GOA-PGN: GLOBAL ORIENTATION-ALIGNED PHYSICS GRAPH NETWORK FOR QUANTITATIVE ANTIMICROBIAL PEPTIDE DISCOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

The rise of antibiotic-resistant bacteria drives urgent interest in antimicrobial peptides (AMPs). However, current computational discovery methods are hindered by single-modality representations that fail to capture the complex spatial mechanisms of membrane disruption. In this work, we propose a global orientation-aligned physics graph network (GOA-PGN) for predicting the minimum inhibitory concentration (MIC) of AMPs. GOA-PGN operates through two parallel paths to integrate localized sequence motifs with structural biophysical properties. One path leverages a peptide-tuned language model and partial convolutions to extract fine-grained sequential patterns. Simultaneously, the other path is used to construct coordinate-free topological graphs on the basis of the cosine similarity of side chain orientations to bypass static conformational noise. We integrate these dual representations via an attention mechanism to comprehensively capture global spatial amphipathicity, which is crucial for membrane disruption. Furthermore, the network incorporates a deep imbalanced regression (DIR) paradigm to mitigate the extreme label skewness prevalent in real-world bioactivity data. This paradigm uses dynamically weighted multitask loss to jointly predict the MIC and auxiliary physicochemical properties and employs a two-stage optimization strategy of adapter warm-up and subsequent joint fine-tuning to explicitly prioritize highly potent candidates. When evaluated on the DBAASP and CAMP datasets under the sequence homology constraints established by the quantitative mapping of antibacterial peptide (QMAP) benchmark, GOA-PGN demonstrates robust predictive accuracy across multiple bacterial strains. The code is publicly available at https://anonymous.4open.science/r/AMP-F316.

1 INTRODUCTION

The discovery of antibiotics was a monumental medical milestone, but the overuse of antimicrobial drugs accelerated the global spread of multidrug-resistant (MDR) pathogens, a crisis projected to cause ten million annual deaths by 2050 (Yuan et al., 2019; Murray et al., 2022; Ibisanmi et al., 2026; Centers for Disease Control and Prevention (CDC), 2024; Naghavi et al., 2024). Consequently, antimicrobial peptides (AMPs) have emerged as promising therapeutic alternatives because of their broad-spectrum activity and membrane-disrupting mechanisms, which substantially reduce the evolutionary pressure for resistance (Yasir et al., 2018; Oliveira Júnior et al., 2025). However, conventional wet-lab AMP discovery is resource intensive and low throughput, underscoring the urgent need for computational approaches.

Artificial intelligence (AI) has revolutionized the landscape of peptide discovery (Bae et al., 2025). Early AMP prediction models primarily relied on physicochemical features and classical deep learning architectures, such as deep neural networks (DNNs), convolutional neural networks (CNNs), and long short-term memory (LSTM) networks. However, these approaches were largely restricted to binary classification tasks (Krizhevsky et al., 2017; Hochreiter & Schmidhuber, 1997; Graves & Schmidhuber, 2005). While these models can autonomously extract sequence features to facilitate the discovery of novel peptides (Zhang et al., 2021; Dee, 2022; Lee et al., 2023), they fall short for tasks requiring granular quantitative profiling, such as predicting the minimum inhibitory concentration (MIC) against specific bacterial strains (Chiu et al., 2026). To overcome this limitation, pretrained

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

protein language models (PLMs)—most notably the ESM family—have been widely adopted (Hayes et al., 2024). By leveraging pretraining across billions of sequences, PLMs can capture deep evolutionary semantics without requiring explicit structural supervision. Recent studies have further employed peptide-optimized ESM-2 models to efficiently explore the peptide space, demonstrating the substantial promise of PLMs for the quantitative modelling of antimicrobial activity (Lin et al., 2023; Bai et al., 2026). Despite these transformative advancements, precise MIC prediction remains hindered by three severe challenges. First, the global self-attention mechanisms inherent to PLMs are computationally intensive and struggle to explicitly capture the localized physicochemical motifs— such as cationic charges and spatial amphipathicity— that are essential for membrane disruption (Li et al., 2026). Second, real-world biochemical datasets exhibit extreme label skewness, where highly active peptides are extraordinarily rare. This imbalance causes standard regression models to overfit the dense majority of weak AMPs and fail to navigate the biochemical "activity cliff" phenomenon (Shu et al., 2025). Consequently, as evidenced by the quantitative mapping of antibacterial peptide (QMAP) benchmark, existing models exhibit near-random predictive performance for highly potent targets (Pham et al., 2026). Third, traditional random data splitting frequently introduces homologous sequence leakage, thereby artificially inflating generalization metrics and obscuring true model performance (Lavertu et al., 2026).

To address these limitations, we propose GOA-PGN, a novel, homology-aware deep learning framework for accurate AMP MIC prediction. We utilize a peptide-pretrained ESM-2 backbone alongside a lightweight partial convolution (PConv) (Chen et al., 2023) module to extract localized sequence motifs. Because 1D features alone are insufficient to represent the complex spatial amphipathicity crucial for membrane disruption, we integrate a coordinate-free, physics-informed graph encoder. Instead of relying on explicit 3D distance matrices—, which often introduce structural noise—, our graph topology is constructed exclusively from the cosine similarity of side-chain orientations to establish robust geometric inductive biases. This combined representation is then adaptively recalibrated via an attention mechanism (Hu et al., 2018). Furthermore, we employ a deep imbalanced regression (DIR) (Yang et al., 2021) architecture with a two-stage training paradigm (adapter warm-up and joint fine-tuning) and a dynamically weighted multitask loss. In terms of the strictly homology-reduced QMAP benchmark, GOA-PGN provides a highly efficient computational workflow for screening next-generation AMPs.

Our contributions are summarized as follows:

- **Identification of Core Bottlenecks.** This work delineates three core bottlenecks in quantitative AMP MIC prediction, namely localized motif extraction inefficiency, extreme label skewness, and homologous sequence leakage.
- **A Novel Regression Pipeline.** We propose GOA-PGN, a coordinate-free regression pipeline integrating a peptide-pretrained ESM-2, a PConv module, and an orientation-based physics graph. This architecture circumvents 3D conformational noise, enabling the model to capture critical spatial amphipathicity. Furthermore, we develop a two-stage DIR paradigm with a dynamically weighted multi-task loss to explicitly model high-potency boundaries.
- **Superior Generalization.** Extensive evaluations on the QMAP benchmark, alongside the DBAASP and CAMP datasets, demonstrate that GOA-PGN achieves superior generalization and predictive accuracy, particularly for highly potent AMPs.

