

PepCCD: A Contrastive Conditioned Diffusion Framework for Target-Specific Peptide Generation

Jun Zhang^{1,2*}, Yangyang Zhou^{1,2}, Tiantian Zhu^{1,2}, Zexuan Zhu^{1,2*}

¹School of Artificial Intelligence, Shenzhen University, Shenzhen 518060, China

²National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, China

junzhang@szu.edu.cn, zhuzx@szu.edu.cn

Abstract

Peptide-based drug design targeting “undruggable” proteins remains one of the most critical challenges in modern drug discovery. Conventional peptide-discovery pipelines rely on low-throughput experimental screening, which is both time-consuming and prohibitively expensive. Moreover, existing computational approaches for designing peptides against target proteins typically depend on the availability of high-quality structural information. Although recent structure-prediction tools such as AlphaFold3 have achieved breakthroughs in protein modeling, their accuracy for functional interfaces remains limited. The acquisition of high-resolution structures is often expensive, time-intensive, and particularly challenging for targets with dynamic conformations, further restricting the efficient development of peptide therapeutics. Additionally, current sequence-based generative methods follow a paradigm that relies on known templates, which limits the exploration of sequence space and results in generated peptides lacking diversity and novelty. To address these limitations, we propose a contrastive conditioned diffusion framework for target-specific peptide generation, referred to as PepCCD. It employs a contrastive learning strategy between proteins and peptides to extract sequence-based conditioning representations of target proteins, which serve as precise conditions to guide a pre-trained diffusion model to generate peptide sequences with the desired target specificity. Extensive experiments on multiple benchmark target proteins demonstrate that the peptides designed by PepCCD exhibit strong binding affinity and outperform state-of-the-art methods in terms of diversity and generation efficiency.

Introduction

In modern drug development, more than 80% of pathogenic proteins are considered “undruggable” by traditional small-molecule inhibitors due to the lack of stable binding pockets (Dang et al. 2017; Behan et al. 2019). As a therapeutic modality positioned between small molecules and antibodies, peptide-based drugs have emerged as promising candidates to target these undruggable proteins, due to their favorable target specificity, biocompatibility, and low toxicity profiles (Bodanzky 1988; Craik et al. 2013; Fosgerau and Hoffmann 2015a; Gomes et al. 2018; Muttenthaler et al.

2021a). Peptides can achieve precise modulation of disease-related targets through mechanisms such as disrupting protein–protein interactions (PPI), inhibiting enzymatic activity, or inducing targeted protein degradation (Fosgerau and Hoffmann 2015b). To date, over 100 peptide drugs have been approved for the treatment of various diseases, including cancer, diabetes, and HIV (Kaspar and Reichert 2013; Henninot, Collins, and Nuss 2018; Lee et al. 2019; Wang et al. 2022; Chen et al. 2024b). Traditionally, the discovery of therapeutic peptides has relied on high-throughput techniques such as yeast display (Muttenthaler et al. 2021b) or computational tools tailored to specific peptide properties (Lee et al. 2017; Lee, Wong, and Ferguson 2018) to explore the vast sequence space. However, the combinatorial space of potential peptides is huge, and only a small subset of sequences satisfies therapeutic requirements, making brute-force screening approaches both time-consuming and expensive.

To overcome the limitations of traditional brute-force screening, increasing attention has turned to peptide drug design guided by specific information about the target protein. However, designing peptides for specific targets remains a highly challenging task. First, conventional computational design methods (Wang et al. 2024; Watson et al. 2023; Anishchenko et al. 2021) heavily rely on high-resolution three-dimensional protein structures. Although recent deep learning-based structure prediction tools, such as AlphaFold3 (Abramson et al. 2024), have achieved remarkable breakthroughs, they still suffer from limited accuracy in predicting critical binding interfaces and modeling proteins with multiple conformational states (Pereira et al. 2025; Chakravarty, Lee, and Porter 2025). High-quality structural data are still challenging to obtain in practice, especially for intrinsically disordered proteins that lack stable conformations (Oldfield and Dunker 2014; Maiti et al. 2024; Trivedi and Nagarajaram 2022). Second, the peptide sequence space is vast, making the design and search process highly challenging (Jenson et al. 2018). Recently, several innovative sequence-only design approaches have been proposed (Bhat et al. 2025; Chen et al. 2024a); however, these computational methods typically rely heavily on initial template sequences, and the scarcity of high-quality peptide–protein binding data further limits the exploration of the sequence space. As a result, the generated peptides often

*Corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

lack diversity and novelty.

Motivated by the above challenges, we propose a Contrastive Conditioned Diffusion framework for target-specific peptide generation, named **PepCCD**. Unlike existing sequence-guided approaches, PepCCD adopts a generative modeling paradigm, directly using the target protein sequence as a conditioning input to guide the generation of peptide sequences. Overall, the main contributions of this work are as follows:

- We are the first to introduce diffusion models for the design of peptides guided by target protein sequences. PepCCD directly trains a conditional diffusion model, in which target protein sequences are encoded as condition vectors using a contrastive learning framework. The model generates peptide sequences in an end-to-end manner without the need to construct candidate peptide libraries or perform post hoc screening.
- We construct a large-scale synthetic peptide fragment dataset to pre-train the conditional diffusion model. This dataset enhances the model’s prior knowledge of peptide sequences and compensates for the scarcity of experimentally determined protein-peptide complex data, thereby improving the diversity of the generated peptides.
- Extensive experiments on benchmark target proteins demonstrate that PepCCD outperforms current state-of-the-art sequence-guided baselines in generating target-specific peptides, particularly excelling in terms of targeting accuracy, diversity, and generation efficiency.

