu}@nju.edu.cn

Abstract

Designing proteins with specific attributes offers an important solution to address biomedical challenges. Pre-trained protein large language models (LLMs) have shown promising results on protein sequence generation. However, to control sequence generation for specific attributes, existing work still exhibits poor functionality and structural stability. In this paper, we propose a novel controllable protein design method called CtrlProt. We finetune a protein LLM with a new multi-listwise preference optimization strategy to improve generation quality and support multi-attribute controllable generation. Experiments demonstrate that CtrlProt can meet functionality and structural stability requirements effectively, achieving state-of-the-art performance in both single-attribute and multi-attribute protein sequence generation.

Introduction

The objective of protein design (Jumper et al. 2021; Zheng et al. 2023) is to create proteins with specific biochemical functions. Designing and exploring novel proteins offers a highly promising approach to addressing challenges in various fields, e.g., drug discovery (Śledź and Caflisch 2018; Cao et al. 2022), vaccine and enzyme design (Correia et al. 2014; Richter et al. 2011). Generating high-quality proteins that not only possess the desired functions but also exhibit structural stability has become increasingly important. Recently, several works based on deep models have achieved notable success in protein sequence generation (Ferruz, Schmidt, and Höcker 2022; Nijkamp et al. 2023) using protein large language models (LLMs). They leverage the similarities between protein sequences and natural language to generate biologically relevant and functional protein sequences with high accuracy.

Directly finetuning protein language models is an effective way to constrain their outputs to meet the specific function attributes (Luo et al. 2023; Nathansen et al. 2024). While some works (Wang et al. 2023; Fang et al. 2024) align protein space with natural language space to achieve controllable outputs, yielding promising results in protein sequence understanding (Zhuo et al. 2024), finetuning on downstream

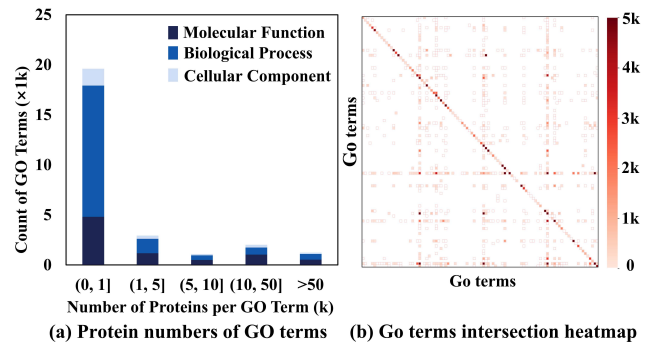


Figure 1: Gene Ontology (GO) terms analysis

tasks remains essential to ensure the generation of specific functional proteins (Lv et al. 2024).

Finetuning LLMs for controllable protein generation has two major challenges. First, existing works based on finetuning do not explicitly constrain structural stability and functionality during training. As a result, the finetuned model may still generate proteins that are functionally irrelevant or incapable of folding into stable structures. Second, the effectiveness of finetuning depends heavily on the quality and quantity of the dataset. We analyze the protein number of each term in Gene Ontology (GO) annotations¹, which provides descriptions of a protein’s molecular functions, biological processes, and cellular components. As shown in Fig. 1(a), most terms have data quantities concentrated below 5,000, and there are very few terms with a sufficient number of proteins. Furthermore, we randomly sample 100 terms and count the number of proteins with shared attributes between any two terms. The results in Fig. 1(b) show that fewer proteins meet both attributes, as indicated by the lighter color in the intersection of the heatmap. So, it is harder to construct sufficiently large data for finetuning when aiming to generate proteins with multiple attributes.

To address the issues mentioned above, we propose a preference optimization-based method to enhance the quality of controllable protein sequence generation. We design two metrics to assess the protein’s structural stability and functionality. For functionality, given the critical relationship be-

*Corresponding author

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹<https://current.geneontology.org/annotations/>

tween a protein’s structure and its function, we extract structural representations of sequences using a pre-trained structure encoder to assess their function similarity. For structural stability, we evaluate the conformational energy using the Rosetta energy score (Alford et al. 2017) to determine its stability. Moreover, we propose a multi-listwise preference optimization method, which fully considers the ranking information among a large number of candidate data based on the above two metrics. We introduce the ranking information as a regularization term to ensure that the optimization process focuses on data pairs with significant differences between preferred and non-preferred samples, thereby improving the optimization results. Specifically, we first employ prefix-tuning on the LLM to avoid the overfitting caused by limited data. Then, the model generates candidate data for evaluating and constructing the preference optimization dataset. Thus, preference optimization expands and augments the original finetuned dataset. We further extend our method to multi-attribute generation. By concatenating single-attribute prefixes and applying our preference optimization, we enhance the quality of multi-attribute generation without the need of multi-attribute supervision data.

We construct the training and evaluation datasets for six attributes based on GO annotations. We conduct comprehensive experiments with widely-used validation metrics. Experimental results demonstrate that our method outperforms baselines, achieving state-of-the-art results on both single-attribute and multi-attribute controllable generation.

In summary, the main contributions of this paper are listed as follows:

- We propose a preference optimization-based method CtrlProt to enhance the quality of controllable protein sequence generation, with optimization metrics designed from functionality and structural stability perspectives.
- We design a multi-listwise preference optimization method that leverages ranking information from large datasets, focusing on pairs with significant quality differences during training to improve the performance.
- Experiments show that CtrlProt performs well in both single-attribute and multi-attribute generation tasks. Additionally, the detailed analyses confirm the rationality and effectiveness of CtrlProt. Datasets and source code are available at <https://github.com/nju-websoft/CtrlProt>.