2 Related Work

Antimicrobial Peptide Discovery. The escalating crisis of antimicrobial resistance has led to the computational discovery of AMPs as viable therapeutic alternatives. Early computational efforts predominantly framed AMP discovery as a binary classification problem—distinguishing AMPs from non-AMPs using hand-crafted physicochemical descriptors and classical machine learning models. With the proliferation of deep learning, models such as CNNs and RNNs (Veltri et al., 2018; Sharma et al., 2021) have drastically improved sequence-to-function mapping. Recently, large-scale PLMs such as the ESM family (Lin et al., 2023) have been extensively utilized to extract evolutionary semantics, achieving remarkable accuracy in binary or multilabel functional classification tasks (Lee et al., 2023). Building upon these predictive foundations, recent computational efforts have rapidly expanded toward automated *de novo* design. This paradigm shift clearly progresses from foundational human-level reward design by coding LLMs (Ma et al., 2024) to the deployment of autonomous

LLM agents for general organic molecule generation (Hu et al., 2025) and, most recently, to closed-loop multiagent collaboration systems specifically tailored for multiobjective AMP discovery (Zhou et al., 2026). While these advanced generative frameworks often incorporate MIC estimators to guide sequence generation, quantitative prediction is frequently treated as an auxiliary scoring function. Dedicated, high-precision MIC prediction models remain remarkably scarce. For true clinical translation, highly specialized and robust quantitative efficacy profiling is urgently needed to complement existing generative pipelines.

Bottlenecks in Quantitative MIC Modeling. Predicting the exact MIC of an AMP represents a critical but highly complex challenge, primarily hindered by two major biochemical and computational bottlenecks. First, real-world biochemical screening datasets exhibit extreme label skewness, conforming to long-tailed distributions where highly potent candidates are extraordinarily rare (Macarron et al., 2011). Standard regression models overfit the dense majority of weak AMPs and suffer from catastrophic failure on minority samples—, a problem exacerbated by biochemical "activity cliffs" where minor amino acid substitutions cause drastic potency shifts (Cruz-Monteagudo et al., 2014). Second, precise MIC prediction fundamentally relies on capturing the spatial mechanisms of membrane disruption, which 1D sequence models often miss. While GNNs (Gilmer et al., 2017) and 3D modelling (Ruff & Pappu, 2021) excel at molecular representation (Yang et al., 2019; Satorras et al., 2021; Schütt et al., 2017; Jing et al., 2021), explicitly relying on predicted 3D distance matrices is fundamentally problematic for AMPs. Current structural predictors typically output flexible, steady-state conformations in aqueous solutions, which drastically differ from the active, membrane-induced structures required for disruption. Consequently, directly incorporating these explicit 3D distances introduces severe conformational noise, causing models to overfit to biologically irrelevant spatial states.

3 METHODOLOGY

We created GOA-PGN, a novel deep learning framework, that is designed to achieve accurate MIC estimation for AMPs. GOA-PGN executes the workflow through four interconnected stages, beginning with an input encoding stage and concluding with a multitask prediction stage. The only input required from the user is a raw peptide sequence, from which the framework explicitly captures spatial amphipathicity to evaluate its antimicrobial potency. Figure 1 outlines the entire workflow, and it is explained in detail below.

3.1 SEMANTIC AND STRUCTURAL ENCODING

The Input & Backbones module extracts foundational representations from both semantic and biophysical modalities.

Semantic encoding. We leverage a peptide-tuned ESM-2 model, originally developed and pretrained by Bae et al. (2025), as our foundational sequence encoder. Through targeted masked language modelling of peptide sequences, this optimized model can capture peptide-specific patterns more effectively, thereby enhancing the overall model's representational capacity. Given a tokenized peptide sequence X of length L, this model computes dense contextual embeddings through multilayer bidirectional transformer blocks. We extract the representations from the final hidden layer, denoted as $E \in \mathbb{R}^{L \times d_{esm}}$, where d_{esm} represents the hidden dimension of the model. To compact this high-dimensional space for physicochemical property extraction, we introduce a projection block consisting of a linear transformation followed by layer normalization, GELU activation, and dropout. The projected representation $H_{proj} \in \mathbb{R}^{L \times d_c}$ is formulated as follows:

$$H_{proj} = \text{Dropout}(\text{GELU}(\text{LayerNorm}(EW_p + b_p))), \tag{1}$$

where $W_p \in \mathbb{R}^{d_{esm} \times d_c}$ and $b_p \in \mathbb{R}^{d_c}$ are the learnable weight matrix and bias vector of the linear transformation, respectively, with the projected dimension d_c.

Structural encoding. Simultaneously, to establish a coordinate-free structural foundation, an ESMFold backbone processes the same sequence X to predict its 3D conformation. Rather than utilizing the entire coordinate set, we strictly extract the side-chain direction vectors $V \in \mathbb{R}^{L \times 3}$. Each vector denotes the directional displacement from the C_α atom to the estimated C_β atom, laying the biophysical groundwork for subsequent topological construction.

