Multi-Objective Peptide Design via Token-Aligned Preference Optimization

Michaela Areti Zervou[1,2] **Felix Teufel**[3] **Yannis Pantazis**[4]
Panagiotis Tsakalides[1,2] **Ole Winther**[3,5]

[1] Institute of Computer Science, FORTH
[2] Department of Computer Science, University of Crete
[3] Department of Biology, University of Copenhagen
[4] Institute of Applied and Computational Mathematics, FORTH
[5] Department of Applied Mathematics and Computer Science, Technical University of Denmark
zervou@ics.forth.gr

Abstract

Protein language models have recently shown promise for de novo protein and peptide design, but they lack mechanisms for controllable optimization of functional properties. This limitation is particularly critical in therapeutic peptide discovery, where candidates must simultaneously satisfy multiple, often conflicting, biochemical constraints. We present a token-aligned preference optimization framework that adapts a pretrained protein language model using pairwise sequence preferences conditioned on property-specific control tokens. By learning from comparative feedback rather than scalar rewards, our approach enables multi-objective control and shows generalization to novel property combinations. As a case study, we apply the method to antimicrobial peptide (AMP) design, a clinically relevant but challenging testbed. Our approach achieves substantial improvements in jointly satisfying biochemical constraints, demonstrating the potential of preference alignment for controllable peptide design.

1 Introduction

Designing peptides with precise biochemical properties remains a central challenge in therapeutic development. Antimicrobial peptides (AMPs) are particularly attractive due to their broad-spectrum activity, low propensity for resistance, and potential applications in biofilm disruption, immune modulation, and synergistic treatment with conventional antibiotics [1, 2]. Despite this promise, translating AMPs into viable therapeutics is difficult. Functional candidates must simultaneously satisfy multiple constraints, including strong antimicrobial activity, low host toxicity, high aqueous solubility, and physicochemical stability. These requirements often interact antagonistically. For example, increased hydrophobicity may promote membrane disruption but at the cost of higher toxicity, while enhanced stability can compromise solubility [3]. This inherent trade-off makes AMP design a fundamentally multi-objective problem.

Computational design offers a route to accelerate discovery and reduce the cost of experimental screening. Advances in protein language models (PLMs) trained on large sequence corpora have enabled the generation of syntactically valid and diverse peptides [4, 5, 6]. However, steering these models toward functional candidates remains challenging. Existing approaches tend to optimize a single objective [7, 8, 9, 10] or, at most, a narrow subset such as activity and toxicity [11, 12]. Multi-objective control is typically attempted through scalar reward signals or filtering after generation, both of which can be brittle to balancing conflicting constraints. Furthermore, adversarial or reinforcement-

Published in the AI for Science Workshop (NeurIPS 2025).

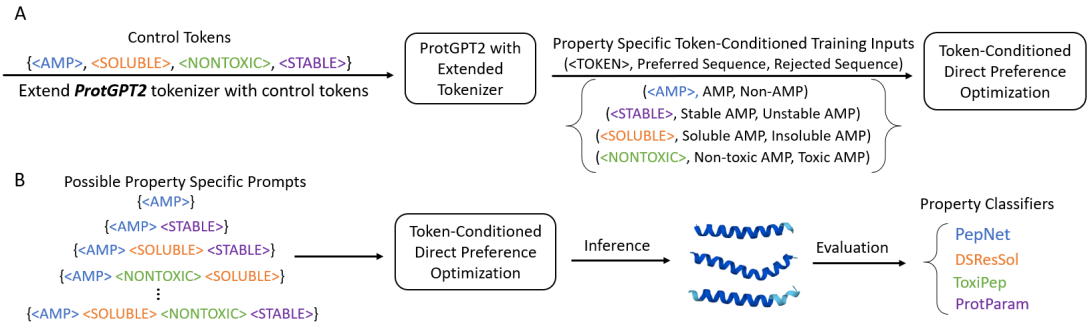

Figure 1: Multiobjective Token-Aligned DPO Pipeline. The pipeline consists of two stages. **(A)** Training: The pretrained ProtGPT2 model's tokenizer is extended with property-specific control tokens. It is then fine-tuned via Direct Preference Optimization (DPO) using paired preferences conditioned on these tokens. **(B)** Inference & Evaluation: The trained model generates peptides based on a prefix of one or more control tokens. The resulting peptides are then validated using external Property Classifiers to verify their properties.

based training strategies tend to introduce instability or reduce diversity [13, 14], making it hard to satisfy all objectives at once. As a result, current methods cannot reliably achieve simultaneous satisfaction of the full set of biochemical properties needed for therapeutic AMPs.

