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

Alireza Noroozi, Ozlem keskin, Attila gursuy
Computational Sciences and Engineering, Koc University

Code project

genbio2026 publication

Introduction

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 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 across almost all available AMP-classifier metrics while maintaining diversity.

Method

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. The Macrel-derived AMP probability and non-hemolytic safety score are embedded through an MLP and injected into each transformer block using adaLN-Zero modulation.

Generation starts from random discrete noise 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.

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.

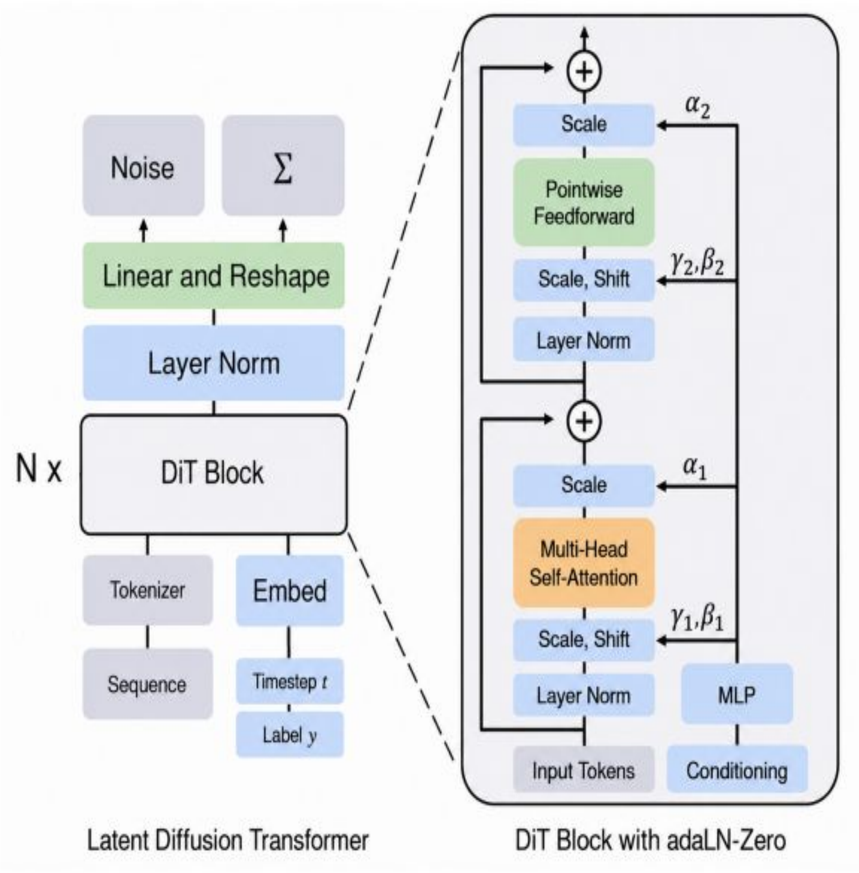
Algorithm 1 Conditional Discrete Diffusion for AMP Generation