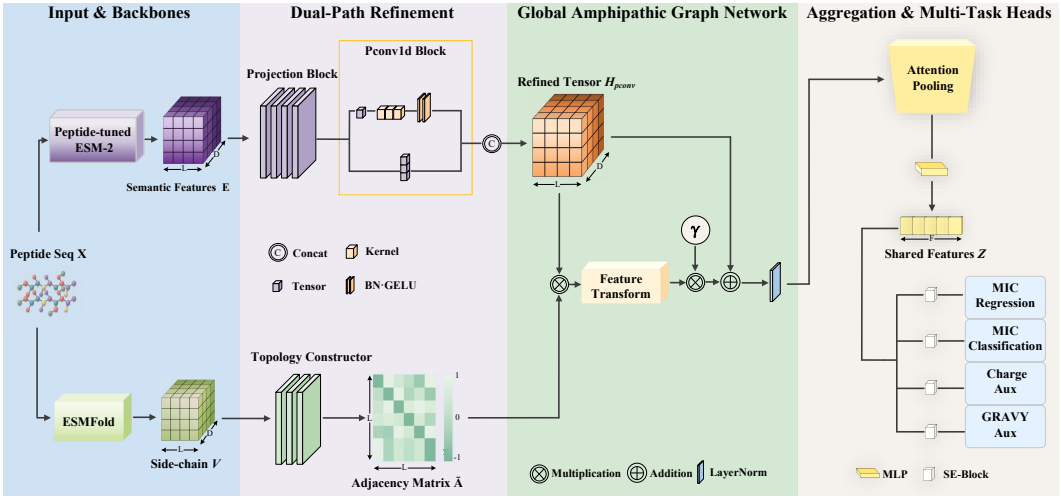


Figure 1: Overall architecture of the GOA-PGN. The model integrates semantic and structural features through a dual-path orientation graph to capture spatial amphipathicity, ultimately performing joint MIC and auxiliary biophysical predictions via multitask heads.

3.2 Dual-Path Refinement and Global Amphipathic Graph Network

To capture both localized physicochemical motifs and global spatial amphipathicity, the framework employs a dual-path refinement strategy leading to the global amphipathic graph network.

1D convolutional refinement. In the first path, the projected semantic feature H_{proj} is channeled into a 1D PConv block. We partition its channels into a representative subset $H^{(1)} \in \mathbb{R}^{L\times d_{cp}}$ and an untouched subset $H^{(2)}$. A 1D convolution with a kernel size of K, followed by batch normalization and the GELU, is applied exclusively to $H^{(1)}$. The output is concatenated with $H^{(2)}$ to yield a refined 1D feature map $H_{pconv} \in \mathbb{R}^{L\times d_c}$:

$$\tilde{H}^{(1)} = \text{GELU}(\text{BatchNorm1D}(\text{Conv1D}(H^{(1)}))), \tag{2}$$

$$H_{pconv} = [\tilde{H}^{(1)}; H^{(2)}] \in \mathbb{R}^{L\times d_c}. \tag{3}$$

This design enriches local physicochemical motifs while preserving global semantic integrity.

Topology construction. In the parallel path, the topology constructor transforms the side-chain vectors V into a physically informed adjacency matrix $\tilde{A}$. By computing the pairwise cosine similarity of side-chain orientations, we intentionally prioritize directionality over spatial proximity:

$$A_{i,j} = \text{ReLU}\left(\frac{v_i \cdot v_j}{\|v_i\|\|v_j\|}\right), \tag{4}$$

where v_i and v_j denote the side chain vectors of the i-th and j-th residues, respectively. To construct a valid message-passing matrix, we define a binary spatial mask $M \in \{0,1\}^{L\times L}$. Specifically, $M_{i,j} = 1$ if both residues i and j are valid (nonpadding) tokens and $i \neq j$, effectively eliminating the impact of padded regions and self-loops. The masked adjacency matrix is then subjected to rowwise degree normalization to yield the final physical transition matrix $\tilde{A}$:

$$\tilde{A}_{i,j} = \frac{A_{i,j}\cdot M_{i,j}}{\sum_{k=1}^{L} A_{i,k}\cdot M_{i,k} + \epsilon}, \tag{5}$$

Global amphipathic graph network. Within the global amphipathic graph network, H_{pconv} is aggregated over this topological structure and projected via a feature transform weight $W_g \in \mathbb{R}^{d_c\times d_c}$:

$$Z_{phys} = \text{GELU}(\tilde{A}H_{pconv}W_g). \tag{6}$$

To ensure training stability, a zero-initialized learnable scalar γ gates the residual connection. The final physics-aware representation is computed as follows:

$$H_{phys} = \text{LayerNorm}(H_{pconv} + \gamma \cdot Z_{phys}). \tag{7}$$

This coordinate-free gating mechanism organically intertwines sequential motifs with 3D-aware global topologies. By circumventing explicit distance predictions in our definitive architecture, the module immunizes the learned representations against conformational noise. Note that while this distance-agnostic formulation serves as our primary model for highly diverse, large-scale datasets, we also formulate a distance-relaxed variant—incorporating an explicit distance scaling factor ($\exp(-dist/\tau)$)—to evaluate its adaptability to smaller, structurally rigid peptide distributions (discussed in Section 4.2.2). Ultimately, this module delivers a highly robust and structurally enriched feature space for the downstream attention pooling mechanism.

3.3 Aggregation and multitask prediction heads

Attention pooling. To distill the residue-level graph features H_{phys} into a fixed-size representation, we deploy an attention pooling mechanism. By applying a two-layer projection with an additive padding mask $m \in \{0, -\infty\}^{L\times 1}$, we generate valid attention probabilities $\alpha \in \mathbb{R}^{L\times 1}$:

$$\alpha = \text{Softmax}(\text{Tanh}(H_{phys}W_{a1})W_{a2} + m). \tag{8}$$

The sequence-level representation is then computed as the weighted sum of the valid residual features. To stabilize the multitask learning process and enhance feature expressiveness, this pooled vector is passed through a shared feature extractor (denoted as $\text{Ext}(\cdot)$), which sequentially applies a linear transformation, layer normalization, GELU activation, and dropout. The final shared representation $z \in \mathbb{R}^{d_{hidden}}$ is computed as follows:

$$z = \text{Ext}\left(\alpha^\top H_{phys}\right). \tag{9}$$

Multitask heads. This shared representation z is subsequently broadcast to four distinct multitask heads: MIC Regression, MIC Classification, Charge Aux, and grand average of hydropathy (GRAVY) Aux. Within each head, an SE block acts as a channelwise attention mechanism to adaptively recalibrate feature responses. The calibrated vector z' is computed as follows:

$$z' = z \odot \text{Sigmoid}(W_2\text{ReLU}(W_1 z)), \tag{10}$$

where $W_1 \in \mathbb{R}^{\frac{d_{hidden}}{r}\times d_{hidden}}$ and $W_2 \in \mathbb{R}^{d_{hidden}\times \frac{d_{hidden}}{r}}$ are learnable weight matrices without bias terms and r is the reduction ratio. This calibrated vector is ultimately processed by an MLP to yield task-specific predictions.