Related Works

Protein Language Models

Recent advances in deep learning-based protein sequence modeling have opened new avenues for peptide drug discovery. Protein language models (pLMs) pre-trained on large-scale protein sequence databases can now capture rich semantic, structural, and functional patterns, enabling structure-free peptide design (Shin et al. 2021). These breakthroughs have driven significant progress in two key areas: (1) substantial improvements in the effectiveness of protein sequence embeddings (Hayes et al. 2025; Brandes et al. 2022; Lin et al. 2023); and (2) strong adaptability to downstream tasks such as protein design after fine-tuning on task-specific data (Madani et al. 2023; Ferruz, Schmidt, and Höcker 2022; Lv et al. 2025). Inspired by these advances, this study leverages protein language models fine-tuned specifically for peptide generation, enabling *de novo* design of target-specific peptides using only the target protein sequence.

Diffusion Models

Diffusion models, which generate new samples by gradually adding noise to input data and learning to reverse the process from a prior distribution, have emerged as a robust generative framework across various domains, including image and text synthesis (Song and Ermon 2019; Trippel et al.

2022). Recently, their potential has been explored in protein design. For example, Liu et al. (Liu et al. 2025) introduced a text-conditional diffusion model to generate proteins based on natural language prompts. Hoogeboom et al. (Hoogeboom et al. 2022) developed ProtDiff, which leverages E(3)-equivariant graph neural networks to model the diverse distributions of protein backbone coordinates. Despite these advances, the integration of diffusion models into peptide-protein sequence design remains largely unexplored. In particular, no prior work has addressed target protein sequence-guided peptide generation using diffusion models. This study presents the first diffusion-based framework conditioned on protein sequences for *de novo* design of target-specific peptides, filling a critical gap in the field.

Contrastive Learning

As a widely used approach in self-supervised learning, contrastive learning enables models to capture hidden patterns in data without requiring explicit labels. Its core principle is to optimize the interaction of samples in the embedding space, pulling positive pairs closer while pushing negative pairs farther apart. Yuan et al. (Yuan et al. 2021) proposed a multimodal contrastive learning framework to align textual and visual data by enhancing the similarity of related text-image pairs and suppressing the similarity of unrelated ones, thereby achieving cross-modal alignment. Zhang et al. (Zhang et al. 2023) applied conformational augmentation to protein structures, thereby maximizing the similarity between embeddings of the same protein and the dissimilarity between different proteins, which enables more discriminative representations. Inspired by these contrastive learning strategies, we design a protein-peptide sequence contrastive learning framework that extracts informative features from protein sequences and uses them as conditional inputs to guide the generation of target-specific peptides.

Methodology

Datasets

To support both pre-training and fine-tuning of the model, we constructed two complementary datasets: a protein-peptide interaction dataset (S1) for fine-tuning, and a large-scale simulated peptide dataset (S2) for pre-training. The construction of each dataset is described as follows.

Fine-tuning Dataset S1 To train the model to recognize and generate peptides with target-binding capabilities, we followed the PepPrCLIP protocol and constructed a high-quality protein-peptide interaction dataset based on co-crystal structures from the RCSB Protein Data Bank. We defined valid interactions as those with a buried surface area (BSA) greater than or equal to 50 square angstroms. This dataset includes a wide range of binding modes, ranging from strong to weak affinities, effectively capturing diverse real-world interactions. We clustered all protein sequences using CD-HIT (Fu et al. 2012) to ensure data diversity and reduce redundancy. The final dataset consists of 15,110 complex pairs for training and 5,480 for testing.

Pre-training Dataset S2 Inspired by transfer learning strategies, we constructed a simulated peptide pre-training dataset based on UniProt to enhance the model’s generalization to peptide sequences and provide it with broad prior knowledge. We applied a sliding window technique, with a window size of 30 amino acids and a stride of 15 amino acids, to segment sequences longer than a specified threshold. Sequences containing non-canonical amino acids (B, J, O, U, X, Z) were removed. This process generated a total of 68,958,049 simulated peptide sequences that were shorter than 30 amino acids in length. The sliding window approach is widely used in peptide function prediction and screening tasks, and its effectiveness and validity in sequence modeling have been demonstrated in previous studies.

Overview of PepCCD

The overall framework of the proposed method is illustrated in Figure 1(A), where a potential target-specific peptide can be designed based solely on the input target protein sequence. The protein encoder is a pre-trained ESM-2 protein language model, fine-tuned via protein–peptide semantic alignment, and is used to extract features from the protein sequence as a conditional vector to guide peptide generation. By sampling Gaussian noise to obtain hidden states, which are then concatenated with the target condition vector to serve as the input to the conditional diffusion model. The model then denoises this input to produce a latent representation, which is subsequently decoded into the final peptide sequence. In this study, we introduce a three-stage training strategy that progressively enhances the model’s ability to generate peptides with both diversity and target specificity.

Stage 1: Protein-Peptide Alignment

Inferring latent peptides solely from a protein sequence is a challenging task that requires identifying the correct protein–peptide pairing within a many-to-many mapping. To establish a semantic bridge between proteins and peptides, we introduce a joint embedding framework built on contrastive learning. To capture deep sequential and structural semantics while avoiding discrepancies arising from different protein language models, we employ two identical ESM-2 encoders, one for protein sequences and one for peptide sequences. ESM-2 is a large-scale pre-trained protein language model that encodes rich structural, functional, and evolutionary information embedded in amino-acid sequences, and has proven effective across diverse downstream tasks. This property enables PepCCD to design structurally meaningful peptides even in the absence of target protein structures.