In this work, we introduce tDPO-ProtGPT2, a token-aligned preference optimization [15] framework for de novo AMP generation. As shown in Figure 1, our method aligns a pretrained autoregressive protein language model (ProtGPT2 [4]) with property-specific sequence preferences through the use of control tokens. This design enables direct control over multiple biochemical objectives during generation, including objectives not observed jointly during training. Our main contributions are:

- We introduce a novel framework that combines token-based control with a preference optimization approach to enable direct, multi-objective peptide design. This approach facilitates controllable generation via prompts and enables generalization to unseen property combinations.
- The proposed model outperforms baselines and prior AMP generators in joint constraint satisfaction.

To the best of our knowledge, this is the first method to combine token-based control with preference optimization for the simultaneous regulation of multiple biochemical properties in peptide design, and specifically in the context of de novo AMP generation. Although demonstrated on AMPs, the framework is potentially applicable to peptide and protein engineering tasks that require the joint control of multiple functional constraints.

2 Related Work

In this work, we perform multi-objective de novo AMP design by aligning a pretrained protein language model through token-conditioned preference optimization. This constitutes a prompting-style alignment method, and as such, we compare primarily against other prompting-based LLM approaches for AMP design. Most notably, AMP-Designer [8] adapts a general-purpose large language model with contrastive prompt tuning, knowledge distillation, and minimum inhibitory concentration (MIC)-guided reinforcement learning to achieve strong antimicrobial activity. However, this approach is predominantly single-objective. Considering multi-objective generative frameworks, we compare against MPOGAN [11], which employs property-specific discriminators within an adversarial training scheme to jointly optimize antimicrobial activity and cytotoxicity. More recently, HMAMP [12] introduced a hypervolume-driven reinforcement learning framework to optimize antimicrobial activity and hemolysis simultaneously, but its limited data availability and absence of complete code hinder reproducibility.

3 Methods

3.1 Dataset Construction

We assembled our dataset by aggregating experimentally validated AMP sequences from seven public databases: APD3 [16], DRAMP [17], CAMP [18], LAMP2 [19], dbAMP [20], YADAMP [21], and XUAMP [22]. After deduplication and standardization, this initial collection contained approximately 28,000 unique AMP sequences. To ensure biological and therapeutic relevance, we filtered sequences to retain only those between 5 and 50 amino acids in length, excluding both short fragments and longer peptides that may exhibit protein-like properties. This reduced the dataset to approximately 9,100 sequences which were then annotated using predictive tools for toxicity (ToxinPred3 [23]), and solubility (CamSol [24]). For stability, we applied ProtParam [25], classifying peptides with values below 40 as stable [26]. From this set, we retained the 1,191 AMP sequences that were predicted to be simultaneously non-toxic, soluble, and stable. We note that since these labels are derived from computational predictions rather than experimental measurements, serve only as proxies for true biological properties and inevitably introduce some degree of noise and potential bias into the training signal. The reported accuracies of each classifier are provided in Appendix Section A.1.

To train the model, we constructed preference triplets of the form (token, preferred sequence, rejected sequence) for four objectives: AMP-likeness, toxicity, solubility, and stability. Each of the 1,191 AMPs is labeled with all four properties, allowing us to derive a distinct preference pair for each property. In effect, this yields four parallel preference datasets of equal size, one per objective, which together form the basis for multi-objective training. For AMP-likeness, negative sequences were sampled from UniProt, matched in length but not annotated as antimicrobial peptides. To reduce redundancy, non-AMPs were clustered using MMseqs2 [27] at 50% sequence identity, and one representative per cluster was retained to ensure diversity. For the remaining properties, both preferred and rejected sequences were AMPs, where the preferred sequence satisfied the target property, and the rejected one did not. For example, in the toxicity objective, the preferred sequence was a non-toxic AMP, while the rejected sequence was a toxic AMP.

3.2 Token-Conditioned Preference Optimization

Problem setup. Peptide generation is modeled as conditional sequence generation. Let Σ denote the amino acid alphabet and $y \in \Sigma^{\leq T}$ a peptide sequence. To control biochemical properties, we extend the tokenizer with the discrete control tokens

$$\mathcal{T} = \{\texttt{<AMP>}, \texttt{<NONTOXIC>}, \texttt{<SOLUBLE>}, \texttt{<STABLE>}\},$$

corresponding to antimicrobial activity, non-toxicity, solubility, and stability. Given a token $t \in \mathcal{T}$, the model defines a conditional policy $\pi_\theta(y \mid t)$. Each training example is a triplet (t, y^+, y^-), where y^+ is preferred to y^- under the property indicated by t. The objective is to adapt π_θ so that y^+ receives higher likelihood than y^- while remaining close to a pretrained reference distribution π_{ref}.