3.4 Deep imbalanced regression and Two-stage Optimization

A persistent challenge in bioactivity modelling is the extreme scarcity of highly potent peptides. The standard MSE uniformly penalizes errors, leading to the underoptimization of the critical minority class. We address this via a DIR objective employing a dynamically weighted MSE loss.

Deep imbalanced regression. We define a high-potency threshold τ in the $\log_{10}$ space. For the i-th sample with the ground-truth log-transformed MIC value y_i, we assign a regression weight ω_i using an indicator function $\mathbb{I}(\cdot)$:

$$\omega_i = 1 + (\lambda_{hp} - 1)\mathbb{I}(y_i < \tau), \tag{11}$$

where $\lambda_{hp} > 1$ is a predefined hyperparameter that amplifies the gradient contribution of highly active samples. The task-specific regression loss is thus defined as follows:

$$\mathcal{L}_{reg} = \frac{1}{N}\sum_{i=1}^{N} \omega_i (y_i - \hat{y}_{reg,i})^2. \tag{12}$$

The total objective integrates this DIR loss with a binary cross-entropy (BCE) classification loss and MSE for the auxiliary physicochemical tasks:

$$\mathcal{L}_{total} = \mathcal{L}_{reg} + \alpha\mathcal{L}_{cls} + \beta(\mathcal{L}_{\text{charge}} + \mathcal{L}_{\text{GRAVY}}). \tag{13}$$

where $\mathcal{L}_{cls}$ denotes the standard BCE loss for the binary MIC classification task, and both $\mathcal{L}_{\text{charge}}$ and $\mathcal{L}_{\text{GRAVY}}$ are standard MSE losses for predicting their respective auxiliary biophysical properties. The hyperparameters α and β control the relative contributions of these auxiliary objectives.

Two-stage optimization. To stably infuse these complex structural biases into the parameter space without triggering catastrophic forgetting, the network is trained via a two-stage optimization paradigm:

- **Adapter warm-up.** The ESM-2 backbone is strictly frozen. The optimizer exclusively updates the PConv module, graph network components, and task-specific heads. This allows the newly initialized adapter modules to learn a stable mapping from the PLM latent space to the physical domain.
- **Joint fine-tuning.** The entire architecture is unfrozen and optimized end-to-end. We apply a differential learning rate strategy, assigning a significantly smaller learning rate to the backbone compared to the adapter modules, thereby preserving generalized sequence knowledge.

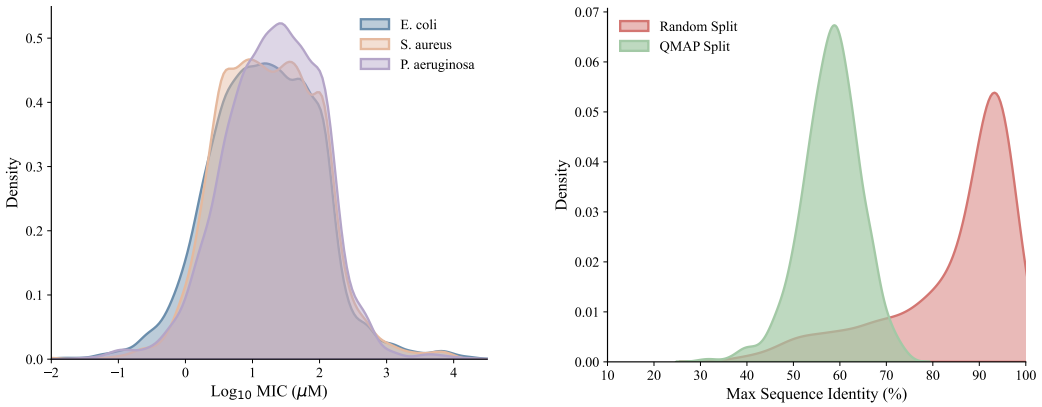
Both stages are governed by a cosine annealing learning rate scheduler to ensure stable convergence.

4 Experiments

To demonstrate the effectiveness of our proposed GOA-PGN, we conduct comprehensive experiments on the task of AMP MIC prediction. We evaluate our method across multiple bacterial strains and compare it against state-of-the-art baselines.

4.1 Experimental Setup

4.1.1 Datasets



(a) Target-Specific MIC Distribution

(b) Sequence Identity Distribution

Figure 2: Dataset characteristics. (a) $\log_{10}$(MIC) distributions for three pathogens. (b) Maximum train–test sequence identity distributions.

We evaluate our method on the DBAASP and CAMP databases. The preprocessing pipeline involved filtering for valid natural monomeric peptides, averaging the MIC values of duplicate sequences, and standardizing activity labels via a base-10 logarithmic transformation of the micromolar (μM) MIC values. For DBAASP, this yielded refined pools of unique sequences targeting three major pathogens: *E. coli* (8,153), *S. aureus* (7,432), and *P. aeruginosa* (5,420). While this transformation normalizes the bulk of the MIC distribution (Figure 2a), extremely potent elite candidates remain sparsely distributed across the extreme left tail. CAMP serves as a supplementary generalization benchmark, where the identical preprocessing pipeline yielded 2,611 unique *E. coli* sequences from an initial pool of 16,886.

Traditional random data splitting suffers from severe leakage, resulting in a substantial density peak at sequence identities between 90% and 100%. To prevent this, we utilized the QMAP graph-based splitting engine (Lavertu et al., 2026). As density estimation confirms (Figure 2b), QMAP effectively mitigates this leakage by shifting and concentrating the primary sequence identity distribution predominantly below the 60% threshold, ensuring an unbiased evaluation environment. Finally, we employed fixed random seeds (1, 3, 7, 12, and 404) to generate five independent train-test splits at an approximate 80:20 ratio. Comprehensive data cleaning and partitioning protocols are detailed in Appendix A.

Table 1: Performance comparison on the DBAASP *E. coli* dataset. The upper panel presents the results evaluated on the full dataset, while the lower panel details the performance on the high-potency subset (MIC < 10 μM). The best results in each category are highlighted in **bold**.