Related Work

Protein Language Models

Protein language models are widely used in protein sequence analysis, function prediction, and protein design. Encoder-based protein language models like ESM (Rives et al. 2021; Lin et al. 2023a), ProtBERT (Elnaggar et al. 2021) and ProtT5 (Elnaggar et al. 2021) are trained on large-scale protein sequence data and capable of capturing the semantic information and latent structural features within protein sequences. Decoder-based models, such as ProtGPT2 (Ferruz, Schmidt, and Höcker 2022) and ProGen2 (Nijkamp et al. 2023), have shown promising results in protein sequence generation. Recent works also

explore using diffusion models for generating protein sequences (Alamdari et al. 2023; Zongying et al. 2024).

The most straightforward method to controllable generation using protein language models is to finetune the model on downstream data (Luo et al. 2023; Nijkamp et al. 2023; Nathansen et al. 2024). It allows to generate proteins with specific functions. Some works aim to guide the model’s generation using natural language instructions. For example, recent studies (Fang et al. 2024; Wang et al. 2024, 2023) construct instruction-tuning datasets based on functional labels or descriptions to align models with natural language, achieving promising results in protein understanding but limited effectiveness and incomplete evaluation on sequence generation. ProLLaMA (Lv et al. 2024) incorporates continual learning of protein sequences before instruction-tuning, yielding better results on generation, but it still requires finetuning on downstream data to improve performance. However, finetuning does not explicitly focus on structural stability and functionality, which affects the quality of the generated sequences. To address it, we propose a preference optimization method to further enhance generation quality.

Preference Optimization

The reinforcement learning from human feedback (RLHF) framework has significantly improved model performance on downstream tasks by aligning with human preferences. Some researchers have shifted towards alternative methods to reduce the complexity of the RLHF framework. Direct preference optimization (DPO) (Rafailov et al. 2023) directly aligns the behavior of language models without using a reward model. ORPO (Hong, Lee, and Thorne 2024) and SimPO (Meng, Xia, and Chen 2024) further simplify the process by eliminating the reference model. Some works also focus on modeling sequence data. For example, PRO (Song et al. 2024) proposes a list maximum likelihood estimation loss for response lists. Lipo- λ (Liu et al. 2024) introduces a training weight to represent the ranking difference. However, they are unsuitable when generating a large number of protein sequences, which is costly for PRO and may lead to an overemphasis on top sequences for Lipo- λ . Additionally, some works (Meng, Xia, and Chen 2024; Park et al. 2024; Qin, Feng, and Yang 2024) explore the use of regularization terms as a margin to increase the reward difference between preferred and non-preferred data, thereby enhancing training effectiveness.

Preliminaries

RLHF. In the RLHF process (Ziegler et al. 2019; Liu et al. 2020), an LLM is first finetuned with instructions, resulting in model π_{ref} . The output of π_{ref} is sampled to obtain responses $y_1, y_2 \sim \pi_{ref}$. Humans then annotate and rank the quality of y_1 and y_2 , and create the preference optimization dataset of $D = \left\{ x^{(i)}, y_w^{(i)}, y_l^{(i)} \right\}_{i=1}^N$, where y_w and y_l denote the preferred and rejected responses, respectively. Based on the Bradley-Terry model (Bradley and Terry 1952), the reward model $r_\phi(x, y)$ is used to model the preference distri-

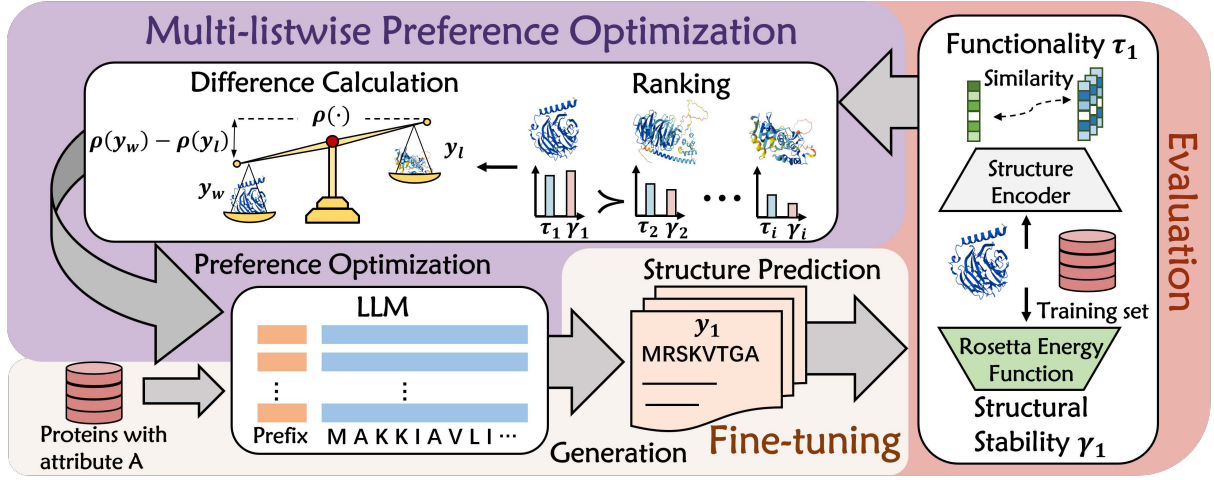


Figure 2: An overview of the proposed method CtrlProt.

bution. The training objective is

$$\mathcal{L}(r_\phi, D) = -\mathbb{E}_{(x, y_w, y_l) \sim D} \left[\log \sigma(r_\phi(x, y_w) - r_\phi(x, y_l)) \right]. \quad (1)$$

During the reinforcement learning phase, we treat the language model as a policy and use the trained $r_\phi(x, y)$ to optimize the LLM parameters. The optimization objective is

$$\max_{\pi_\theta} \mathbb{E}_{x \sim D, y \sim \pi_\theta(y|x)} [r_\phi(x, y)] - \beta \mathbb{D}_{KL}[\pi_\theta(y|x) \parallel \pi_{ref}(y|x)]. \quad (2)$$