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

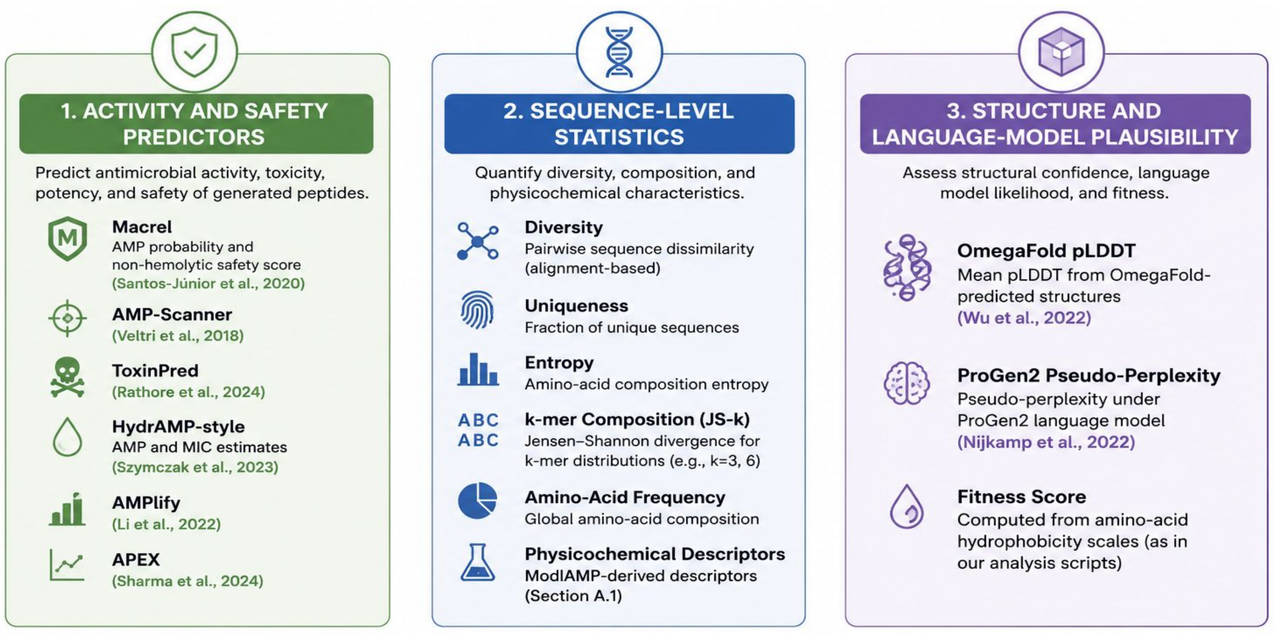
for each training step **do**
  Sample $(x_0, c) \sim \mathcal{D}$
  Sample timestep $t \sim \mathcal{U}(1, T)$
  Sample noisy sequence $x_t \sim q(x_t \mid x_0)$
  Predict $\hat{x}_0 = f_\theta(x_t, t, c)$
  Compute loss:

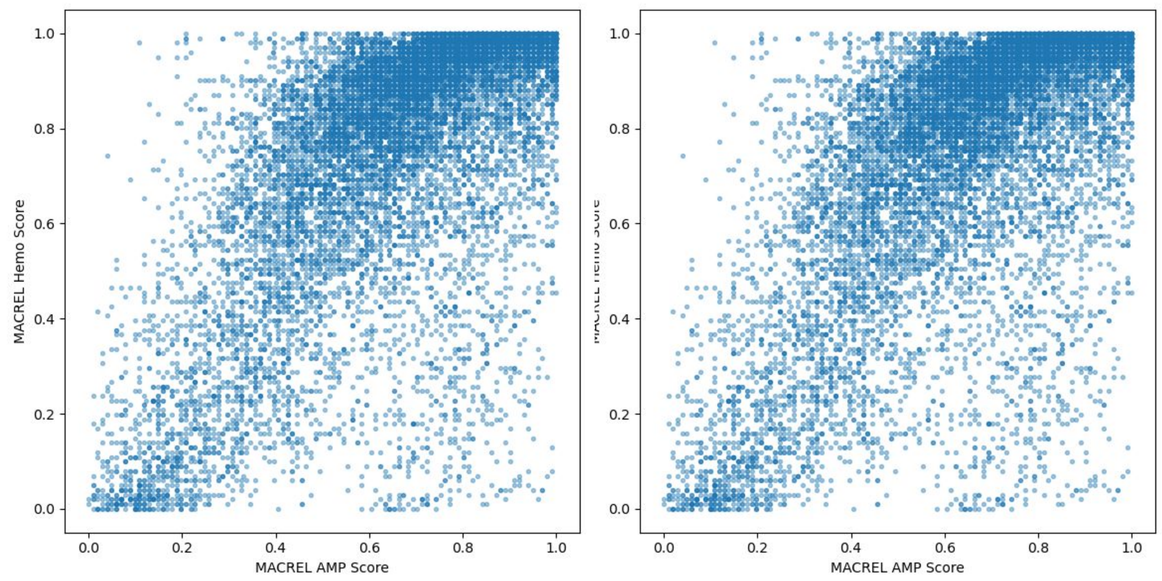
$$\mathcal{L} = \lambda_{\text{vb}}\mathcal{L}_{\text{vb}} + \mathcal{L}_{\text{CE}}$$