Figure 1(B) illustrates the semantic alignment process. During training, both encoders are fine-tuned on the constructed dataset S1. The training objective is to use contrastive learning to ensure that matched protein-peptide pairs exhibit high similarity in the feature space, while unmatched pairs exhibit low similarity. In this stage, we employ the InfoNCE loss to achieve this goal, formulated as follows:

$$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{N} \sum_{i=1}^N \log \frac{\exp(\mathbf{T}_i \cdot \mathbf{P}_i)}{\sum_{j=1}^N \exp(\mathbf{T}_i \cdot \mathbf{P}_j)} \quad (1)$$

Here, \mathbf{T}_i and \mathbf{P}_i represent the normalized embedding vectors of the target protein and peptide, respectively. Essentially, the objective is to maximize the probability of the positive pair among all possible candidates. In this stage, we evaluated the performance of different types of encoders for the protein–peptide semantic alignment task. The detailed results are shown in the Appendix.

Stage 2: Unconditional Peptide Pre-training

In Stage 1, we obtained a pair of encoders that can accurately align and capture protein–peptide semantic relationships. To further enhance the model’s stability and improve the diversity of peptide generation, we introduce a large-scale unlabeled peptide dataset S2 in this stage for the unconditional pre-training of the diffusion model. The diffusion model comprises two primary processes: a noise addition process and a denoising process. In the noise addition phase, peptide embeddings undergo a forward diffusion transformation, gradually injected with Gaussian noise. In the denoising phase, a Transformer-based architecture is used to reconstruct the embeddings. This Transformer consists of multiple modules, each incorporating multilayer perceptrons (MLPs) and multi-head self-attention, enabling the model to capture complex noise patterns by modeling dependencies across sequence positions.

As shown in Figure 1(C), we first use the pre-trained peptide encoder to extract embedding representations from the unlabeled peptide sequences, allowing the model to adapt to the target-conditioned generation task. Gaussian noise is then added to the embeddings over 500 time steps, simulating the diffusion process as a Markov chain. The diffusion model learns to reconstruct the original embeddings from the noisy representations, thereby capturing the prior distribution of peptide features.

At this stage, we utilize the Mean Squared Error (MSE) as the loss function. The training objective is to constrain the diffusion model to minimize the discrepancy between the reconstructed and original embeddings, enabling it to effectively learn the complex distribution within the peptide embedding space and acquire rich semantic representations. The loss function is defined as follows:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N \left\| \hat{\mathbf{P}}_i - \mathbf{P}_i \right\|^2 \quad (2)$$

where \mathbf{P}_i and $\hat{\mathbf{P}}_i$ denote the original embedding matrix extracted by the pre-trained peptide encoder and the embedding matrix reconstructed from noise by the diffusion model, respectively.

Stage 3: Target-Guided Conditioned Fine-Tuning

In this stage, pre-trained encoders extract target protein condition vectors and peptide embedding matrices from the fine-tuning dataset. The condition vectors guide the diffusion model to generate target-specific peptide sequences.

During denoising, the condition vector c , noisy latent X_t , and time embedding t are concatenated and fed into a Transformer-based denoiser to predict noise. Conditioned

