AMP-DiT: Antimicrobial Peptide Design with AMP-classifier Conditional Diffusion Transformers

Anonymous Authors[1]

Abstract

Existing AMP generation methods largely rely on protein language models such as ESM-2, which are trained on full-length proteins rather than peptide-specific distributions, leading to representation mismatch. Moreover, current approaches struggle to generate sequences that align with AMP classifier signals, despite activity prediction being the most critical objective in AMP design. We introduce AMP-DiT, a conditional discrete diffusion framework that generates peptides directly in sequence space, avoiding reliance on pre-trained protein language models. Our method uses classifier conditioning from *Macrel* (Santos-Júnior et al., 2020) to bias generation toward high antimicrobial activity, improving functional quality during sampling. Conditioning on a single predictor generalizes across multiple AMP prediction models, indicating that AMP-DiT captures underlying antimicrobial features rather than overfitting. Overall, AMP-DiT outperforms existing AMP design methods, in almost all of the AMP-classifer methods that are available, while keeping diversity.

1. Introduction

Biological background and problem statement. Antimicrobial resistance (AMR) is a major global health threat, motivating the development of new therapeutic strategies beyond conventional antibiotics. Antimicrobial peptides (AMPs) are short, diverse peptides with broad activity against bacteria, fungi, parasites, and viruses, making them promising candidates for combating drug-resistant pathogens. (Fjell et al., 2012) However, AMP design remains challenging because antimicrobial activity depends on a delicate balance of sequence composition, charge, hydrophobicity, amphipathicity, toxicity, stability, and target specificity. Traditional discovery pipelines rely heavily on high-throughput screening, rational modification, and post-generation filtering, which are costly and often inefficient. (Wan et al., 2024) These limitations highlight the need for controllable generative models that can directly produce peptide candidates aligned with antimicrobial activity and suitable for downstream experimental validation.

AI-based AMP generation and protein design. Recent advances in generative modeling have opened new opportunities for biological sequence and structure design. In protein design, diffusion-based methods such as RFdiffusion have demonstrated the ability to generate protein backbones, binders, and functional scaffolds by conditioning the denoising process on structural or functional constraints (Watson et al., 2023). More recently, binder design frameworks such as BoltzGen (Stark et al., 2025) have extended generative protein design toward all-atom protein and peptide binders, suggesting that generative models can be used not only to sample plausible biomolecules but also to optimize them for target-specific molecular recognition. These advances are particularly relevant to AMR, where resistance is often mediated by bacterial proteins such as enzymes, transporters, or regulatory factors; designing peptides that either act as AMPs or bind and inhibit AMR-related proteins could provide a powerful route toward new antimicrobial therapeutics.

In parallel, several deep learning based methods have been proposed specifically for AMP generation. AMP-Diffusion introduced a latent diffusion framework that uses ESM-2 representations to generate candidate antimicrobial peptides (Torres et al., 2025). OmegAMP proposed a controllable diffusion-based AMP design framework with biological conditioning and downstream classifier-based filtering (Soares et al., 2026). These methods demonstrate the promise of diffusion models for AMP discovery, but they also expose important limitations. Many existing AMP generators depend on protein language models such as ESM-2 (Lin et al., 2023), which are trained primarily on full-length proteins rather than peptide-specific distributions. This can introduce a representation mismatch, since AMPs are short, compositionally biased, often disordered, and governed by different

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

Submitted to the 2026 Workshop on Generative and Agentic AI for Biology (ICML 2026). Do not distribute.

functional constraints than globular proteins. Moreover, when generation occurs in continuous latent spaces, it may be harder to directly control discrete peptide sequences and align them with activity predictors.