Direct Preference Optimization. We adopt Direct Preference Optimization (DPO) [15]. For each triplet (t, y^+, y^-) the loss is

$$\mathcal{L}_{\text{DPO}}(t, y^+, y^-) = -\log \sigma(\beta_t \left[\Delta_\theta(t) - \Delta_{\text{ref}}(t)\right]),$$

with

$$\Delta_\theta(t) = \log \pi_\theta(y^+ \mid t) - \log \pi_\theta(y^- \mid t), \quad \Delta_{\text{ref}}(t) = \log \pi_{\text{ref}}(y^+ \mid t) - \log \pi_{\text{ref}}(y^- \mid t).$$

Here $\log \pi_\theta(y \mid t)$ is the sequence log-likelihood, computed as the sum of token log-probabilities when the control token t is prepended. Control tokens and padding symbols are masked from the loss so that gradients are computed only on peptide tokens.

Multi-objective training. The dataset provides triplets (t, y^+, y^-) for all $t \in \mathcal{T}$. Mini-batches are stratified across tokens to ensure balanced training. To balance heterogeneous supervision signals, a token-specific scaling factor β_t is introduced. For more details, see Appendix Section A.2. The full training objective is

$$\mathcal{L} = \frac{1}{|\mathcal{B}|} \sum_{(t, y^+, y^-) \in \mathcal{B}} -\log \sigma(\beta_t \left[\Delta_\theta(t) - \Delta_{\text{ref}}(t)\right]),$$

where $\mathcal{B}$ is a balanced batch across $\mathcal{T}$.

Table 1: Diversity, novelty, and entropy of generated sequences. Novelty is defined as the percentage of sequences with less than 80% identity to training data.

Model	Diversity	Novelty (80%)	Entropy
MPOGAN	**83.5%**	96.1%	65.1%
AMP-Designer (n-tokens)	82.2%	99.9%	66.4%
AMP-Designer (top-k)	78.1%	**100%**	59.3%
ProtGPT2	75.1%	**100%**	56.4%
SFT-ProtGPT2	80.1%	**100%**	60.5%
DPO-ProtGPT2	70.9%	**100%**	66.5%
tDPO-ProtGPT2 (This work)	71.8%	**100%**	**69.2%**

Inference. At inference, conditioning is specified by prefixing the desired control token or a concatenation of tokens. For example, `<AMP>` directs generation toward antimicrobial peptides, while `<AMP> <NONTOXIC>` produces peptides that are simultaneously antimicrobial and non-toxic. Because optimization is performed jointly across tokens, the model generalizes to compositional prompts without additional finetuning. Sequences are decoded using stochastic sampling with fixed hyperparameters for comparability (temperature = 1.0, top-p = 0.95, top-k = 50).

3.3 Baselines

Our evaluation focuses on baselines directly relevant to multi-objective generation and prompting-based language models for de novo AMP design. We include MPOGAN and AMP-Designer as primary points of comparison. For AMP-Designer, we evaluate two modes described in the original work: (i) AMP-Prompt-TopK, which integrates prompting with top-k sampling, and (ii) AMP-Prompt (n-tokens), which applies contrastive prompt tuning with a fixed set of property tokens. As additional baselines, we consider: (i) ProtGPT2 in a zero-shot setting to assess whether a general protein language model exhibits inherent controllability without task-specific alignment, (ii) SFT-ProtGPT2, obtained by supervised fine-tuning on AMPs that jointly satisfy non-toxicity, solubility, and stability, and (iii) DPO-ProtGPT2, trained with unconditioned preference optimization. For each model, we generate and evaluate 1,000 peptide sequences under identical sampling and evaluation protocols.

4 Results

4.1 Sequence Novelty, Diversity, and Entropy

Table 1 summarizes sequence-level diversity, novelty, and entropy across models. Diversity captures the proportion of unique sequences, novelty is defined as the percentage of sequences with less than 80% identity to training data, and entropy measures the token-level distributional spread across generated outputs. For fairness, all baselines (ProtGPT2, SFT-ProtGPT2, and DPO-ProtGPT2) are assessed using identical decoding parameters, ensuring that diversity and entropy-related metrics are directly comparable.

MPOGAN achieves the highest diversity score (83.5%), indicating a broad exploration of peptide space. In contrast, our tDPO-ProtGPT2 model shows a diversity of 71.8%. The plain DPO-ProtGPT2 variant also exhibits a similar diversity of 70.9%, which is a consequence of the stronger conditioning imposed by property-specific control tokens. While diversity is somewhat reduced due to targeting specific property combinations, this is an expected outcome of focusing the model on feasible regions of sequence space. Novelty remains at 100% across all deep learning models, confirming that none of the approaches rely on rote memorization of the training data. This suggests that preference optimization and token conditioning preserve generative originality.