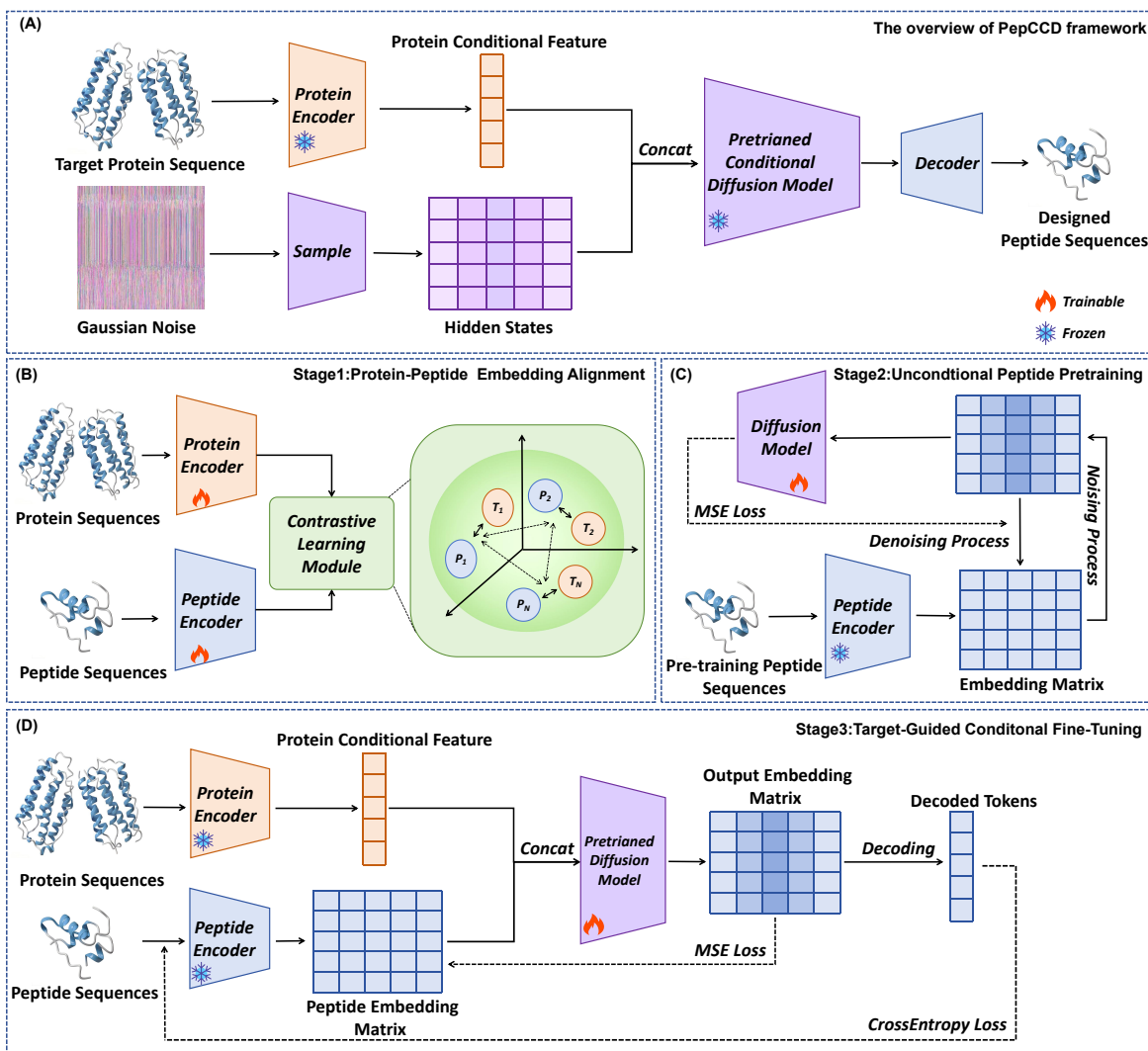


Figure 1: Overview of the PepCCD framework. (A) The overall architecture for generating target-specific peptides from protein sequences. (B) Protein-peptide semantic alignment via contrastive learning using dual ESM-2 encoders. (C) Unconditional pre-training of the diffusion model on unlabeled peptides. (D) Target-guided fine-tuning for peptide generation using conditional denoising.

on the target protein, the model minimizes the mean squared error (MSE) between predicted and true noise:

$$\mathcal{L}_{\text{MSE}} = \mathbb{E}_{t, X_0, \epsilon} [\|\epsilon - \epsilon_{\theta}(X_t, t, c)\|^2], \quad (3)$$

where $\epsilon_{\theta}(\cdot)$ denotes the predicted noise.

As shown in Figure 1(D), the model reconstructs peptide embeddings from Gaussian noise under the target condition, which are then decoded into amino acid sequences. A cross-entropy loss measures the difference between generated and true sequences:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(q_i) + (1 - y_i) \log(1 - q_i)], \quad (4)$$

where y_i is the true token and q_i is the predicted probability at position i .

The total training objective combines both losses:

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \mathcal{L}_{\text{CE}}. \quad (5)$$

This joint optimization improves peptide reconstruction and enhances the model's target-specific generative capability. During inference, Gaussian noise and condition vectors from target proteins guide sampling to produce diverse, biologically relevant peptides.

Experiments

Experiment Setup

Baselines To evaluate the effectiveness of PepCCD, we compared it against two state-of-the-art approaches that represent distinct paradigms in peptide design: one that utilizes

Methods	Type	ipTM(best) \uparrow	ipTM(avg) \uparrow	RT-score(best) \downarrow	RT-score(avg) \downarrow
RFdiffusion	Structure	0.7838***	0.5960***	-394.422***	-242.238***
PepPrCLIP	Sequence	0.7141***	0.4901***	-363.481*	-157.573***
PepCCD	Sequence	<u>0.7418</u>	<u>0.5557</u>	<u>-373.731</u>	<u>-188.933</u>

Table 1: Performance on target-oriented peptide design. In the column headers, (**best**) denotes the best single peptide designed for each target, whereas (**avg**) denotes the average performance across all targets. All statistical analyses were conducted using a two-sided paired Student’s t-test to evaluate the significance of performance differences between models. Statistical significance levels are indicated directly in tables ($p \leq 0.05$ (*), $p \leq 0.01$ (**), $p \leq 0.001$ (***)).

Methods	Sequence similarity \downarrow	Structure similarity \downarrow	Bioactivity \uparrow	Instability \downarrow	Inference Time (one peptide) \downarrow
RFdiffusion	0.3529***	0.3846***	0.2012***	45.462***	66.865s
PepPrCLIP	<u>0.2666***</u>	0.3453***	<u>0.2981***</u>	52.341***	6.699s
PepCCD	0.1895	0.2701	0.4419	37.663	<u>1.244s</u>

Table 2: Overall Performance and Efficiency Comparison. Inference Time is the wall-clock time averaged over the full test set, measured on an NVIDIA GTX 4090 GPU.

the structural information of the target protein (structure-guided), and another that conditions on the protein sequence (sequence-guided). These two methods, RFdiffusion and PepPrCLIP, are considered among the most competitive models for structure-based and sequence-based peptide generation, respectively.

It is worth noting that RFdiffusion was not initially designed for peptide sequence generation; rather, its primary focus lies in protein backbone modeling and structural design. However, due to its strong capabilities in structure generation, it has demonstrated promising potential in structure-guided peptide design tasks. Therefore, we adopt RFdiffusion as the representative structure-based method, and utilize ProteinMPNN (Dauparas et al. 2022) to generate amino acid sequences from its predicted backbones, following the same protocol as its official Colab implementation.