Architecture	Spearman ↑	Pearson ↑	Kendall's τ ↑	RMSE ↓	MAE ↓	MSE ↓
Panel A: Full Dataset						
Witten Witten & Witten (2019)	0.452 ± 0.037	0.478 ± 0.041	0.315 ± 0.026	0.695 ± 0.032	0.544 ± 0.019	0.484 ± 0.045
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.292 ± 0.061	0.303 ± 0.041	0.199 ± 0.043	0.754 ± 0.036	0.597 ± 0.023	0.569 ± 0.056
ESM2-Linear Lin et al. (2023)	0.428 ± 0.049	0.432 ± 0.051	0.298 ± 0.036	0.789 ± 0.024	0.602 ± 0.009	0.623 ± 0.039
MBC-Attention Yan et al. (2023)	0.471 ± 0.028	0.506 ± 0.039	0.358 ± 0.043	0.676 ± 0.030	0.531 ± 0.019	0.458 ± 0.041
BERT-AMPEP60-ES Cai et al. (2025)	0.507 ± 0.056	0.519 ± 0.062	0.358 ± 0.043	0.688 ± 0.028	0.528 ± 0.020	0.474 ± 0.040
LLAMP Bae et al. (2025)	0.514 ± 0.051	0.538 ± 0.040	0.364 ± 0.040	0.679 ± 0.027	0.524 ± 0.012	0.462 ± 0.037
ANIA Chiu et al. (2026)	0.395 ± 0.071	0.429 ± 0.069	0.274 ± 0.051	0.743 ± 0.038	0.586 ± 0.022	0.554 ± 0.056
GOA-PGN (Ours)	$\mathbf{0.549 \pm 0.042}$	$\mathbf{0.567 \pm 0.032}$	$\mathbf{0.391 \pm 0.033}$	$\mathbf{0.646 \pm 0.013}$	$\mathbf{0.501 \pm 0.005}$	$\mathbf{0.417 \pm 0.017}$
Panel B: High-Potency Subset (MIC < 10 μM)						
Witten Witten & Witten (2019)	0.207 ± 0.034	0.165 ± 0.069	0.140 ± 0.023	0.779 ± 0.041	0.607 ± 0.032	0.608 ± 0.067
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.186 ± 0.082	0.237 ± 0.062	0.127 ± 0.057	0.853 ± 0.028	0.715 ± 0.024	0.728 ± 0.049
ESM2-Linear Lin et al. (2023)	0.189 ± 0.058	0.192 ± 0.108	0.129 ± 0.040	0.843 ± 0.056	0.644 ± 0.046	0.713 ± 0.094
MBC-Attention Yan et al. (2023)	0.277 ± 0.060	0.276 ± 0.100	0.190 ± 0.041	0.727 ± 0.032	0.563 ± 0.029	0.530 ± 0.046
BERT-AMPEP60-ES Cai et al. (2025)	0.239 ± 0.091	0.213 ± 0.061	0.163 ± 0.064	0.814 ± 0.055	0.641 ± 0.052	0.665 ± 0.090
LLAMP Bae et al. (2025)	0.240 ± 0.063	0.224 ± 0.086	0.164 ± 0.043	0.757 ± 0.083	0.577 ± 0.056	0.578 ± 0.127
ANIA Chiu et al. (2026)	0.203 ± 0.103	0.157 ± 0.106	0.138 ± 0.070	0.861 ± 0.124	0.695 ± 0.117	0.753 ± 0.220
GOA-PGN (Ours)	$\mathbf{0.313 \pm 0.056}$	$\mathbf{0.315 \pm 0.072}$	$\mathbf{0.216 \pm 0.037}$	$\mathbf{0.697 \pm 0.034}$	$\mathbf{0.524 \pm 0.020}$	$\mathbf{0.486 \pm 0.047}$

4.1.2 Implementation and Evaluation Metrics

Our framework is implemented in PyTorch and trained on a single NVIDIA RTX 4090 GPU. We utilize a two-stage training curriculum: an initial 10-epoch warm-up optimizing only the GOA-PGN adapter (learning rate 5×10^{-4}), followed by 30 epochs of joint fine-tuning with differential learning rates (2×10^{-5} for the backbone, 1×10^{-4} for the adapter) using a cosine annealing scheduler.

Following standard conventions for molecular property prediction, we evaluate the predictive performance using a comprehensive suite of metrics. The regression quality is primarily assessed via Spearman's rank correlation and the RMSE. We also report the Pearson correlation, Kendall's τ, MAE, and MSE for completeness. Note that the coefficient of determination (R^2) is explicitly omitted from our evaluation. As demonstrated in the QMAP study (Lavertu et al., 2026), calculating R^2 on highly skewed domains or subsets with truncated targets often yields strongly negative values, which distorts overall performance comparisons.

4.2 Experimental Performance Comparison

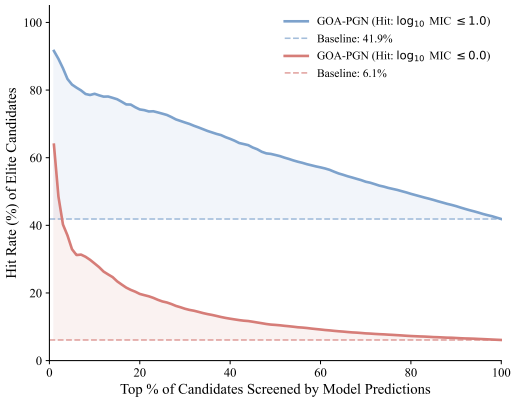


Figure 3: Dual-threshold enrichment curves for *E. coli* test set.

Table 2: Performance comparison on the CAMP *E. coli* dataset. The upper panel presents the results evaluated on the full dataset, while the lower panel details the performance on the high-potency subset (MIC < 10 μM). The best results in each category are highlighted in **bold**.