Entropy was computed as the normalized Shannon entropy of each peptide sequence, scaled to [0,1] by the maximum possible entropy of the amino acid alphabet. tDPO-ProtGPT2 achieved the highest mean entropy (69.2%), clearly higher than SFT-ProtGPT2 (60.5%) and modestly above plain DPO-ProtGPT2 (66.5%). This suggests that token conditioning maintains local variability relative to preference optimization alone, while also avoiding the reduction in variability observed under supervised fine-tuning. For comparison, MPOGAN achieved higher sequence-level diversity but a

lower entropy score, indicating that uniqueness across sequences does not necessarily correspond to greater token-level variability.

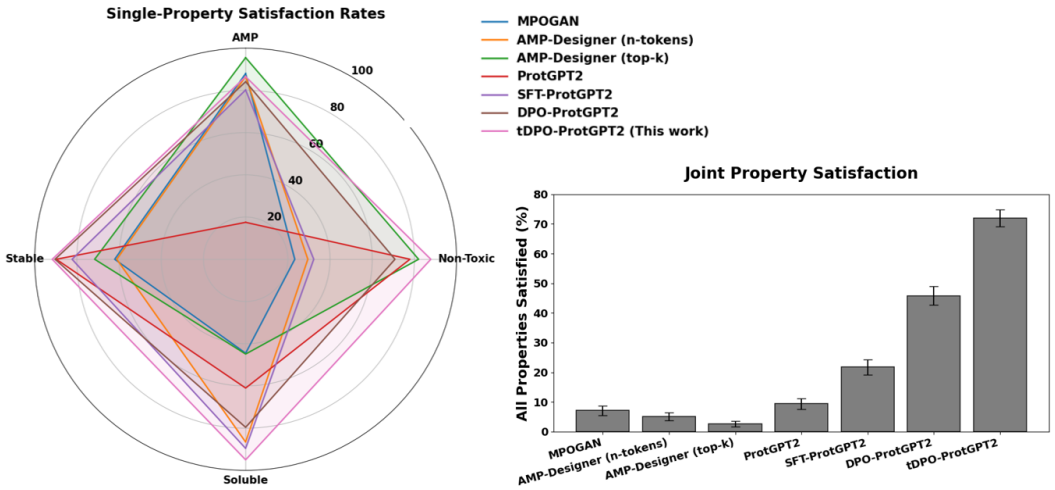

Figure 2: Single-property satisfaction rates (left) and joint satisfaction across all four properties (right) for the compared models. Error bars indicate 95% binomial confidence intervals (n = 1000).

4.2 Property Satisfaction Analysis

We evaluated the quality of the generated peptides on four properties: antimicrobial activity, toxicity, solubility, and physicochemical stability. The first three were evaluated using leading predictive models such as PepNet [28], ToxiPep [29], and DSResSol [30]. These models differ from the classifiers employed during dataset construction to ensure unbiased validation. For stability, we applied the ProtParam instability index, keeping the stability threshold consistent between training and evaluation so that results are directly comparable. To complement these property-based evaluations, we also assessed the structural plausibility of representative peptides generated by our model using AlphaFold3 [31]. While not a primary benchmark, these visualizations provide qualitative confirmation that the designed sequences can adopt realistic conformations. The resulting structures are included in Appendix A.3.

Figure 2 presents a comparative evaluation of single- and multi-objective performance. The radar plot on the left visualizes single-property performance across the four biochemical objectives. It is important to note that our model is evaluated on all four properties simultaneously. As provided, baselines such as MPOGAN and AMP-Designer demonstrate uneven performance, highlighting the difficulty of achieving a balanced trade-off in multi-objective design. In contrast, our tDPO-ProtGPT2 model's profile expands uniformly across all four axes, demonstrating consistent alignment with each objective simultaneously.

The bar plot on the right of Figure 2 captures the stricter criterion of joint satisfaction, measuring the percentage of peptides that simultaneously fulfill all four biochemical constraints. Baselines such as MPOGAN, AMP-Designer, and ProtGPT2 remain below 10%, confirming that methods optimized for single objectives do not generalize well to complex multi-objective scenarios. SFT-ProtGPT2 and plain DPO-ProtGPT2 show significant improvements, reaching approximately 21.8% and 45.8% respectively, but still fail to ensure reliable simultaneous satisfaction. In sharp contrast, tDPO-ProtGPT2 reaches 72%, markedly higher than all other methods, underlining its superior performance in joint optimization.

The large performance gap between plain DPO and token-conditioned DPO highlights a key advantage of our method. In the token-free setting, all preference signals are aggregated into a single model distribution, which makes it difficult to preserve information about distinct objectives. This leads to a compromise solution that, while better than other baselines, still struggles with joint satisfaction. In contrast, explicit control tokens partition the preference space, providing a context that isolates supervision for each property. This structure allows the model to align with multiple objectives

simultaneously rather than collapsing to partial compromises. Moreover, even when prompted with all four objectives, tDPO-ProtGPT2 performs comparably to baselines optimized for single properties, in some cases matching or slightly exceeding them.