Direct preference optimization. Although RLHF is effective in adapting LLMs to human preferences, it involves four sub-models, making the training complex and costly. DPO derives a simple approach for policy optimization using preferences directly. With the optimal solution to the KL-constrained reward maximization objective in Eq. (2), DPO rearranges it to formulate the reward function:

$$r^*(x, y) = \beta \log \frac{\pi_\theta(y|x)}{\pi_{ref}(y|x)} + \beta \log Z(x), \quad (3)$$

where $Z(x) = \sum_y \pi_{ref}(y|x) \exp\left(\frac{r(x, y)}{\beta}\right)$ is the partition function. $r^*(x, y)$ can be substituted back into Eq. (1), and get the following optimization objective:

$$\mathcal{L}_{DPO}(\pi_\theta; \pi_{ref}) = -\mathbb{E}_{(x, y_w, y_l) \sim D} \left[\log \sigma \left(\beta \frac{\pi_\theta(y_w|x)}{\pi_{ref}(y_w|x)} - \beta \frac{\pi_\theta(y_l|x)}{\pi_{ref}(y_l|x)} \right) \right]. \quad (4)$$

Methods

In this section, we provide a detailed description of our proposed method CtrlProt. The framework is shown in Fig. 2. The goal of CtrlProt is to enhance the structural stability and functionality of LLM controllable sequence generation with preference optimization. Our method is divided into three

main steps. First, we finetune an LLM on a protein sequence dataset with specific attributes. Then, we generate and evaluate candidate sequences with functionality and stability metrics to build preference optimization datasets. Finally, we perform multi-listwise preference optimization to improve the performance of the LLM.

Supervised Finetuning

Suppose that we have several protein sequences related to the attribute A . We use prefix-tuning (Li and Liang 2021) for supervised finetuning on the attribute A . Let $y = (a_1, \dots, a_l)$ be a protein sequence with l amino acids. The generation probability of y can be formulated as $p(y) = \prod_{i=1}^l p(a_i | a_{<i})$ and we optimize the LLM by minimizing the negative log-likelihood as follows:

$$L_{sft} = -\sum_{i=1}^k \log p_\theta(a_i | a_{<i}, P_A), \quad (5)$$

where P_A denotes the prefix related to the attribute A .

Although prefix-tuning can effectively leverage pre-trained knowledge and finetune data to generate sequences related to the attribute A , we observe from the experimental results that solely finetuning on attribute sequences may still result in protein sequences with poor structural stability and functionality. This indicates that finetuning alone is insufficient to fully leverage the model’s understanding of protein structural semantics acquired during pre-training. Therefore, we formulate this issue of improving the quality of the generated proteins as a preference optimization problem on structural stability and functionality.

DPO Data Construction

We generate abundant candidate sequences from the prefix-tuned model $\pi_{ref}(\cdot | P_A)$. To fully evaluate the quality of sequence $y_i \sim \pi_{ref}(\cdot | P_A)$, we design two dimensions of evaluation metrics for stability and functionality.

Stability. The Rosetta energy function (Alford et al. 2017) is a widely used metric for assessing conformational energy which reflects the structural stability of the protein (Nijkamp et al. 2023; Ferruz, Schmidt, and Höcker 2022). It includes interactions and force fields like van der Waals forces, charge interactions, hydrogen bonds, and virtual side-chain conformations (Alford et al. 2017). Typically, protein structures that achieve lower scores are more likely to approximate stable structures. To compare the stability between proteins with different lengths, we normalize the raw Rosetta scores by lengths and use the per-residue Rosetta score (Kim, Seffernick, and Lindert 2018) e_i of y_i to calculate the structural stability score γ_i :

$$\gamma_i = 1 - \frac{e_i - e_{\min}}{e_{\max} - e_{\min}}, \quad (6)$$

where e_{\max} and e_{\min} are the maximum and minimum Rosetta energy scores of all sequences, respectively.

Functionality. The structural similarity between protein sequences indicates their functionality relevance (Hamamsy et al. 2024). We use a pre-trained encoder to obtain representations from the protein structures and measure the functionality relevance between y_i and the sequences in the training set based on the similarity of their structural representations. The functionality score τ_i^A is:

$$\tau_i^A = \frac{1}{M} \sum_{j=1}^M \cos(\text{Encoder}(y_i), \text{Encoder}(y_j^{\text{train}})), \quad (7)$$

where $\cos(\cdot)$ measures cosine similarity and y_j^{train} is from the training set containing M sequences. We use Protein-MPNN (Dauparas et al. 2022) as the structural encoder.

Based on the two scores γ_i and τ_i^A , we can build the preference optimization dataset D_A for the attribute A :

$$D_A = \{(y_w, y_l, \gamma_w, \gamma_l, \tau_w^A, \tau_l^A) \mid \gamma_w > \gamma_l \wedge \tau_w^A > \tau_l^A\}, \quad (8)$$

where $y_w, y_l \sim \pi_{\text{ref}}(\cdot \mid P_A)$ and both scores of y_w are greater than y_l in D_A .

Multi-listwise DPO

Since D_A is obtained by extensive sampling and simultaneous evaluation of functionality and structural stability, our preference optimization method should also consider both multidimensionality and sequentiality.

Sequentiality. All sequences are evaluated based on determined scores, which can be compared and ranked. It is natural to incorporate ranking information into DPO, making the process more attentive to pairs with significant differences while reducing the intensity of optimization for pairs with minor differences (Song et al. 2024). Here we use $G(y_i, \gamma_i)$ to calculate the weighted stability score of y_i :

$$G(y_i, \gamma_i) = F(\gamma_i)(2^{\gamma_i} - 1), \quad (9)$$

where $F(\cdot)$ denotes the cumulative distribution function of the distribution of γ , which provides the rank of γ_i within the entire dataset. Even when scores are relatively concentrated, $F(\gamma_i)$ can still help capture and utilize the information of the subtle differences between samples. We use the beta distribution to fit the score and calculate $F(\cdot)$.