Architecture	Spearman ↑	Pearson ↑	Kendall's τ ↑	RMSE ↓	MAE ↓	MSE ↓
Panel A: Full Dataset						
Witten Witten & Witten (2019)	0.404 ± 0.041	0.408 ± 0.036	0.279 ± 0.027	0.717 ± 0.031	0.556 ± 0.028	0.515 ± 0.044
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.278 ± 0.068	0.274 ± 0.061	0.187 ± 0.046	0.741 ± 0.033	0.581 ± 0.031	0.550 ± 0.049
ESM2-Linear Lin et al. (2023)	0.217 ± 0.076	0.209 ± 0.049	0.148 ± 0.052	1.384 ± 0.152	1.017 ± 0.074	1.935 ± 0.428
MBC-Attention Yan et al. (2023)	0.451 ± 0.028	0.456 ± 0.034	0.314 ± 0.021	0.684 ± 0.054	0.540 ± 0.043	0.470 ± 0.072
BERT-AMPEP60-ES Cai et al. (2025)	0.470 ± 0.042	0.461 ± 0.031	0.327 ± 0.032	0.696 ± 0.043	0.544 ± 0.041	0.486 ± 0.059
LLAMP Bae et al. (2025)	0.493 ± 0.073	0.488 ± 0.074	0.346 ± 0.055	0.674 ± 0.054	0.524 ± 0.043	0.457 ± 0.075
ANIA Chiu et al. (2026)	0.379 ± 0.041	0.394 ± 0.057	0.259 ± 0.032	0.747 ± 0.057	0.586 ± 0.047	0.560 ± 0.084
GOA-PGN (Ours)	$\mathbf{0.514 \pm 0.053}$	$\mathbf{0.502 \pm 0.057}$	$\mathbf{0.361 \pm 0.042}$	$\mathbf{0.668 \pm 0.060}$	$\mathbf{0.516 \pm 0.039}$	$\mathbf{0.449 \pm 0.079}$
Panel B: High-Potency Subset (MIC < 10 μM)						
Witten Witten & Witten (2019)	0.263 ± 0.085	0.241 ± 0.093	0.180 ± 0.056	0.746 ± 0.018	0.569 ± 0.038	0.557 ± 0.027
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.122 ± 0.132	0.078 ± 0.119	0.083 ± 0.090	0.807 ± 0.045	0.639 ± 0.036	0.654 ± 0.073
ESM2-Linear Lin et al. (2023)	0.152 ± 0.080	0.152 ± 0.041	0.102 ± 0.052	1.474 ± 0.293	1.080 ± 0.139	2.240 ± 0.897
MBC-Attention Yan et al. (2023)	0.242 ± 0.101	0.229 ± 0.091	0.167 ± 0.071	0.704 ± 0.091	0.541 ± 0.076	0.502 ± 0.129
BERT-AMPEP60-ES Cai et al. (2025)	0.240 ± 0.097	0.211 ± 0.101	0.165 ± 0.063	0.752 ± 0.106	0.581 ± 0.092	0.575 ± 0.164
LLAMP Bae et al. (2025)	$\mathbf{0.303 \pm 0.115}$	$\mathbf{0.262 \pm 0.102}$	$\mathbf{0.208 \pm 0.076}$	0.715 ± 0.093	0.541 ± 0.075	0.518 ± 0.135
ANIA Chiu et al. (2026)	0.221 ± 0.068	0.254 ± 0.093	0.151 ± 0.047	0.766 ± 0.126	0.586 ± 0.103	0.600 ± 0.195
GOA-PGN (Ours)	0.288 ± 0.126	0.246 ± 0.148	0.200 ± 0.087	$\mathbf{0.691 \pm 0.132}$	$\mathbf{0.508 \pm 0.097}$	$\mathbf{0.491 \pm 0.188}$

4.2.1 Performance on the DBAASP Dataset

We first evaluate the models on the DBAASP *E. coli* subset, reporting performance on both the full dataset and a high-potency subset (MIC < 10 μM). Although this top echelon comprises approximately 40% of our candidate pool, its elite candidates remain extremely sparse and heavily right-skewed—a regime where standard models typically suffer catastrophic degradation. As Table 1 demonstrates, GOA-PGN establishes state-of-the-art results across the board. On the full dataset, the substantial reduction in RMSE and MAE highlights that explicitly encoding topological graphs via relative side-chain co-orientations outperforms purely 1D sequence models. Conversely, baselines like ANIA degrade severely under strict homology partitioning due to their reliance on specific preprocessing and lack of pre-trained embeddings. In the stringent high-potency regime, while sequence-based models (e.g., LLAMP) underperform previously dominant methods (e.g., MBC-Attention), GOA-PGN consistently retains the lead. This resilience confirms that capturing explicit biophysical amphipathicity is indispensable.

To demonstrate practical utility in drug discovery, we evaluated GOA-PGN's virtual screening enrichment (Figure 3). The model achieves an initial hit rate approaching 90% for standard high-potency candidates ($\log_{10}$ MIC ≤ 1.0). More remarkably, for the extremely rare ultrapotent subset ($\log_{10}$ MIC ≤ 0.0, baseline 6.1%), top-ranked predictions yield a hit rate exceeding 60%—a nearly 10-fold enrichment over random screening. This confirms that our method effectively mitigates regression shrinkage, successfully prioritizing rare elite candidates for wet-lab validation.

Finally, we evaluated GOA-PGN on the *P. aeruginosa* and *S. aureus* subsets (Appendices B and C), which present distinct biophysical challenges. For *P. aeruginosa*, its complex outer-membrane permeability barrier and specific lipopolysaccharide (LPS) interactions cause a noticeable degradation in ranking correlations across all architectures. Conversely, the gram-positive *S. aureus* lacks an outer lipid membrane but features a thick peptidoglycan layer, meaning membrane penetration is heavily dictated by local sequence motifs. Nevertheless, despite this shift in correlation dominance, GOA-PGN robustly maintains the lowest absolute prediction errors.

4.2.2 Performance on the CAMP Dataset

We evaluate GOA-PGN on the CAMP *E. coli* dataset to rigorously test generalization under distribution shifts (Table 2). In the critical high-potency regime, while purely sequence-based models like LLAMP exhibit competitive rank correlations, they suffer from substantial absolute prediction

deviations. In contrast, GOA-PGN robustly minimizes these deviations, maintaining superior error bounds and achieving a lower RMSE.