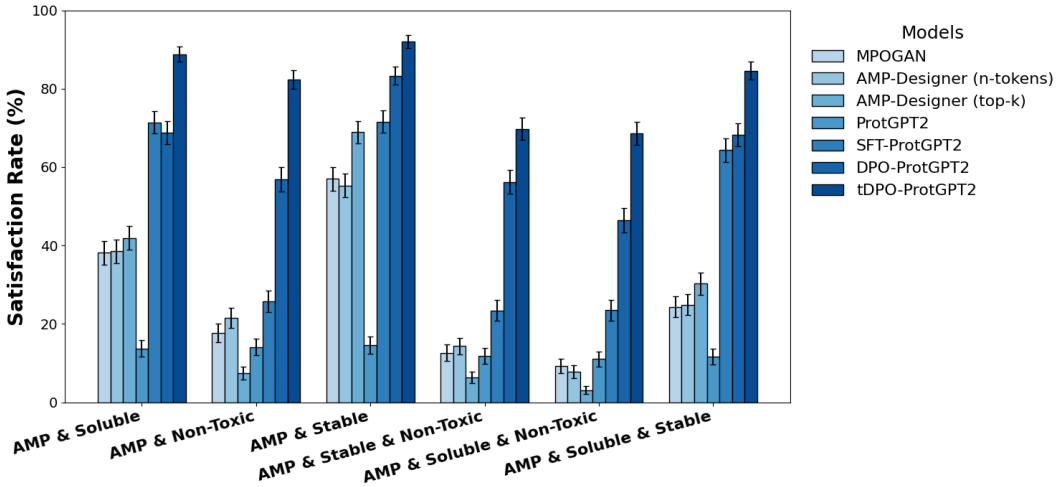

Figure 3: Multi-property satisfaction rates across models. Bars show the percentage of generated peptides satisfying pairwise and triple combinations of biochemical properties, with 95% binomial confidence intervals (n=1000).

4.3 Satisfaction under Combined Property Constraints

Unlike existing baselines, which cannot be prompted for arbitrary combinations of objectives, our token-conditioned model can be explicitly directed toward subsets of constraints. Evaluating these intermediate cases shows how performance scales as more biochemical requirements are added, and highlights that tDPO-ProtGPT2 generalizes beyond single-objective or full four-objective settings.

As shown in Figure 3, most baselines remain limited even when only two or three objectives are imposed simultaneously. MPOGAN and AMP-Designer achieve moderate success on pairs (30–50%) but drop below 10–15% on any triple combination. ProtGPT2 performs uniformly poorly across all settings, confirming the lack of controllability without fine-tuning. SFT-ProtGPT2 performs substantially better, especially for two-property combinations, but satisfaction declines rapidly when a third property is introduced, revealing limited compositional generalization. Plain DPO-ProtGPT2 shows stronger consistency, with scores in the 50–80% range across pairs and three-way combinations, but it still exhibits degradation as constraints accumulate. By contrast, tDPO-ProtGPT2 maintains high performance even as requirements compound. Across two-property combinations, satisfaction exceeds 80–90%, and critically, the model sustains 68–70% satisfaction even on three-property subsets.

5 Conclusion

This work introduces tDPO-ProtGPT2, a token-aligned preference optimization framework that achieves multi-objective peptide generation. By conditioning a pretrained protein language model on property-specific tokens and aligning it through comparative preferences, we demonstrate simultaneous optimization across antimicrobial activity, toxicity, solubility, and stability. Empirical results show that our approach outperforms prior methods, achieving much higher rates of joint constraint satisfaction while maintaining sequence diversity and comparable entropy. The ability to generalize to unseen property combinations demonstrates the flexibility of token-conditioned preference alignment, indicating its promise for real-world peptide discovery where multiple requirements must be met. A limitation of this study is that preference pairs were constructed using predictive classifiers, which inevitably introduce error and may not perfectly reflect biological ground truth. While our results highlight the potential of token-conditioned preference optimization, experimental validation will be required to fully establish reliability.

6 Code and Data Availability

Code and data is available at `https://github.com/aretiz/tDPO`.

Acknowledgments and Disclosure of Funding

This work is supported by the TITAN ERA Chair project (contract no. 101086741) within the Horizon Europe Framework Program of the European Commission.

References

[1] Maria Magana, Muthuirulan Pushpanathan, Ana L Santos, Leon Leanse, Michael Fernandez, Anastasios Ioannidis, Marc A Giulianotti, Yiorgos Apidianakis, Steven Bradfute, Andrew L Ferguson, et al. The value of antimicrobial peptides in the age of resistance. *The lancet infectious diseases*, 20(9):e216–e230, 2020.