Limitations of classifier-based evaluation and filtering. A central challenge in AMP generation is that antimicrobial activity is difficult to evaluate computationally in a standardized way. Most generative AMP methods use their own activity predictors, discriminators, or filtering pipelines to select promising candidates after generation.(Szymczak et al., 2023; Soares et al., 2026) While such classifiers are useful, this creates a comparison problem: each method may appear strong under the classifier it was optimized or filtered with, but weaker under independent predictors. This can lead to biased filtering, inflated in silico performance, and poor comparability across methods. Therefore, an important goal for AMP generation is not only to produce diverse and novel sequences, but also to generate peptides that consistently align with independent AMP classifier signals. Since predicted AMP activity is one of the most important computational metrics before wet-lab testing, incorporating classifier guidance directly into the generation process may be more effective than relying only on post hoc filtering.

AMP-DiT In this work, we introduce **AMP-DiT**, a conditional discrete diffusion transformer for antimicrobial peptide generation. Unlike methods that sample from protein language model latent spaces, AMP-DiT generates peptide sequences directly in discrete token space and is trained on AMP-specific data. To improve functional alignment, AMP-DiT incorporates conditioning from an external AMP predictor, Macrel (Santos-Júnior et al., 2020), during generation, explicitly steering the denoising process toward sequences with high predicted antimicrobial activity.

Our key hypothesis is that AMP activity should not be treated only as a post-generation filtering criterion, but as a direct conditioning signal during peptide generation. By conditioning the diffusion process with classifier values, AMP-DiT biases sampling toward high-fitness antimicrobial peptides while maintaining sequence diversity. We further show that conditioning on a single AMP predictor improves not only the conditioning classifier score but also scores from other independent AMP prediction models, suggesting that AMP-DiT captures general antimicrobial features rather than simply overfitting to one predictor. AMP-DiT is a class-conditional diffusion model that uses a transformer backbone That is inspired by the DiT model (Peebles & Xie, 2022).

2. Methods

2.1. AMP-DiT

We model peptide design as the generation of a discrete token sequence $x_0 \in \{0, \dots, N-1\}^L$, where L is a fixed maximum sequence length and N is the token vocabulary size. Sequences are encoded at the character level and padded to length L.

Data and Conditioning Signal For each training peptide, we compute two external biological scores using *Macrel*: (i) AMP probability and (ii) a non-hemolytic safety score.(Santos-Júnior et al., 2020) The second score is oriented so that larger values indicate lower predicted hemolytic risk, not higher hemolysis. These values form a conditioning vector $c \in \mathbb{R}^2$, which is used during training as a soft attribute signal.

Forward Diffusion Process We employ a D3PM-style (Austin et al., 2021) categorical noising process with T diffusion steps. At each step t, a transition matrix Q_t mixes each token with a near-uniform distribution according to a cosine schedule β_t. The cumulative transition $\bar{Q}_t$ defines:

$$q(x_t \mid x_0) = \mathrm{Cat}(x_t; \bar{Q}_t x_0).$$

Sampling from this categorical distribution is implemented via the Gumbel-max trick (Jang et al., 2017).

Reverse Model The reverse transition is modeled by a transformer denoiser conditioned on x_t, t, and c. It predicts logits for the clean sequence:

$$\hat{x}_0 = f_\theta(x_t, t, c).$$

Training Objective At each iteration, we sample a timestep t, corrupt $x_0 \to x_t$, and predict $\hat{x}_0$. The model is trained using a hybrid objective:

- Variational bound term: KL divergence between $q(x_{t-1} \mid x_t, x_0)$ and $q(x_{t-1} \mid x_t, \hat{x}_0)$,
- Cross-entropy loss on predicted token logits.

The total loss is:

$$\mathcal{L} = \lambda_{\mathrm{vb}} \mathcal{L}_{\mathrm{vb}} + \mathcal{L}_{\mathrm{CE}}.$$

Sampling and Inference Generation starts from random discrete noise $x_T \sim \mathrm{Unif}(\{0, \dots, N-1\}^L)$ and iteratively applies the reverse transitions from $t = T$ to 1. Generated sequences are decoded to characters. In our current implementation, conditioning is fixed to $c = [1, 1]$ for all samples, corresponding to high antimicrobial activity and high non-hemolytic safety.