  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

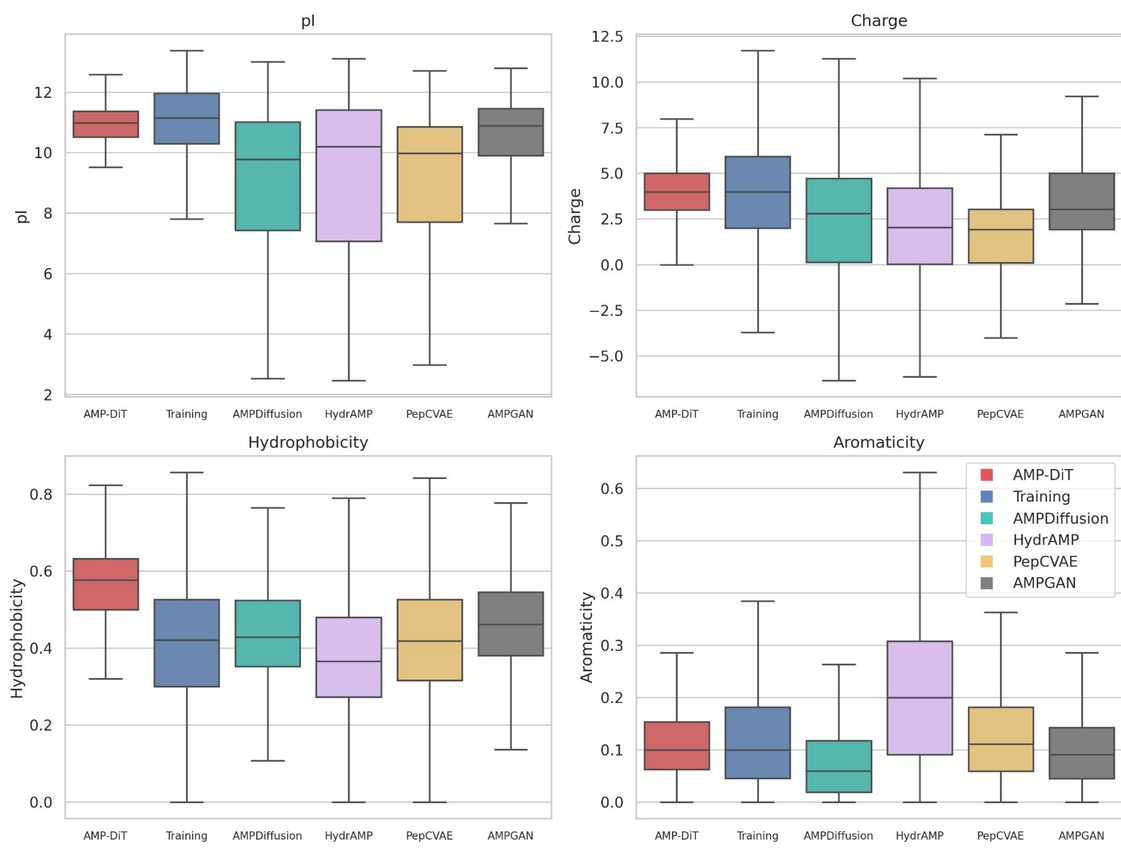


Evaluation Metrics





Training and sampling Macrel values distribution



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

Results

* AMP-DiT consistently outperforms baseline generative models across multiple independent AMP predictors, including AMP-Scanner, AMPlify, HydrAMP, APEX, and Macrel.
* Achieves near-perfect predicted AMP activity, with 1.00 AMP-Scanner and 1.00 AMPlify AMP rates.
* Shows strong predicted potency with HydrAMP AMP = 0.97 and HydrAMP MIC = 0.80, while achieving the lowest APEX MIC (236) among compared methods.
* Maintains 100% uniqueness and the highest diversity (47), indicating that improved activity does not come at the cost of sequence diversity.
* Conditioning only on Macrel also improves scores from independent predictors, suggesting that AMP-DiT learns general antimicrobial characteristics rather than simply overfitting to its conditioning classifier.

Method	AMP-Scanner$_{\text{AMP Rate}}$	AMP-Scanner$_{\text{Prob}}$	AMPlify$_{\text{AMP Rate}}$	AMPlify$_{\text{Prob}}$	APEX$_{\text{MIC}}$	OmegaFold pLDDT$_{\text{mean}}$	HydrAMP$_{\text{AMP}}$	HydrAMP$_{\text{MIC}}$	MACREL$_{\text{AMP}}$	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**

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 nonhemolytic safety scores. Across benchmark metrics, our model maintains high uniqueness 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, targeting specific bacteria, and extending this framework to broader protein and peptide engineering tasks.