[2] Sainan Zheng, Yuhan Tu, Bin Li, Gaoer Qu, Anqi Li, Xuemei Peng, Shijun Li, and Chuanfeng Shao. Antimicrobial peptide biological activity, delivery systems and clinical translation status and challenges. *Journal of Translational Medicine*, 23(1):292, 2025.

[3] Axel Hollmann, Melina Martínez, Martín E Noguera, Marcelo T Augusto, Anibal Disalvo, Nuno C Santos, Liliana Semorile, and Paulo C Maffía. Role of amphipathicity and hydrophobicity in the balance between hemolysis and peptide–membrane interactions of three related antimicrobial peptides. *Colloids and Surfaces B: Biointerfaces*, 141:528–536, 2016.

[4] Noelia Ferruz, Steffen Schmidt, and Birte Höcker. Protgpt2 is a deep unsupervised language model for protein design. *Nature communications*, 13(1):4348, 2022.

[5] Ali Madani, Ben Krause, Eric R Greene, Subu Subramanian, Benjamin P Mohr, James M Holton, Jose Luis Olmos Jr, Caiming Xiong, Zachary Z Sun, Richard Socher, et al. Large language models generate functional protein sequences across diverse families. *Nature biotechnology*, 41(8):1099–1106, 2023.

[6] Talal Widatalla, Rafael Rafailov, and Brian Hie. Aligning protein generative models with experimental fitness via direct preference optimization. *bioRxiv*, pages 2024–05, 2024.

[7] Paulina Szymczak, Marcin Możejko, Tomasz Grzegorzek, Radosław Jurczak, Marta Bauer, Damian Neubauer, Karol Sikora, Michał Michalski, Jacek Sroka, Piotr Setny, et al. Discovering highly potent antimicrobial peptides with deep generative model hydramp. *nature communications*, 14(1):1453, 2023.

[8] Jike Wang, Jianwen Feng, Yu Kang, Peichen Pan, Jingxuan Ge, Yan Wang, Mingyang Wang, Zhenxing Wu, Xingcai Zhang, Jiameng Yu, et al. Discovery of novel antimicrobial peptides with notable antibacterial potency by a llm-based foundation model. *arXiv preprint arXiv:2407.12296*, 2024.

[9] Shuwen Jin, Zihan Zeng, Xiyan Xiong, Baicheng Huang, Li Tang, Hongsheng Wang, Xiao Ma, Xiaochun Tang, Guoqing Shao, Xingxu Huang, et al. Ampgen: an evolutionary information-reserved and diffusion-driven generative model for de novo design of antimicrobial peptides. *Communications Biology*, 8(1):839, 2025.

[10] Diogo Soares, Leon Hetzel, Paulina Szymczak, Fabian Theis, Stephan Günnemann, and Ewa Szczurek. Targeted amp generation through controlled diffusion with efficient embeddings. *arXiv preprint arXiv:2504.17247*, 2025.

[11] Jiaming Liu, Tao Cui, Tao Wang, Xi Zeng, Yinbo Niu, Shaoqing Jiao, Dazhi Lu, Jun Wang, Shuyuan Xiao, Dongna Xie, et al. A multi-property optimizing generative adversarial network for de novo antimicrobial peptide design. *bioRxiv*, pages 2024–11, 2024.

[12] Li Wang, Yiping Li, Xiangzheng Fu, Xiucai Ye, Junfeng Shi, Gary G Yen, and Xiangxiang Zeng. Hmamp: Hypervolume-driven multi-objective antimicrobial peptides design. *arXiv preprint arXiv:2405.00753*, 2024.

[13] Massimo Caccia, Lucas Caccia, William Fedus, Hugo Larochelle, Joelle Pineau, and Laurent Charlin. Language gans falling short. *arXiv preprint arXiv:1811.02549*, 2018.

[14] Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep reinforcement learning that matters. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[15] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

[16] Guangshun Wang, Xia Li, and Zhe Wang. Apd3: the antimicrobial peptide database as a tool for research and education. *Nucleic acids research*, 44(D1):D1087–D1093, 2016.

[17] Guobang Shi, Xinyue Kang, Fanyi Dong, Yanchao Liu, Ning Zhu, Yuxuan Hu, Hanmei Xu, Xingzhen Lao, and Heng Zheng. Dramp 3.0: an enhanced comprehensive data repository of antimicrobial peptides. *Nucleic acids research*, 50(D1):D488–D496, 2022.

[18] Ulka Gawde, Shuvechha Chakraborty, Faiza Hanif Waghu, Ram Shankar Barai, Ashlesha Khanderkar, Rishikesh Indraguru, Tanmay Shirsat, and Susan Idicula-Thomas. Campr4: a database of natural and synthetic antimicrobial peptides. *Nucleic Acids Research*, 51(D1):D377–D383, 2023.