To ensure fairness and accuracy, we strictly followed the official implementation procedures of both RFdiffusion and PepPrCLIP, and deployed them in the local environment for consistent evaluation.

Evaluation Protocol We conducted peptide drug design experiments on 209 target proteins, consistent with the evaluation set used by PepPrCLIP. For each target, every model was tasked with generating 10 novel peptide sequences. The quality of the generated peptides was assessed using the following metrics:

- **Interface TM score (ipTM)**: A key evaluation metric from the AlphaFold3 framework used to assess the binding affinity between the protein and peptide.
- **Rosetta-total-score (RT-score)** (Leaver-Fay et al. 2011): Measures the overall energy stability of the predicted complex, reflecting the overall stability of the complex interface by integrating various energy terms.
- **Sequence similarity**: Assesses the global sequence similarity between the generated peptide and its corresponding natural peptide, reflecting sequence-level diversity.

- **Structure similarity**: Quantifies the structural similarity between the generated and native peptides using the TM-score (Zhang et al. 2022), reflecting structural diversity.
- **Bioactivity** (Mooney et al. 2012): Predicts whether the generated peptide has potential biological activity, including functions such as antimicrobial effects or signal regulation.
- **Instability** (Gasteiger et al. 2005): Estimates the stability of the generated peptide under in vitro conditions, indicating its potential for experimental expression and application.

Detailed information on the tested target proteins, metrics, and implementation can be found in the Appendix. We only reported the average metrics over all generated peptides for each method in the experimental results. The code, data, and appendix are available at <https://github.com/ZhouYyang/PepCCD>.

Experimental Results

Performance Comparison

In the results of targeted peptide design performance, as shown in Table 3, PepCCD demonstrates significant advantages over a similar baseline method that is solely guided by sequence information, in terms of multiple key indicators evaluating the interaction between the peptide and target protein. This reflects its strong generative ability and modeling accuracy. Although PepCCD still slightly lags behind the state-of-the-art structure-guided model, RFDiffusion, which excels in predicting structured targets, this is primarily attributed to its direct utilization of protein three-dimensional structural information. Notably, all test targets in the benchmark are sourced from the PDB database and possess experimentally validated three-dimensional structures. It is plausible that these or structurally similar proteins were present in RFDiffusion’s training data, potentially offering it an advantage in these evaluations.

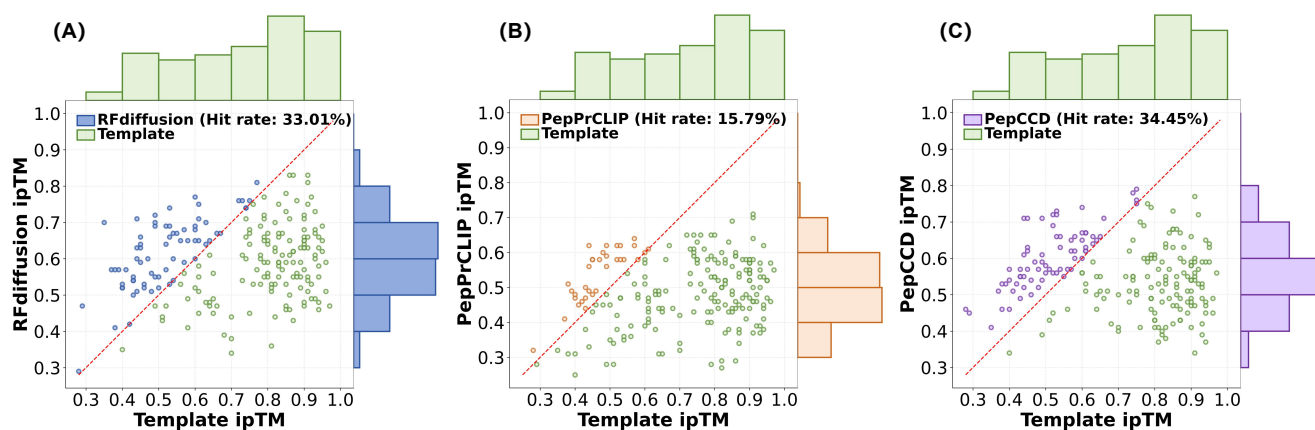


Figure 2: The hit rate is defined as the proportion of generated peptides whose ipTM(avg) scores are greater than or equal to those of their corresponding templates, serving as a metric to evaluate the model’s capability in producing high-quality peptide structures. Points above the diagonal line are considered hits. Comparisons of different models across various targets are presented in the Appendix.

On the more generalizable metric, such as hit rate (calculated using ipTM (avg)), PepCCD even surpasses RFDiffusion, as shown in Figure 2. This demonstrates that, in settings where explicit structural information is unavailable and only sequence-based conditioning is used, PepCCD demonstrates exceptional stability and reliability in generating high-quality peptide candidates, ranking among the top-performing methods currently available.

In terms of overall performance and efficiency (as shown in Table 2), PepCCD achieves the lowest sequence and structural similarity, reduced instability, and the best bioactivity scores among all baseline methods, indicating strong generalization ability in generating diverse and stable peptide candidates. More importantly, PepCCD significantly outperforms existing complex structure-based modeling approaches in terms of inference efficiency, thereby greatly enhancing scalability in practical applications.

We further evaluated the Global Amino-acid Composition Discrepancy (GACD). As shown in Figure 3, peptides generated by PepCCD exhibit amino acid distributions closest to those of natural templates, indicating strong biological plausibility. In terms of overall performance and efficiency (as shown in Table 2), PepCCD achieves the lowest sequence and structural similarity, reduced instability, and the best bioactivity scores among all baseline methods, indicating strong generalization ability in generating diverse and stable peptide candidates. It also outperforms existing complex structure-based modeling approaches in terms of inference efficiency, thereby greatly enhancing scalability in practical applications.