Multidimensionality. We aim to simultaneously consider the functionality and structural stability of protein sequences during the optimization process. Thus, the difference between chosen and rejected sequences should be considered by τ_i^A and γ_i , respectively. We obtain the quality score $\rho(y_i)$ of the sequence y_i by

$$\rho(y_i) = G(y_i, \gamma_i) + G(y_i, \tau_i^A), \quad (10)$$

where $G(y_i, \tau_i)$ is generated in the same way as Eq. (9).

Preference optimization. We rewrite the optimization objective based on Eq. (2) with the quality score $\rho(\cdot)$:

$$\begin{aligned} \max_{\pi_\theta} \mathbb{E}_{x \sim D, y \sim \pi_\theta(y \mid x)} [r^*(x, y) + \rho(y)] \\ - \beta \mathbb{D}_{KL}[\pi_\theta(y \mid x) \parallel \pi_{\text{ref}}(y \mid x)], \end{aligned} \quad (11)$$

where we use $r^*(x, y) + \rho(y)$ equivalently replaces the originally hypothesized latent reward function.

Following the same derivation in DPO, we get the mapping from reward functions to optimal policies:

$$r^*(x, y) = \beta \log \frac{\pi_{r^*}(y \mid x)}{\pi_{\text{ref}}(y \mid x)} + \beta \log Z(x) - \rho(y). \quad (12)$$

Finally, substituting $r^*(x, y)$ into Eq. (1), we eliminate the partition function $Z(x)$ and obtain the final loss:

$$\begin{aligned} \mathcal{L}_{MLPO}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{x, y_w, y_l \sim D} \left[\log \sigma \left(\beta \frac{\pi_\theta(y_w \mid x)}{\pi_{\text{ref}}(y_w \mid x)} \right. \right. \\ \left. \left. - \beta \frac{\pi_\theta(y_l \mid x)}{\pi_{\text{ref}}(y_l \mid x)} - \alpha(\rho(y_w) - \rho(y_l)) \right) \right], \end{aligned} \quad (13)$$

where we add α to adjust the intensity. $\alpha(\rho(y_w) - \rho(y_l))$ denotes the difference between preference optimization pairs and as a regularization term, influences the training. Specifically, when y_w and y_l have a significant difference, the gradient on the pairs during training will increase. Conversely, when y_w and y_l are similar in functionality and stability, the training gradient will decrease (Park et al. 2024). This can also be understood from a margin perspective (Amini, Vieira, and Cotterell 2024) that pairs with larger differences between y_w and y_l will be penalized more heavily.

Generalization to Multi-attribute Generation

Assuming we have sufficient protein sequences with multiple attributes, it can be treated as a generation task similar to the single-attribute scenario and optimized using the aforementioned method. However, in most cases, the amount of protein data with multi-attribute labels is limited, making it difficult to directly optimize with a multi-attribute dataset. Therefore, we aim to extend the above method and leverage several single-attribute datasets to enhance the effectiveness of multi-attribute generation.

Suppose that we need to generate protein sequences with K attributes. After prefix-tuning, we can get the prefix for single attribute P_1, P_2, \dots, P_K . For multi-attribute generation, a naive method is to directly concatenate all the prefixes and use $P_{\text{multi}} = [P_1; P_2; \dots; P_K]$ to generate sequences with all K attributes (Luo et al. 2023).

	CLS	TM	RMSD	pLDDT	CLS	TM	RMSD	pLDDT	CLS	TM	RMSD	pLDDT
Models	MFO: Metal ion binding				BPO: Phosphorylation				CCO: Cytoplasm			
ESM-1b	50.1	54.9	7.17	68.8	64.5	60.2	5.29	68.3	48.4	52.8	6.74	57.8
ESM-2	55.4	67.7	6.67	70.7	77.2	<u>75.0</u>	<u>3.31</u>	70.1	50.4	62.3	4.27	60.8
EvoDiff	69.7	58.7	7.62	60.4	87.1	67.0	5.45	66.3	62.9	50.5	7.44	55.6
PrefixProt	57.7	67.9	8.33	<u>70.8</u>	82.0	67.3	3.77	<u>76.6</u>	51.1	47.0	6.77	59.5
ProGen2	62.5	<u>69.7</u>	7.52	68.1	85.5	66.3	3.82	76.0	59.1	<u>63.5</u>	4.45	<u>64.7</u>
ProLLaMA	74.3	56.9	<u>5.52</u>	55.9	<u>87.3</u>	66.8	4.79	64.0	72.4	55.9	<u>3.64</u>	61.2
CtrlProt (ours)	<u>70.5</u>	74.2	4.38	72.4	97.4	94.1	1.18	83.8	<u>65.4</u>	85.7	2.80	81.0
Models	MFO: RNA binding				BPO: Translation				CCO: Nucleus			
ESM-1b	51.2	48.6	6.83	49.1	70.4	39.8	9.77	52.9	66.4	35.5	14.40	56.0
ESM-2	53.9	38.9	9.19	52.6	73.3	46.7	9.12	51.7	74.8	<u>37.4</u>	9.67	61.4
EvoDiff	<u>69.3</u>	47.4	7.33	53.6	87.7	64.7	5.39	60.2	72.0	29.0	17.79	49.0
PrefixProt	<u>67.3</u>	51.7	9.52	<u>62.7</u>	83.3	<u>71.0</u>	5.15	<u>68.7</u>	<u>77.9</u>	26.7	24.88	<u>61.9</u>
ProGen2	64.2	<u>53.8</u>	<u>6.46</u>	55.6	87.5	68.5	6.45	63.8	76.4	30.2	16.14	59.8
ProLLaMA	69.2	42.2	8.42	57.9	<u>88.2</u>	59.3	<u>4.82</u>	67.2	71.2	23.4	27.50	57.8
CtrlProt (ours)	76.1	79.3	2.86	69.9	89.0	88.2	2.54	73.0	78.4	44.9	<u>12.05</u>	63.4