[19] Guizi Ye, Hongyu Wu, Jinjiang Huang, Wei Wang, Kuikui Ge, Guodong Li, Jiang Zhong, and Qingshan Huang. Lamp2: a major update of the database linking antimicrobial peptides. *Database*, 2020:baaa061, 2020.

[20] Lantian Yao, Jiahui Guan, Peilin Xie, Chia-Ru Chung, Zhihao Zhao, Danhong Dong, Yilin Guo, Wenyang Zhang, Junyang Deng, Yuxuan Pang, et al. dbamp 3.0: updated resource of antimicrobial activity and structural annotation of peptides in the post-pandemic era. *Nucleic acids research*, 53(D1):D364–D376, 2025.

[21] Stefano P Piotto, Lucia Sessa, Simona Concilio, and Pio Iannelli. Yadamp: yet another database of antimicrobial peptides. *International journal of antimicrobial agents*, 39(4):346–351, 2012.

[22] Jing Xu, Fuyi Li, André Leier, Dongxu Xiang, Hsin-Hui Shen, Tatiana T Marquez Lago, Jian Li, Dong-Jun Yu, and Jiangning Song. Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides. *Briefings in bioinformatics*, 22(5):bbab083, 2021.

[23] Anand Singh Rathore, Shubham Choudhury, Akanksha Arora, Purva Tijare, and Gajendra PS Raghava. Toxinpred 3.0: An improved method for predicting the toxicity of peptides. *Computers in biology and medicine*, 179:108926, 2024.

[24] Pietro Sormanni, Francesco A Aprile, and Michele Vendruscolo. The camsol method of rational design of protein mutants with enhanced solubility. *Journal of molecular biology*, 427(2):478–490, 2015.

[25] Elisabeth Gasteiger, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R Wilkins, Ron D Appel, and Amos Bairoch. Protein identification and analysis tools on the expasy server. In *The proteomics protocols handbook*, pages 571–607. Springer, 2005.

[26] Kunchur Guruprasad, BV Bhasker Reddy, and Madhusudan W Pandit. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*, 4(2):155–161, 1990.

[27] Martin Steinegger and Johannes Söding. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35(11):1026–1028, 2017.

[28] Jiyun Han, Tongxin Kong, and Juntao Liu. Pepnet: an interpretable neural network for anti-inflammatory and antimicrobial peptides prediction using a pre-trained protein language model. *Communications Biology*, 7(1):1198, 2024.

[29] Jiahui Guan, Peilin Xie, Dian Meng, Lantian Yao, Dan Yu, Ying-Chih Chiang, Tzong-Yi Lee, and Junwen Wang. Toxipep: Peptide toxicity prediction via fusion of context-aware representation and atomic-level graph. *Computational and Structural Biotechnology Journal*, 2025.

[30] Mohammad Madani, Kaixiang Lin, and Anna Tarakanova. Dsressol: A sequence-based solubility predictor created with dilated squeeze excitation residual networks. *International Journal of Molecular Sciences*, 22(24):13555, 2021.

[31] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

A Appendix

A.1 Performance of Classifiers

Table 2: Reported performance of classifiers used for dataset construction and validation.

Model	ACC	MCC	Recall	Precision	Specificity	AUC
PepNet	0.819	0.666	0.940	0.705	–	–
ToxiPep (CD-HIT 0.9)	0.850	0.701	0.827	–	0.873	–
DSResSol	0.796	0.589	0.769	0.817	0.782	–
ToxinPred3.0	0.930	0.860	0.920	0.930	0.930	0.980

Table 2 summarizes the reported performance of the predictive models used to construct and validate our training data. While all methods achieve reasonably high accuracies (0.80–0.93), they differ in sensitivity to false positives and false negatives. For instance, PepNet attains high recall (0.94) but relatively modest precision (0.705), suggesting that some sequences labeled as antimicrobial may be false positives. ToxiPep achieves balanced accuracy and specificity, but precision values were not reported, limiting interpretability of its error profile. DSResSol exhibits solid sensitivity and selectivity for solubility classification (ACC = 0.796), but its lower MCC (0.589) indicates that misclassifications remain substantial. ToxinPred3.0 provides the strongest overall metrics (ACC = 0.93, MCC = 0.86, AUC = 0.98), yet, like all predictors, it remains an imperfect proxy for experimental validation.

These limitations highlight an inherent source of noise in our training data. Consequently, some preference pairs may be mislabeled, potentially biasing the optimization process. Nevertheless, the high aggregate performance of these tools, especially when combined, provides sufficient signal to guide preference learning.