Ablation Study

To evaluate the necessity of each module, we compared PepCCD with three ablated variants: (1) PepCCD (w/o Align) that removes the align stage, (2) PepCCD (w/o Pre-training) that removes the pre-training stage, and (3) PepCCD (w/o Align & Pre-training) that removes both align and pre-training stages. The evaluation metrics used in the

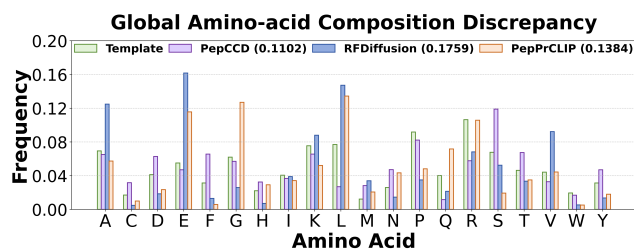


Figure 3: GACD is used to assess the difference in amino acid composition between peptide sequences and template peptides. The closer the frequency distribution is to the natural templates, the higher the biological relevance of the generated peptides will be. Detailed calculation procedures are provided in the Appendix.

ablation study differ from those in the main comparison experiments. Since PepCCD is entirely sequence-based in both representation and generation, we avoided the high computational cost associated with AlphaFold3 complex structure prediction. Instead, we employed a more efficient sequence-level metric, **Superior ratio**, as a substitute for ipTM and RT-score (both of which rely on predicted structures and are computationally expensive). In addition, we introduced two new metrics, including Intra Similarity (**Intra-Sim**) and Inter Similarity (**Inter-Sim**), to evaluate the model’s performance in terms of intra-target diversity and inter-target specificity, respectively. The detailed calculation methods of these three new metrics are provided in the Appendix.

As shown in Table 3, removing the align module results in performance degradation across all metrics, with the Superior ratio exhibiting the most significant decline. Eliminating the pre-training module also results in an overall decline, particularly with a noticeable decrease in sequence diversity. When both modules are removed, the model’s performance deteriorates even further. Overall, the align and pre-

Methods	Intra-Sim↓	Inter-Sim↓	Superior ratio↑	GACD↓	Bioactivity↑	Instability↓
PepCCD (w/o Align & Pre-training)	0.3532	0.4097	21.05	0.3729	0.2720	57.68
PepCCD (w/o Pre-training)	0.3347	0.3670	37.79	0.2804	0.3439	52.94
PepCCD (w/o Align)	0.2845	0.3513	22.96	0.2418	0.3383	44.63
PepCCD	0.2392	0.2445	44.97	0.1754	0.4419	37.66

Table 3: Ablation study results of PepCCD and its three variants across six evaluation metrics. Note that Intra Similarity measures the sequence similarity among peptides generated for the same target, reflecting diversity; Inter Similarity measures the similarity between peptides generated for different targets, reflecting the model’s ability to distinguish between targets; Superior ratio evaluates the target-specificity of the generated peptides.

training modules play essential roles in enhancing the specificity and diversity of the generated peptides, respectively. PepCCD achieves optimal performance when all modules are included, and the absence of any single component compromises both the biological plausibility and diversity of the generated peptide candidates.

Molecular Dynamics Simulation

To further evaluate the binding affinity and structural stability of the designed peptides, we performed all-atom molecular dynamics (MD) simulations using the GROMACS software package (Abraham et al. 2015), and calculated binding free energies using MM/GBSA and MM/PBSA methods (Genheden and Ryde 2015). The simulations focused on interactions with the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein (reference structure: PDB ID:6M0J). For each of the three methods, 10 representative peptides were selected. Their structures were predicted using AlphaFold3 and docked into the RBD binding pocket (Yan et al. 2020) via Rosetta. The most stable complex conformations were then subjected to standard preprocessing (Jo et al. 2008) and 100 ns production simulations to reflect physiological conditions. Full details of the simulation setup, including force field selection, solvation models, equilibration protocols, and free energy calculation parameters, are provided in the Appendix.

As shown in Figure 4, sequence-guided methods perform excellently in MD simulations, with peptides designed by PepCCD exhibiting the lowest binding free energies under both MM/GBSA and MM/PBSA calculations. Additionally, these peptides show significantly lower variance in binding energies, indicating better stability and consistency throughout the simulations. Notably, although RFDiffusion performs well on static target-binding metrics, it demonstrates the poorest performance in molecular dynamics simulations. This discrepancy may be due to the selected complex structure (PDB: 6M0J) lacking a native peptide ligand for the RBD, which prevents RFDiffusion from fully leveraging the advantages of its pre-trained structural information.