Notably, while our distance-agnostic graph optimally mitigates structural noise on the massive DBAASP dataset, we empirically found that incorporating explicit ESMFold distances ($\exp(-dist/\tau)$) actually aligns better with the CAMP distribution in certain aspects, as detailed in Appendix D. We attribute this divergence to intrinsic peptide flexibility: static 3D vacuum predictions introduce confounding noise for the diverse, intrinsically disordered AMPs prevalent in DBAASP, yet provide beneficial localized constraints for smaller, well-characterized rigid peptides typical of CAMP. Given DBAASP's comprehensive scope, we establish the distance-agnostic architecture as our definitive model. Its superior overall performance strongly corroborates the biological consensus that AMP membrane insertion is driven by global amphipathic topology rather than isolated 1D motifs.

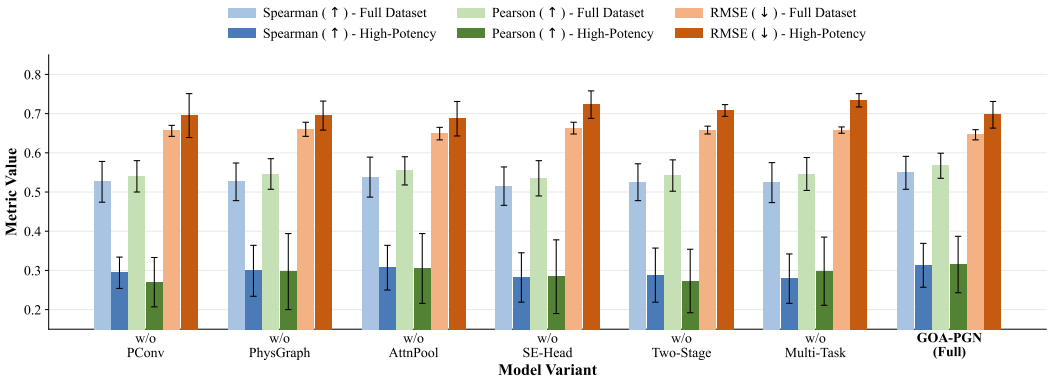


Figure 4: Ablation study of GOA-PGN on the DBAASP *E. coli* dataset.

4.2.3 Ablation Study

To systematically validate GOA-PGN's architectural and training components, we conducted ablation studies on the DBAASP *E. coli* dataset (full results in Appendix E). On the full dataset, omitting the SE-Head severely degrades ranking capacity, while discarding PConv or PhysGraph uniformly drops all metrics, confirming that localized motif filtering, explicit topological amphipathicity, and feature recalibration provide synergistic biophysical cues. Similarly, replacing adaptive attention with global average pooling degrades performance, validating the dynamic weighting of anchor residues.

The high-potency subset reveals a rank-error trade-off. Removing AttentionPool or PhysGraph yields nominal error reductions but compromises ranking accuracy—the most actionable metric for virtual screening—justifying the full model's prioritization of robust rank-ordering. In this constrained regime, bypassing PConv drastically collapses the Pearson correlation, underscoring its necessity for capturing linear trends among elite candidates. Finally, our training paradigms prove indispensable: skipping the two-stage warm-up disrupts pretrained representations, and removing the multitask objective triggers severe error spikes. These stark degradations confirm that explicit biophysical priors are vital to prevent catastrophic overfitting to sequence biases.

5 Conclusion

In conclusion, GOA-PGN is a novel coordinate-free graph neural network that redefines quantitative AMP MIC prediction by prioritizing topological amphipathicity. By constructing a physics-informed graph exclusively from side-chain orientation similarities and employing a two-stage DIR strategy, explicit 3D conformational noise and extreme label skewness can be effectively mitigated. GOA-PGN achieves state-of-the-art performance on the homology-reduced QMAP benchmark, consistently outperforming existing baselines in both rank-ordering and absolute error minimization, particularly within the clinically vital high-potency regime. More broadly, it provides a robust, highly generalizable paradigm for integrating biophysical priors into next-generation therapeutic peptide design.

AI USE STATEMENT

In this work, we used generative AI tools for literature search, brainstorming, interpreting complex analytical results, and refining the presentation of theoretical frameworks. Additionally, we utilized generative AI as a programming assistant to help reproduce outdated baseline experiments, strictly guiding the AI to adapt previous open-source logic into a unified PyTorch framework for fair benchmarking.

We have not used generative AI tools for dataset generation, for fabricating references, for forging experimental data, or for directly generating research hypotheses, scientific conclusions, or core innovations. Furthermore, no AI tools were used to plagiarize other researchers' ideas or to provide direct, overarching conceptual guidance.

We have rigorously reviewed all AI-assisted work. For instance, all AI-assisted codebase migrations for baseline reproductions were manually verified, debugged, and tested for correctness by the authors to ensure strict consistency with their original algorithmic intent. Any biological interpretations or analytical insights assisted by AI were cross-validated against biological ground truths. We take full responsibility for the final content of this work, including all text, claims, code, and computational artifacts produced with the aid of generative AI, ensuring that they meet the standards of scientific rigor and integrity.

REFERENCES

D. Bae, M. Kim, J. Seo, and H. Nam. Ai-guided discovery and optimization of antimicrobial peptides through species-aware language model. *Briefings in Bioinformatics*, 26(4):bbaf343, 2025. doi: 10.1093/bib/bbaf343.

W. Bai, W. Yang, Y.-Q. Chen, H. Ji, L. Brennan, and L. Wang. Ampban: A deep learning framework integrating protein sequence and structural features for antimicrobial peptide prediction. *bioRxiv*, 2026. doi: 10.64898/2026.01.20.700468.

Jianxiu Cai, Jielu Yan, Chonwai Un, Yapeng Wang, François-Xavier Campbell-Valois, and Shirley W. I. Siu. Bert-ampep60: A bert-based transfer learning approach to predict the minimum inhibitory concentrations of antimicrobial peptides for escherichia coli and staphylococcus aureus. *Journal of Chemical Information and Modeling*, 65(7):3186–3202, 2025. ISSN 1549-9596. doi: 10.1021/acs.jcim.4c01749. URL https://doi.org/10.1021/acs.jcim.4c01749.

Centers for Disease Control and Prevention (CDC). Antimicrobial resistance. Available online at: https://www.cdc.gov/antimicrobial-resistance/data-research/facts-stats/index.html, 2024.

J. Chen, X. Fan, K. Xu, et al. Run, don't walk: chasing higher flops for faster neural networks. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12021–12031, 2023.