A.2 Training Details

We trained two variants of our preference-optimized models: an unconditioned DPO model and a token-conditioned DPO model. Both were initialized from the same pretrained ProtGPT2 backbone and trained under comparable conditions to ensure a fair comparison. Training was conducted for three epochs with a learning rate of 5×10^{-5} using AdamW optimization, gradient clipping at 1.0, and early stopping with a patience of 10 validation intervals. For each objective (AMP-likeness, non-toxicity, solubility, and stability), we constructed balanced preference pairs and sampled fixed-size batches to prevent bias toward any single property.

In the unconditioned setting, preference signals from all objectives were aggregated into a single model distribution. In the conditioned setting, we extended the tokenizer with property-specific control tokens, which were prepended during training to isolate supervision for each property while retaining the same optimization strategy. Both models were validated at regular intervals and evaluated on the same held-out data splits to enable direct and unbiased comparison.

Choice of Beta Values: An important design factor in preference optimization is the per-property scaling coefficient β, which balances the relative contribution of each objective to the DPO loss. We explored multiple β configurations to mitigate the dominance of any single property. Without scaling, the model showed a marked tendency to favor non-toxic and soluble peptides, often suppressing antimicrobial activity and stability. This likely reflects that solubility and non-toxicity are comparatively easier constraints for the model to satisfy.

The most effective trade-off was achieved with $\beta_{\text{AMP}} = 1.5$, $\beta_{\text{Stable}} = 1.0$, and $\beta_{\text{Non-Toxin}} = \beta_{\text{Soluble}} = 0.8$. This configuration increased the emphasis on antimicrobial activity, maintained sufficient pressure on stability, and reduced the over-representation of non-toxic and soluble peptides. It consistently produced the highest joint satisfaction rates across objectives and was therefore adopted for all reported experiments.

A.3 Structural Analysis of Generated Peptides

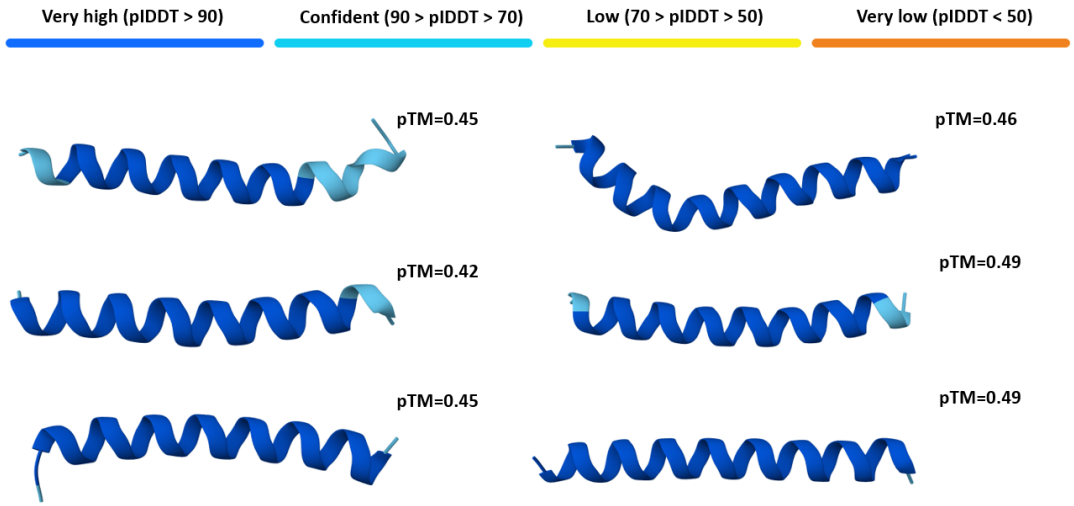

Figure 4: Representative AlphaFold3 structural predictions of generated peptides of tDPO-ProtGPT2. Structures are colored by per-residue confidence (pLDDT), with blue indicating very high confidence (>90) and orange indicating very low confidence (<50). The predicted TM-score (pTM) for each peptide is shown below its structure, reflecting overall global fold reliability.

To complement the property-based evaluations, we assessed the structural plausibility of representative peptides generated by our tDPO-ProtGPT2 framework using AlphaFold3 [31]. As shown in Figure 4, the predicted structures consistently form α-helical motifs, which are a common feature of many naturally occurring antimicrobial peptides. The high confidence scores (pLDDT > 90 across most residues) indicate reliable local structural assignments, suggesting that the designed sequences are not only biochemically optimized but also conformationally stable.

The predicted pTM-scores indicate that the generated peptides are not strongly aligned to existing structural templates (pTM<0.5). This outcome is expected in the context of de novo design, where the objective is to generate novel sequences and conformations rather than reproduce known folds. However, we treat these results as complementary evidence rather than definitive proof. Experimental validation will ultimately be required to confirm both structural and functional properties.