Algorithm 1 Conditional Discrete Diffusion for AMP Generation

Require: Dataset $\mathcal{D} = \{(x_0, c)\}$, diffusion steps T

for each training step **do**
  Sample $(x_0, c) \sim \mathcal{D}$
  Sample timestep $t \sim \mathcal{U}(1, T)$
  Sample noisy sequence $x_t \sim q(x_t \mid x_0)$
  Predict $\hat{x}_0 = f_\theta(x_t, t, c)$
  Compute loss:

$$\mathcal{L} = \lambda_{\text{vb}} \mathcal{L}_{\text{vb}} + \mathcal{L}_{\text{CE}}$$

  Update θ via gradient descent
end for
Sampling:
Initialize $x_T \sim \text{Unif}(\{0, \ldots, N-1\}^L)$
for $t = T$ to 1 **do**
  Sample $x_{t-1} \sim p_\theta(x_{t-1} \mid x_t, c)$
end for
Return x_0

2.2. Classifier Conditioning

Diffusion models need to process conditional inputs, like diffusion timesteps or class labels. we used the transformer block with adaptive layer norm layers (adaLN) same as the DiT model. Importantly, these adaLN layers also modulate the activations immediately prior to any residual connections within the block, and are initialized such that each ViT block is the identity function. the architecture is shown in the figure 1. This change was the only one we needed to get good performance; otherwise, AMP-DiT is a fairly standard transformer model or the architecture is mostly same as AMP-Diffuion paper (Torres et al., 2025).

2.3. Evaluation

Generated peptides are assessed along three axes: *activity and safety predictors*, *sequence-level statistics*, and *language-model plausibility*.

- **Predictors:** Macrel (AMP probability and non-hemolytic safety score), AMPScanner, ToxinPred, HydrAMP-style MIC estimates, Amplify, and APEX where applicable;
- **Statistics:** diversity, entropy, k-mer composition, amino-acid frequency, and ModlAMP-derived descriptors;
- **Language models:** pseudo-perplexity under ProGen2 and a fitness score computed from amino-acid hydrophobicity scales (as in our analysis scripts).

more details in the appendix.

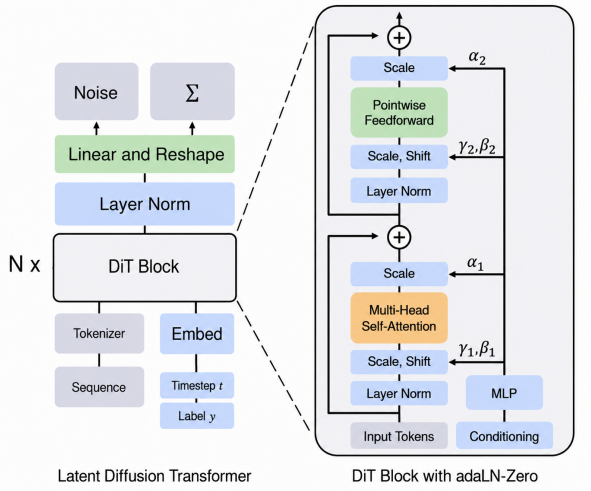

Figure 1. Architecture of AMP-DiT. Peptide sequences are tokenized and corrupted through a categorical diffusion process. A DiT-style transformer denoiser predicts clean amino-acid tokens from noisy tokens, timestep embeddings, and biological conditioning values. AMP probability and hemolysis-related scores are embedded through an MLP and injected into each transformer block using adaLN-Zero modulation.

3. Dataset

We use the Therapeutic Peptides dataset (Xiao et al., 2025) for training and evaluation. we used only common Amino Acids and the length is less than 30 and higher than 5.

4. Results

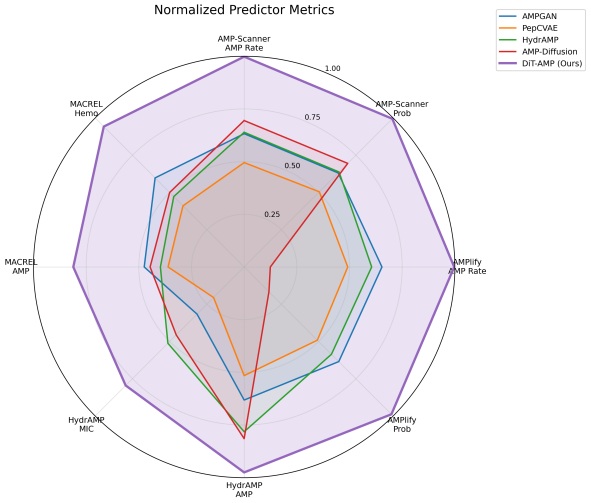

Figure 2. Comparison of AMP generation methods across multiple external predictors.