Table 1: Results of single-attribute generation

However, the generation quality with P_{multi} is suboptimal and fails to balance multiple functionalities. Therefore, we employ a similar preference optimization method to further refine the model. First, we generate candidate sequences $y_1, y_2, \dots, y_n \sim \pi_{ref}(\cdot | P_{multi})$, and evaluate structural stability with Eq. (6) and functionality on each attribute with Eq. (7). Then, we construct the preference optimization dataset D_{multi} for multi-attribute generation similar to Eq. (8). We can get the new quality score:

$$\rho_{multi}(y_i) = G(y_i, \gamma_i) + \frac{1}{K} \sum_{k=1}^K G(y_i, \tau_i^k). \quad (14)$$

We substitute $\rho_{multi}(y_i)$ into Eq. (13) as the final loss.

Experiments and Results

Experiment Setup

Dataset construction. We extract protein sequences with Gene Ontology (GO) terms from the UniProtKB database² and corresponding structures from the AlphaFold protein structure database³. We choose six terms as attributes from three different aspects: molecular function ontology (MFO): metal ion binding and RNA binding; biological process ontology (BPO): phosphorylation and translation; cellular component ontology (CCO): cytoplasm and nucleus. Each attribute contains 10k protein sequences for training.

Evaluation metrics. To evaluate the quality of sequences comprehensively, we use CLS-score, TM-score, and RMSD to assess their functionality and pLDDT for structural stability. For each attribute, we extract 100k sequences from UniProtKB as the evaluation set, excluding the training set to ensure no data leakage. We construct a database for each attribute on the evaluation set for alignment with Foldseek (Van Kempen et al. 2024). **CLS-score:** We finetune

a classifier based on the ESM-2 (Lin et al. 2023a) model on the evaluation set, using the classification probabilities as the classifier score (CLS-score). **TM-score** and **RMSD:** Following previous works (Lv et al. 2024; Zongying et al. 2024), we use Foldseek (Van Kempen et al. 2024) to assess structural similarity with Template Modeling score (TM-score) (Zhang and Skolnick 2004) and Root Mean Square Distance (RMSD) (Betancourt and Skolnick 2001). Higher CLS-score or TM-score, or a lower RMSD, indicate greater structural similarity between the generated sequences and the evaluation set. **pLDDT:** The predicted Local Distance Difference Test (pLDDT) is used to assess the confidence of protein structure predictions. A higher pLDDT indicates higher prediction confidence and greater structural stability.

Baselines. We compare CtrlProt to six competitive baselines: The encoder-based methods including ESM-1b (Rives et al. 2021) and ESM-2 (Lin et al. 2023a). The diffusion-based model is EvoDiff (Alamdari et al. 2023). We also compare with the state-of-the-art decoder-based methods. PrefixProt (Luo et al. 2023) uses prefix-tuning to finetune ProtGPT2 (Ferruz, Schmidt, and Höcker 2022). ProGen2 (Nijkamp et al. 2023) is pre-trained on a large corpus of protein sequences, and we finetune it using the same prefix-tuning setting. ProLLaMa (Lv et al. 2024) leverages the alignment of natural language to generate corresponding protein sequences. All baselines are finetuned on the training set.

Settings. For prefix-tuning, we finetune ProtGPT2 with following settings: batch size (16), learning rate (1e-4), prefix token number (100). For preference optimization, we use 5k pairs on each attribute and set the learning rate (5e-5), β (0.1), and α (0.05). The maximum generation length is 400. We use ProteinMPNN as the structural encoder and ESM-Fold (Lin et al. 2023b) for structure prediction, both with default parameters. The Rosetta score is calculated using the weight configuration of ref2015 (Park et al. 2016). All training and generation are conducted on a single A800 GPU.

²<https://www.uniprot.org/>

³<https://alphafold.ebi.ac.uk/>

Variants	CLS	TM	RMSD	pLDDT
CtrlProt (full)	79.5	77.7	<u>4.30</u>	73.9
w/o γ	75.3	75.9	4.41	70.2
w/o τ	71.9	66.2	4.90	70.5
w/ DPO	76.4	72.0	4.20	72.0
w/ ORPO	74.0	70.2	4.80	72.4
w/ Lipo- λ	74.4	70.1	4.92	<u>73.5</u>

Table 2: Results of ablation and alternative studies

Single-attribute Generation Results

To assess our method CtrlProt’s performance in controllable sequence generation, for each attribute, we generate 500 sequences using our method and baselines. We compare the generated results with natural proteins from the evaluation set using Foldseek. The results, as shown in Table 1, indicate that our method outperforms the baselines on six single-attribute controllable generation datasets. Notably, CtrlProt exhibits a significant advantage in pLDDT and TM-score, suggesting that our generated sequences have greater structural stability, and are structurally more similar to natural proteins with the same attributes, which implies similar functionality. This verifies that CtrlProt effectively improves the quality of the generated proteins.

Overall, the performance of decoder-based models after finetuning is superior to encoder-based models, highlighting their significant advantage in sequence generation tasks. ESM-1b, ESM-2, and EvoDiff generate sequences by predicting masked tokens. It is challenging to generate protein sequences from scratch. Therefore, we provide 10% of amino acids to assist in the generation process. It is worth noting that, while ProLLaMA has a larger number of parameters, its performance is not particularly remarkable. This may be because its original instruction tuning limits the sequence length to 256. Although we do not impose such a restriction, it is still affected, leading to decreased performance in generating longer protein sequences.