Conclusion

In this work, we introduced PepCCD, a novel contrastive conditioned diffusion framework designed for target-specific peptide generation based solely on protein sequence information. Unlike traditional approaches that rely

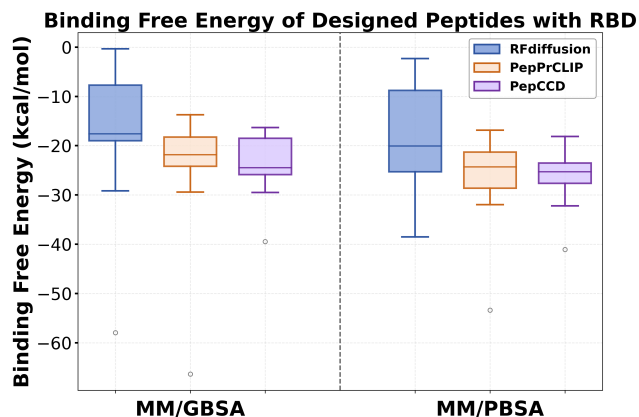


Figure 4: Based on the simulation trajectories, the binding free energies of each peptide–protein complex were calculated using the MM/GBSA and MM/PBSA methods, in order to assess the binding affinity between the peptides and the target protein.

on structural templates or limited sequence-guided heuristics, PepCCD integrates contrastive learning and diffusion-based generative modeling to achieve effective semantic alignment between proteins and peptides, thereby generating highly diverse, biologically relevant peptide sequences.

Extensive experiments on multiple benchmark datasets show that PepCCD outperforms existing sequence-based and even structure-based peptide generation methods in terms of binding affinity, diversity, bioactivity, and inference time. Furthermore, ablation studies validate the critical role of each component in enhancing specificity and diversity. At the same time, molecular dynamics simulations confirm the binding stability of peptides generated by PepCCD in realistic biological contexts. Overall, PepCCD offers a scalable and efficient structure-free solution for *de novo* peptide generation, especially valuable for targets lacking high-resolution structural data.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62302311, 62406199, and 62471310), the National Key Research and Development Program of China (No. 2022YFF1202104), the Guangdong Basic and Applied Basic Research Foundation (No.

2024A1515011681), Shenzhen Research Initiation Program for High-Caliber Critical Talent (No. 827-000932), and the Internal Fund of National Engineering Laboratory for Big Data System Computing Technology (No. SZU-BDSC-IF2024-01 and SZU-BDSC-IF2024-10).

References

- Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; and Lindahl, E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1: 19–25.
- Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500.
- 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. 2021. De novo protein design by deep network hallucination. *Nature*, 600(7889): 547–552.
- Behan, F. M.; Iorio, F.; Picco, G.; Gonçalves, E.; Beaver, C. M.; Migliardi, G.; Santos, R.; Rao, Y.; Sassi, F.; Pinnelli, M.; et al. 2019. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature*, 568(7753): 511–516.
- Bhat, S.; Palepu, K.; Hong, L.; Mao, J.; Ye, T.; Iyer, R.; Zhao, L.; Chen, T.; Vincoff, S.; Watson, R.; et al. 2025. De novo design of peptide binders to conformationally diverse targets with contrastive language modeling. *Science Advances*, 11(4): eadr8638.
- Bodanszky, M. 1988. *Peptide chemistry*. Springer.
- Brandes, N.; Ofer, D.; Peleg, Y.; Rappoport, N.; and Linial, M. 2022. ProteinBERT: a universal deep-learning model of protein sequence and function. *Bioinformatics*, 38(8): 2102–2110.
- Chakravarty, D.; Lee, M.; and Porter, L. L. 2025. Proteins with alternative folds reveal blind spots in AlphaFold-based protein structure prediction. *Current Opinion in Structural Biology*, 90: 102973.
- Chen, T.; Dumas, M.; Watson, R.; Vincoff, S.; Peng, C.; Zhao, L.; Hong, L.; Pertsemlidis, S.; Shapers-Cheu, M.; Wang, T. Z.; et al. 2024a. PepMLM: target sequence-conditioned generation of therapeutic peptide binders via span masked language modeling. *ArXiv*, arXiv–2310.
- Chen, Z.; Wang, R.; Guo, J.; and Wang, X. 2024b. The role and future prospects of artificial intelligence algorithms in peptide drug development. *Biomedicine & Pharmacotherapy*, 175: 116709.
- Craik, D. J.; Fairlie, D. P.; Liras, S.; and Price, D. 2013. The future of peptide-based drugs. *Chemical biology & drug design*, 81(1): 136–147.
- Dang, C. V.; Reddy, E. P.; Shokat, K. M.; and Soucek, L. 2017. Drugging the ‘undruggable’ cancer targets. *Nature Reviews Cancer*, 17(8): 502–508.
- 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.
- Ferruz, N.; Schmidt, S.; and Höcker, B. 2022. ProtGPT2 is a deep unsupervised language model for protein design. *Nature communications*, 13(1): 4348.
- Fosgerau, K.; and Hoffmann, T. 2015a. Peptide therapeutics: current status and future directions. *Drug discovery today*, 20(1): 122–128.
- Fosgerau, K.; and Hoffmann, T. 2015b. Peptide therapeutics: current status and future directions. *Drug discovery today*, 20(1): 122–128.
- Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; and Li, W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23): 3150–3152.
- Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; and Bairoch, A. 2005. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, 571–607. Springer.
- Genheden, S.; and Ryde, U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, 10(5): 449–461.
- Gomes, B.; Augusto, M. T.; Felício, M. R.; Hollmann, A.; Franco, O. L.; Gonçalves, S.; and Santos, N. C. 2018. Designing improved active peptides for therapeutic approaches against infectious diseases. *Biotechnology advances*, 36(2): 415–429.
- Hayes, T.; Rao, R.; Akin, H.; Sofroniew, N. J.; Oktay, D.; Lin, Z.; Verkuil, R.; Tran, V. Q.; Deaton, J.; Wiggert, M.; et al. 2025. Simulating 500 million years of evolution with a language model. *Science*, 387(6736): 850–858.
- Henninot, A.; Collins, J. C.; and Nuss, J. M. 2018. The current state of peptide drug discovery: back to the future? *Journal of medicinal chemistry*, 61(4): 1382–1414.
- Hoogeboom, E.; Satorras, V. G.; Vignac, C.; and Welling, M. 2022. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, 8867–8887. PMLR.
- Jenson, J. M.; Xue, V.; Stretz, L.; Mandal, T.; Reich, L. and Keating, A. E. 2018. Peptide design by optimization on a data-parameterized protein interaction landscape. *Proceedings of the National Academy of Sciences*, 115(44): E10342–E10351.
- Jo, S.; Kim, T.; Iyer, V. G.; and Im, W. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*, 29(11): 1859–1865.
- Kaspar, A. A.; and Reichert, J. M. 2013. Future directions for peptide therapeutics development. *Drug discovery today*, 18(17-18): 807–817.
- Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K. W.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; et al. 2011. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. In *Methods in enzymology*, volume 487, 545–574. Elsevier.