For in silico evaluation, we benchmark generated peptides using multiple independent AMP prediction models, including AMP-Scanner, AMPlify, APEX, HydrAMP, and Macrel.

Table 1. Comparison of AMP generation models across diversity, novelty, and fitness metrics.

Method	Diversity	Uniq.	Novelty	Fitness	Entropy μ	JS-3	JS-6
AMPGAN	41	99	100	0.12	2.90	0.89	0.013
PepCVAE	30	**100**	**100**	0.07	3.12	**0.89**	0.009
HydrAMP	30	**100**	**100**	0.09	2.82	0.89	0.011
AMP-Diffusion	36	90	100	0.10	**3.17**	0.79	0.018
DiT-AMP (Ours)	**47**	**100**	**100**	**0.14**	2.80	0.37	**0.022**

Table 2. Comparison of AMP generation methods across multiple evaluation metrics. For $\text{MACREL}_{\text{Hemo}}$, higher values denote greater non-hemolytic safety, i.e., lower predicted hemolytic risk.

Method	$\text{AMP-Scanner}_{\text{AMP Rate}}$	$\text{AMP-Scanner}_{\text{Prob}}$	$\text{AMPlify}_{\text{AMP Rate}}$	$\text{AMPlify}_{\text{Prob}}$	APEX_{MIC}	$\text{pLDDT}_{\text{mean}}$	$\text{HydrAMP}_{\text{AMP}}$	$\text{HydrAMP}_{\text{MIC}}$	$\text{MACREL}_{\text{AMP}}$	$\text{MACREL}_{\text{Hemo}}$
AMPGAN	0.63	0.63	0.65	0.64	344	74.03	0.63	0.32	0.47	0.60
PepCVAE	0.50	0.50	0.49	0.49	426	72.11	0.52	0.21	0.36	0.41
HydrAMP	0.64	0.64	0.61	0.59	361	71.74	0.78	0.51	0.40	0.47
AMP-Diffusion	0.69	0.70	0.12	0.17	413	71.76	0.81	0.46	0.45	0.50
DiT-AMP (Ours)	**1.00**	**0.99**	**1.00**	**0.99**	**236**	**75.25**	**0.97**	**0.80**	**0.81**	**0.94**

These models provide complementary measures of peptide quality: AMP-Scanner and AMPlify estimate antimicrobial likelihood, APEX and HydrAMP provide MIC-related activity estimates, and Macrel provides both AMP activity and a non-hemolytic safety score. Reporting performance across this classifier panel allows us to assess whether generated sequences consistently exhibit antimicrobial properties rather than being optimized for a single scoring function. This is particularly important because AMP generators are often evaluated using method-specific classifiers, which can introduce biased filtering and make comparisons across methods unreliable. By using multiple external predictors, we obtain a more stringent and comparable evaluation of functional alignment, potency, and potential toxicity.

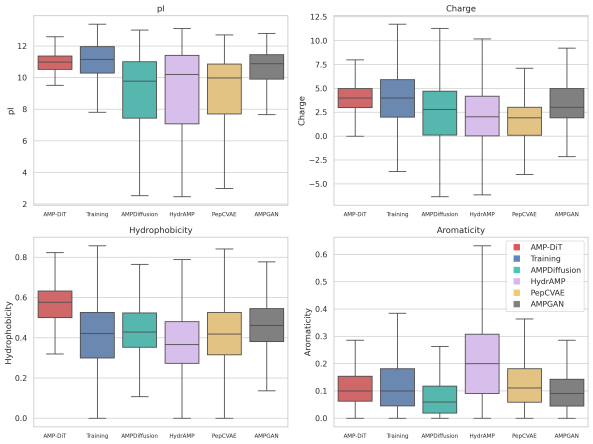

Figure 3. Physicochemical property analysis. the labels are the attributes of the peptides.