Ablation and Alternative Results

We perform ablation and alternative studies to further analyze the effectiveness of CtrlProt. As presented in Table 2, we list the average results across six single-attribute datasets. In the ablation study, we separately remove the functionality metric τ and the structural stability metric γ , and both result in a decline in CtrlProt’s performance. Specifically, removing γ leads to a more significant drop in pLDDT, while removing τ causes larger decreases in CLS-score, TM-score, and RMSD. Therefore, τ and γ play a crucial role in enhancing functionality and structural stability, respectively.

In the alternative study, we focus on comparing several preference optimization methods: DPO, ORPO, and Lipo- λ . Overall, our multi-listwise preference optimization outperforms all other methods. Compared to DPO and ORPO, CtrlProt introduces the difference of the quality scores as a regularization term, allowing our method to pay more attention to the pairs with greater quality differences. This leads to superior results in protein sequence data. On the other hand,

	Inter-output (%)	Training set (%)
Natural proteins	2.53	2.53
ESM-1b	6.15	3.13
ESM2	5.57	<u>2.60</u>
EvoDiff	3.39	2.67
PrefixProt	3.12	2.73
ProGen2	<u>3.26</u>	2.67
ProLLaMA	5.13	2.73
CtrlProt (ours)	3.38	2.55

Table 3: Results of diversity

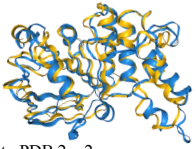
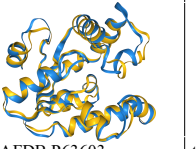
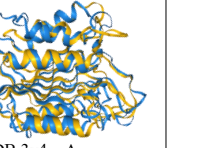
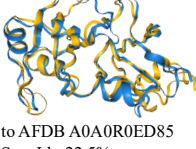

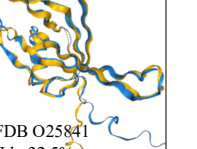
MFO: Metal ion binding  to PDB 2nq2 Seq. Id.: 18.0% TM: 0.824; RMSD: 3.45	BPO: Phosphorylation  to AFDB P63603 Seq. Id.: 28.3% TM: 0.878; RMSD: 2.52	CCO: Cytoplasm  to PDB 3v4o_A Seq. Id.: 19.7% TM: 0.781; RMSD: 4.71
MFO: RNA binding  to AFDB A0A0R0ED85 Seq. Id.: 22.5% TM: 0.851; RMSD: 3.70	BPO: Translation  to AFDB O96173 Seq. Id.: 19.3% TM: 0.760; RMSD: 3.49	CCO: Nucleus  to AFDB O25841 Seq. Id.: 32.5% TM: 0.827; RMSD: 6.98

Figure 3: Case study of single-attribute generation

Lipo- λ , which also employs a listwise method, performs less effectively. This is because Lipo- λ uses the difference of the reciprocal of ranking in its lambda weight. When it comes to ranking and comparing in a large number of protein sequences, it overly emphasizes a few top-ranked proteins and reduces the distinction among lower-ranked proteins. CtrlProt incorporates the CDF of the quality score ρ to introduce ranking information and avoids this issue.

Diversity Analysis

To further verify that CtrlProt can generate high-quality proteins without overfitting to the training set or experiencing mode collapse (Shumailov et al. 2024), we analyze and compare its diversity with other baselines. Following previous works (Kirk et al. 2024; Li et al. 2016), we use n-gram to calculate the similarity ratio between two sequences:

$$Sim(y_i, y_j) = \frac{|Set(y_i) \cap Set(y_j)|}{|Set(y_i)|}, \quad (15)$$

where $Set(\cdot)$ is the set of all 3-gram items. We report the inter-output similarity for the similarity among all generated sequences, and the training set similarity, which shows similarity between generated sequences and the training set.

The results, as shown in Table 3, indicate that CtrlProt gains the lowest similarity to the training set, close to natural proteins, which suggests that CtrlProt does not achieve