- Lee, A. C.-L.; Harris, J. L.; Khanna, K. K.; and Hong, J.-H. 2019. A comprehensive review on current advances in peptide drug development and design. *International journal of molecular sciences*, 20(10): 2383.
- Lee, E. Y.; Lee, M. W.; Fulan, B. M.; Ferguson, A. L.; and Wong, G. C. 2017. What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface focus*, 7(6): 20160153.
- Lee, E. Y.; Wong, G. C.; and Ferguson, A. L. 2018. Machine learning-enabled discovery and design of membrane-active peptides. *Bioorganic & medicinal chemistry*, 26(10): 2708–2718.
- Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637): 1123–1130.
- Liu, S.; Li, Y.; Li, Z.; Gitter, A.; Zhu, Y.; Lu, J.; Xu, Z.; Nie, W.; Ramanathan, A.; Xiao, C.; et al. 2025. A text-guided protein design framework. *Nature Machine Intelligence*, 1–12.
- Lv, L.; Lin, Z.; Li, H.; Liu, Y.; Cui, J.; Chen, C. Y.-C.; Yuan, L.; and Tian, Y. 2025. Prollama: A protein large language model for multi-task protein language processing. *IEEE Transactions on Artificial Intelligence*.
- Madani, A.; Krause, B.; Greene, E. R.; Subramanian, S.; Mohr, B. P.; Holton, J. M.; Olmos Jr, J. L.; Xiong, C.; Sun, Z. Z.; Socher, R.; et al. 2023. Large language models generate functional protein sequences across diverse families. *Nature biotechnology*, 41(8): 1099–1106.
- Maiti, S.; Singh, A.; Maji, T.; Saibo, N. V.; and De, S. 2024. Experimental methods to study the structure and dynamics of intrinsically disordered regions in proteins. *Current Research in Structural Biology*, 7: 100138.
- Mooney, C.; Haslam, N. J.; Pollastri, G.; and Shields, D. C. 2012. Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity.
- Muttenthaler, M.; King, G. F.; Adams, D. J.; and Alewood, P. F. 2021a. Trends in peptide drug discovery. *Nature reviews Drug discovery*, 20(4): 309–325.
- Muttenthaler, M.; King, G. F.; Adams, D. J.; and Alewood, P. F. 2021b. Trends in peptide drug discovery. *Nature reviews Drug discovery*, 20(4): 309–325.
- Oldfield, C. J.; and Dunker, A. K. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual review of biochemistry*, 83(1): 553–584.
- Pereira, G. P.; Gouzien, C.; Souza, P. C.; and Martin, J. 2025. Challenges in predicting PROTAC-mediated protein-protein interfaces with AlphaFold reveal a general limitation on small interfaces. *Bioinformatics Advances*, 5(1): vbaf056.
- Shin, J.-E.; Riesselman, A. J.; Kollasch, A. W.; McMahan, C.; Simon, E.; Sander, C.; Manglik, A.; Kruse, A. C.; and Marks, D. S. 2021. Protein design and variant prediction using autoregressive generative models. *Nature communications*, 12(1): 2403.
- Song, Y.; and Ermon, S. 2019. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32.
- Trippe, B. L.; Yim, J.; Tischler, D.; Baker, D.; Broderick, T.; Barzilay, R.; and Jaakkola, T. 2022. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*.
- Trivedi, R.; and Nagarajaram, H. A. 2022. Intrinsically disordered proteins: an overview. *International journal of molecular sciences*, 23(22): 14050.
- Wang, F.; Wang, Y.; Feng, L.; Zhang, C.; and Lai, L. 2024. Target-specific de novo peptide binder design with DiffPep-Builder. *Journal of Chemical Information and Modeling*, 64(24): 9135–9149.
- Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; and Fu, C. 2022. Therapeutic peptides: current applications and future directions. *Signal transduction and targeted therapy*, 7(1): 48.
- 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. 2023. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976): 1089–1100.
- Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; and Zhou, Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485): 1444–1448.
- Yuan, X.; Lin, Z.; Kuen, J.; Zhang, J.; Wang, Y.; Maire, M.; Kale, A.; and Faieta, B. 2021. Multimodal contrastive training for visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6995–7004.
- Zhang, C.; Shine, M.; Pyle, A. M.; and Zhang, Y. 2022. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature methods*, 19(9): 1109–1115.
- Zhang, Z.; Xu, M.; Lozano, A.; Chenthamarakshan, V.; Das, P.; and Tang, J. 2023. Enhancing protein language model with structure-based encoder and pre-training. In *ICLR 2023-Machine Learning for Drug Discovery workshop*.