The physicochemical property analysis is depicted in the figure 3. The AMP-Diffusion model generates peptides with isoelectric point (pI) and charge distributions that are closely aligned with those of the training dataset, exhibiting a broad interquartile range. This suggests that AMP-Diffusion can generate a diverse array of peptides with pI values that are potentially advantageous for biological activity. In contrast, HydrAMP and AMPGAN exhibit higher pI and charge values; however, these largely fall outside the distribution of natural AMPs, which may put into question their biological relevance. PepCVAE demonstrates a performance comparable to AMP-Diffusion in terms of pI and charge. Regarding hydrophobicity and aromaticity, AMP-Diffusion and other models generally match the distribution of real AMPs, with the exception of HydrAMP. Notably, the peptides from AMP-Diffusion display a wide range of hydrophobicity and aromaticity, indicating the model's ability to synthesize peptides that extend beyond the diversity found in natural AMPs. This expanded range could potentially lead to novel peptides with unique properties suitable for various therapeutic applications.

5. Conclusion

We presented a conditional discrete diffusion framework for antimicrobial peptide design, where a DiT-style denoiser learns reverse transitions in sequence space under biological conditioning from Macrel-derived AMP and non-hemolytic safety scores. Across benchmark metrics, our model maintains high novelty while improving key generation properties such as diversity and fitness relative to baseline methods. Complementary analyses with external AMP predictors and physicochemical descriptors further indicate that generated sequences are both functionally plausible and aligned with natural AMP-like profiles. These results support discrete, controllable diffusion as a practical foundation for AMP discovery. Future work will focus on wet-lab validation, and specific target specific bacteria and extension of this framework to broader protein and peptide engineering tasks.

Impact Statement

Our work on reliable conditional generation of AMPs has the potential to advance antimicrobial discovery, especially in low-data regimes. However, the ability to generate novel bioactive sequences could be misused to design harmful peptides. We do not intend for our research to be used in such a manner and encourage responsible applications aligned with public health and safety. And also our approach goes trough using existing AMP prediction methods to evaluate the generated peptides that makes the comparison more fair and rigorous.

References

Austin, J., Johnson, D. D., Ho, J., Tarlow, D., and Van Den Berg, R. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993, 2021.

Fjell, C. D., Hiss, J. A., Hancock, R. E. W., and Schneider, G. Designing antimicrobial peptides: form follows function. *Nature Reviews Drug Discovery*, 11(1):37–51, 2012. doi: 10.1038/nrd3591.

Jang, E., Gu, S., and Poole, B. Categorical reparameterization with gumbel-softmax. *International Conference on Learning Representations (ICLR)*, 2017.

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., and Rives, A. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574.

Needleman, S. B. and Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48 (3):443–453, 1970. doi: 10.1016/0022-2836(70)90057-4.

Peebles, W. and Xie, S. Scalable diffusion models with transformers. *arXiv preprint arXiv:2212.09748*, 2022.

Santos-Júnior, C. D., Pan, S., Zhao, X., and Coelho, L. P. Macrel: antimicrobial peptide screening in genomes and metagenomes. *PeerJ*, 8:e10555, 2020. doi: 10.7717/peerj.10555.

Soares, D., Hetzel, L., Szymczak, P., Torres, M. D. T., Sommer, J., de la Fuente-Nunez, C., Theis, F. J., Günnemann, S., and Szczurek, E. OmegAMP: Targeted AMP discovery via biologically informed generation, 2026. URL https://openreview.net/forum?id=dGE8GNOLKs.

Stark, H., Faltings, F., Choi, M., Xie, Y., Hur, E., O'Donnell, T., et al. Boltzgen: Toward universal binder design. *bioRxiv*, 2025. doi: 10.1101/2025.11.20.689494. URL https://doi.org/10.1101/2025.11.20.689494.