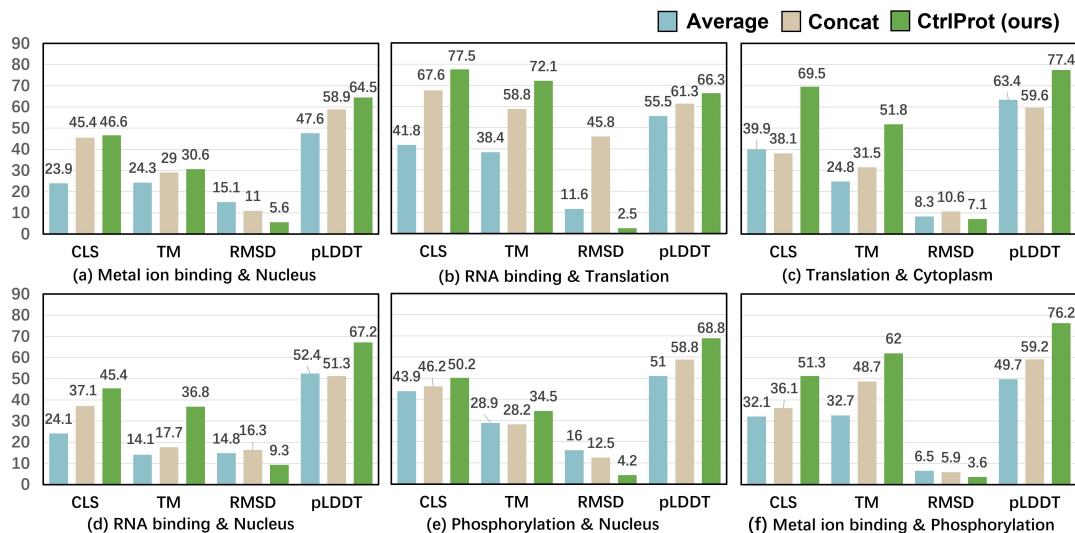


Figure 4: Results of multi-attribute generation

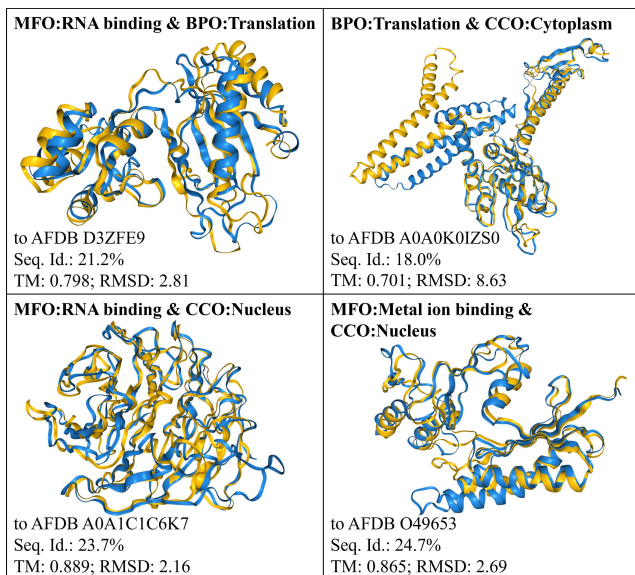


Figure 5: Case study of multi attribute generation

optimal performance through overfitting. Although the inter-output similarity of CtrlProt slightly increases compared to PrefixProt, it remains within a relatively optimal range. It indicates that no mode collapse occurs during preference optimization, meaning CtrlProt does not resort to generating only a few patterns of proteins to obtain optimal results.

Case Study

In Fig. 3, we present proteins generated by CtrlProt (shown in blue) and the most similar natural proteins (shown in yellow). We use Sequence Identity (Seq.Id.) from foldseek to reflect the sequence similarity. The significant overlap in 3D structures and high TM-scores confirm structural similar-

ity between the generated and natural proteins. We observe that these natural proteins also satisfy the corresponding attributes, indicating functional similarity between the generated and natural proteins. The lower Seq.Id. indicates lower amino acid sequence similarity, which means CtrlProt can generate desired attribute proteins with novel sequences.

Multi-attribute Generation Results

CtrlProt can be extended to multi-attribute generation. To validate the effectiveness, we construct six attribute combinations. Due to the lack of studies on multi-attribute generation, we follow PrefixProt by comparing with two straightforward methods: **Average**, where the prefixes are averaged and merged, and **Concat**, where the prefixes are concatenated for generation. We adopt the same metrics used in single-attribute generation. The results, as shown in Fig. 4, demonstrate that CtrlProt achieves higher CLS-score and TM-score, as well as lower RMSD, indicating a significant improvement in functionality. The higher pLDDT suggests stronger structural stability in all attribute combinations.

In Fig. 5, we present four cases of two attribute combinations. The overlap in structures and high TM-scores shows that the generated proteins exhibit structural similarity to natural proteins with related attributes while maintaining low sequence similarity, based on low Seq.Id.

Conclusion

In this paper, we present CtrlProt, a preference optimization-based method that improves the quality of controllable protein sequence generation. We propose multi-listwise preference optimization on functionality and structural stability metrics, targeting pairs with notable quality differences. Experiments show that CtrlProt excels in single-attribute and multi-attribute generation and can generate diverse proteins. However, achieving more precise and programmable generation for certain attribute combinations remains a challenge. We hope to continue exploring this in future work.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62272219).

References

- Alamdari, S.; Thakkar, N.; van den Berg, R.; Lu, A.; Fusi, N.; Amini, A.; and Yang, K. 2023. Protein generation with evolutionary diffusion: sequence is all you need. In *NeurIPS GenBio Workshop*.
- Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. 2017. The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(6): 3031–3048.
- Amini, A.; Vieira, T.; and Cotterell, R. 2024. Direct preference optimization with an offset. *arXiv preprint arXiv:2402.10571*.
- Betancourt, M. R.; and Skolnick, J. 2001. Universal similarity measure for comparing protein structures. *Biopolymers: Original Research on Biomolecules*, 59(5): 305–309.
- Bradley, R. A.; and Terry, M. E. 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4): 324–345.
- Cao, L.; Coventry, B.; Goreschnik, I.; Huang, B.; Sheffler, W.; Park, J. S.; Jude, K. M.; Marković, I.; Kadam, R. U.; Verschuere, K. H.; et al. 2022. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910): 551–560.
- Correia, B. E.; Bates, J. T.; Loomis, R. J.; Baneyx, G.; Carrico, C.; Jardine, J. G.; Rupert, P.; Correnti, C.; Kalyuzhnyi, O.; Vittal, V.; et al. 2014. Proof of principle for epitope-focused vaccine design. *Nature*, 507(7491): 201–206.
- Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I.; Courbet, A.; de Haas, R. J.; Bethel, N.; et al. 2022. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615): 49–56.
- Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; et al. 2021. ProtTrans: Toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10): 7112–7127.
- Fang, Y.; Liang, X.; Zhang, N.; Liu, K.; Huang, R.; Chen, Z.; Fan, X.; and Chen, H. 2024. Mol-Instructions: A Large-Scale Biomolecular Instruction Dataset for Large Language Models. In *ICLR*.
- Ferruz, N.; Schmidt, S.; and Höcker, B. 2022. ProtGPT2 is a deep unsupervised language model for protein design. *Nature Communications*, 13(1): 4348.
- Hamamsy, T.; Morton, J. T.; Blackwell, R.; Berenberg, D.; Carriero, N.; Gligorijevic, V.; Strauss, C. E.; Leman, J. K.; Cho, K.; and Bonneau, R. 2024. Protein remote homology detection and structural alignment using deep learning. *Nature Biotechnology*, 42(6): 975–985.
- Hong, J.; Lee, N.; and Thorne, J. 2024. Reference-free monolithic preference optimization with odds ratio. *arXiv preprint arXiv:2403.07691*.
- Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873): 583–589.
- Kim, S. S.; Seffernick, J. T.; and Lindert, S. 2018. Accurately predicting disordered regions of proteins using Rosetta residuedisorder application. *The Journal of Physical Chemistry B*, 122(14): 3920–3930.
- Kirk, R.; Mediratta, I.; Nalmpantis, C.; Luketina, J.; Hambro, E.; Grefenstette, E.; and Raileanu, R. 2024. Understanding the effects of rlhf on llm generalisation and diversity. In *ICLR*.
- Li, J.; Galley, M.; Brockett, C.; Gao, J.; and Dolan, B. 2016. A Diversity-Promoting Objective Function for Neural Conversation Models. In *NAACL*, 110–119.
- Li, X. L.; and Liang, P. 2021. Prefix-Tuning: Optimizing Continuous Prompts for Generation. In *ACL*, 4582–4597.
- Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. 2023a. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130.
- Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. 2023b. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130.
- Liu, F.; et al. 2020. Learning to summarize from human feedback. In *ACL*.
- Liu, T.; Qin, Z.; Wu, J.; Shen, J.; Khalman, M.; Joshi, R.; Zhao, Y.; Saleh, M.; Baumgartner, S.; Liu, J.; et al. 2024. Lipo: Listwise preference optimization through learning-to-rank. *arXiv preprint arXiv:2402.01878*.
- Luo, J.; Liu, X.; Li, J.; Chen, Q.; and Chen, J. 2023. Controllable Protein Design by Prefix-Tuning Protein Language Models. *bioRxiv*, 2023–12.
- Lv, L.; Lin, Z.; Li, H.; Liu, Y.; Cui, J.; Chen, C. Y.-C.; Yuan, L.; and Tian, Y. 2024. ProLLaMA: A protein large language model for multi-task protein language processing. *arXiv preprint arXiv:2402.16445*.
- Meng, Y.; Xia, M.; and Chen, D. 2024. Simpo: Simple preference optimization with a reference-free reward. *arXiv preprint arXiv:2405.14734*.
- Nathansen, A.; Klein, K.; Renard, B.; Nowicka, M.; and Bartoszewicz, J. M. 2024. Evaluating Tuning Strategies for Sequence Generation with Protein Language Models. In *Machine Learning in Computational Biology*, 76–89.
- Nijkamp, E.; Ruffolo, J. A.; Weinstein, E. N.; Naik, N.; and Madani, A. 2023. Progen2: exploring the boundaries of protein language models. *Cell Systems*, 14(11): 968–978.

- Park, H.; Bradley, P.; Greisen Jr, P.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; Baker, D.; and DiMaio, F. 2016. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *Journal of Chemical Theory and Computation*, 12(12): 6201–6212.
- Park, R.; Rafailov, R.; Ermon, S.; and Finn, C. 2024. Disentangling length from quality in direct preference optimization. *arXiv preprint arXiv:2403.19159*.
- Qin, B.; Feng, D.; and Yang, X. 2024. Towards Understanding the Influence of Reward Margin on Preference Model Performance. *arXiv preprint arXiv:2404.04932*.
- Rafailov, R.; Sharma, A.; Mitchell, E.; Ermon, S.; Manning, C. D.; and Finn, C. 2023. Direct preference optimization: your language model is secretly a reward model. In *NeurIPS*, 53728–53741.
- Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; and Baker, D. 2011. De novo enzyme design using Rosetta3. *PloS One*, 6(5): e19230.
- Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; et al. 2021. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15): e2016239118.
- Shumailov, I.; Shumaylov, Z.; Zhao, Y.; Papernot, N.; Anderson, R.; and Gal, Y. 2024. AI models collapse when trained on recursively generated data. *Nature*, 631(8022): 755–759.
- Śledź, P.; and Caflisch, A. 2018. Protein structure-based drug design: from docking to molecular dynamics. *Current Opinion in Structural Biology*, 48: 93–102.
- Song, F.; Yu, B.; Li, M.; Yu, H.; Huang, F.; Li, Y.; and Wang, H. 2024. Preference ranking optimization for human alignment. In *AAAI*, 18990–18998.
- Van Kempen, M.; Kim, S. S.; Tumescheit, C.; Mirdita, M.; Lee, J.; Gilchrist, C. L.; Söding, J.; and Steinegger, M. 2024. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 42(2): 243–246.
- Wang, C.; Fan, H.; Quan, R.; and Yang, Y. 2024. Protchatgpt: Towards understanding proteins with large language models. *arXiv preprint arXiv:2402.09649*.
- Wang, Z.; Zhang, Q.; Ding, K.; Qin, M.; Zhuang, X.; Li, X.; and Chen, H. 2023. InstructProtein: Aligning human and protein language via knowledge instruction. *arXiv preprint arXiv:2310.03269*.
- Zhang, Y.; and Skolnick, J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4): 702–710.
- Zheng, Z.; Deng, Y.; Xue, D.; Zhou, Y.; Ye, F.; and Gu, Q. 2023. Structure-informed language models are protein designers. In *ICML*, 42317–42338.
- Zhuo, L.; Chi, Z.; Xu, M.; Huang, H.; Zheng, H.; He, C.; Mao, X.-L.; and Zhang, W. 2024. ProtLLM: An interleaved protein-language llm with protein-as-word pre-training. In *ACL*.
- Ziegler, D. M.; Stiennon, N.; Wu, J.; Brown, T. B.; Radford, A.; Amodei, D.; Christiano, P.; and Irving, G. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.
- Zongying, L.; Hao, L.; Liuzhenghao, L.; Bin, L.; Junwu, Z.; Yu-Chian, C. C.; Li, Y.; and Yonghong, T. 2024. TaxDiff: Taxonomic-Guided Diffusion Model for Protein Sequence Generation. *arXiv preprint arXiv:2402.17156*.