Szymczak, P., Możejko, M., Grzegorzek, T., Jurczak, R., Bauer, M., Neubauer, D., Sroka, J., Setny, P., Kamysz, W., and Szczurek, E. Discovering highly potent antimicrobial peptides with deep generative model HydrAMP. *Nature Communications*, 14(1):1453, 2023. doi: 10.1038/s41467-023-36994-z.

Torres, M. D. T., Chen, T., Wan, F., Chatterjee, P., and de la Fuente-Nunez, C. Generative latent diffusion language modeling yields anti-infective synthetic peptides. *bioRxiv*, 2025. doi: 10.1101/2025.01.31.636003. URL https://www.biorxiv.org/content/early/2025/02/01/2025.01.31.636003.

Wan, F., Wong, F., Collins, J. J., and de la Fuente-Nunez, C. Machine learning for antimicrobial peptide identification and design. *Nature Reviews Bioengineering*, 2(5):392–407, May 2024. doi: 10.1038/s44222-024-00152-x.

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., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., Lauko, A., De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T. S., DiMaio, F., Baek, M., and Baker, D. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023. doi: 10.1038/s41586-023-06415-8.

Xiao, B., Zhou, Y., Zhao, L., Huang, H., Fei, X., and Zhang, Y.-B. A comprehensive dataset of therapeutic peptides on multi-function property and structure information. *Scientific Data*, 12(1):1213, 2025. doi: 10.1038/s41597-025-05528-1. URL https://doi.org/10.1038/s41597-025-05528-1.

AA	h	h_x
A	0.25	0.00
R	-1.80	0.21
N	-0.64	0.65
D	-0.72	0.69
C	0.04	0.68
Q	-0.69	0.39
E	-0.62	0.40
G	0.16	1.00
H	-0.40	0.61
I	0.73	0.41
L	0.53	0.21
K	-1.10	0.26
M	0.26	0.24
F	0.61	0.54
P	-0.07	3.16
S	-0.26	0.50
T	-0.18	0.66
W	0.37	0.49
Y	0.02	0.53
V	0.54	0.61

Table 3. Hydrophobicity (h) and helix propensity (h_x) scales used in the fitness calculation.

A. Appendix

A.1. Metrics

For a generated set of sequences $S := \{s_i\}_{i=1}^{N}$, we compute diversity as the complement of the average normalized pairwise alignment between peptides. Alignment is defined as the largest ordered subset of characters shared by two sequences, following Needleman–Wunsch alignment (Needleman & Wunsch, 1970). Specifically,

$$\text{Diversity}(S) = 100 - 100\binom{N}{2}^{-1} \sum_{1 \le i < j \le N} \frac{\text{Alignment}(s_i, s_j)}{\min\left(\text{length}(s_i), \text{length}(s_j)\right)}.$$

We compute uniqueness as the percentage of distinct sequences in the generated set:

$$\text{Uniqueness}(S) = \frac{|\{s_i \mid s_i \notin \{s_1, \ldots, s_{i-1}\}\}|}{N} \times 100.$$

To measure novelty, let H denote the reference set of experimentally validated AMPs. Novelty is the percentage of generated sequences that do not overlap with H:

$$\text{Novelty}(S) = \frac{|\{s_i \mid s_i \notin H\}|}{N} \times 100.$$

For the fitness score, let $\theta := 100\pi/180$, i.e., the equivalent of 100° in radians, and let h and h_x denote the amino-acid scales in Table 3. For a sequence $(a_1, \ldots, a_L)$, we compute amphipathic hydrophobicity along a helical projection and normalize by helix propensity:

$$\text{Fitness}(S) = \frac{1}{N} \sum_{(a_1,\ldots,a_L) \in S} \frac{\sqrt{\left(\sum_{i=1}^{L} h(a_i)\cos(i\theta)\right)^2 + \left(\sum_{i=1}^{L} h(a_i)\sin(i\theta)\right)^2}}{\sum_{i=1}^{L} \exp\left(h_x(a_i)\right)}.$$