Y.-P. Chiu, L. Yao, Y. Tang, C.-R. Chung, Y. Pang, Y.-C. Chiang, and T.-Y. Lee. Ania: an inception-attention network for predicting minimum inhibitory concentration of antimicrobial peptides. *Briefings in Bioinformatics*, 27(1):bbag023, 2026. doi: 10.1093/bib/bbag023.

Maykel Cruz-Monteagudo, José L. Medina-Franco, Yunierkis Pérez-Castillo, Orazio Nicolotti, M. Natália D.S. Cordeiro, and Fernanda Borges. Activity cliffs in drug discovery: Dr jekyll or mr hyde? *Drug Discovery Today*, 19(8):1069–1080, 2014.

W. Dee. Lmpred: Predicting antimicrobial peptides using pretrained language models and deep learning. *Bioinformatics Advances*, 2:vbac021, 2022. doi: 10.1093/bioadv/vbac021.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272, 2017.

A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18:602–610, 2005. doi: 10.1016/j.neunet.2005.06.042.

T. Hayes, R. Rao, H. Akin, et al. Simulating 500 million years of evolution with a language model. *bioRxiv*, 2024. doi: 10.1101/2024.07.01.600583.

S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7132–7141, 2018. doi: 10.1109/CVPR.2018.00745.

Zhaolin Hu, Yixiao Zhou, Zhongan Wang, Xin Li, Weimin Yang, Hehe Fan, and Yi Yang. OSDA agent: Leveraging large language models for de novo design of organic structure directing agents. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=9YNyiCJE3k.

T. A. Ibisanmi, X. Jiang, M. Willcox, and N. Kumar. Recent advances in computational antimicrobial peptide discovery through big data, modeling, and artificial intelligence and their interplay in ushering the next golden era of drug development. *Frontiers in Bioinformatics*, 6:1749404, 2026. doi: 10.3389/fbinf.2026.1749404.

Bowen Jing, Stephan Eismann, Patricia N. Suriana, Raphael J. L. Townshend, and Ron O. Dror. Learning from protein structure with geometric vector perceptrons. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=1YLJDvSx6J4.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60:84–90, 2017. doi: 10.1145/3065386.

A. Lavertu, J. Corbeil, and P. Germain. Qmap: A benchmark for standardized evaluation of antimicrobial peptide mic and hemolytic activity regression. *bioRxiv*, 2026. doi: 10.64898/2026.02.03.703041.

H. Lee, S. Lee, I. Lee, et al. Amp-bert: Prediction of antimicrobial peptide function based on a bert model. *Protein Science*, 32:e4529, 2023. doi: 10.1002/pro.4529.

J. Li, Y. Shao, Y. Li, and Q. Yu. Integrative multi-scale sequence–structure modeling for antimicrobial peptide prediction and design. *bioRxiv*, 2026. doi: 10.64898/2026.02.24.707610.

Z. Lin, H. Akin, R. Rao, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379:1123–1130, 2023. doi: 10.1126/science.ade2574.

Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Eureka: Human-level reward design via coding large language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=IEduRUO55F.

Ricardo Macarron, Martyn N Banks, Dejan Bojanic, David J Burns, Dragan A Cirovic, Tina Garyantes, Darren VS Green, Robert P Hertzberg, William P Janzen, Jeff W Paslay, et al. Impact of high-throughput screening in biomedical research. *Nature Reviews Drug Discovery*, 10(3):188–195, 2011.

C. J. L. Murray, K. S. Ikuta, F. Sharara, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet*, 399(10325):629–655, 2022. doi: 10.1016/S0140-6736(21)02724-0.

M. Naghavi, S. E. Vollset, K. S. Ikuta, et al. Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet*, 404(10459):1199–1226, 2024. doi: 10.1016/S0140-6736(24)01867-1.

N. G. Oliveira Júnior, C. M. Souza, D. F. Buccini, M. H. Cardoso, and O. L. Franco. Antimicrobial peptides: structure, functions and translational applications. *Nature Reviews Microbiology*, 11:1–14, 2025. doi: 10.1038/s41579-025-01200-y.

H. T. Pham, B. Huynh, and T.-H. Nguyen-Vo. Genpept-curated-2025: A benchmark dataset for antimicrobial peptide prediction with homology-controlled partitioning. *bioRxiv*, 2026. doi: 10.64898/2026.04.25.720793.

Kiersten M Ruff and Rohit V Pappu. Alphafold and intrinsically disordered proteins. *Journal of Molecular Biology*, 433(20):167208, 2021.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332, 2021.

Kristof T. Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Ritesh Sharma, Sameer Shrivastava, Sanjay Kumar Singh, Abhinav Kumar, Sonal Saxena, and Raj Kumar Singh. Deep-abppred: identifying antibacterial peptides in protein sequences using bidirectional lstm with word2vec. *Briefings in Bioinformatics*, 22(5):bbab065, 2021. doi: 10.1093/bib/bbab065.

Z. Shu, Y. Deng, H. Zhang, Z. Nie, and J. Chen. Mtpnet: multi-grained target perception for unified activity cliff prediction. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI '25)*, pp. 7733–7741, 2025. doi: 10.24963/ijcai.2025/860.

Taha ValizadehAslani, Zhengqiao Zhao, Bahrad A. Sokhansanj, and Gail L. Rosen. Amino acid k-mer feature extraction for quantitative antimicrobial resistance (amr) prediction by machine learning and model interpretation for biological insights. *Biology*, 9(11):365, 2020. ISSN 2079-7737. doi: 10.3390/biology9110365. URL https://www.mdpi.com/2079-7737/9/11/365.

Daniel Veltri, Uday Kamath, and Amarda Shehu. Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, 34(16):2740–2747, 2018. doi: 10.1093/bioinformatics/bty179.

Jacob Witten and Zack Witten. Deep learning regression model for antimicrobial peptide design. *bioRxiv*, 2019. doi: 10.1101/692681. URL https://www.biorxiv.org/content/early/2019/07/12/692681.

Jielu Yan, Bob Zhang, Mingliang Zhou, François-Xavier Campbell-Valois, and Shirley W. I. Siu. A deep learning method for predicting the minimum inhibitory concentration of antimicrobial peptides against ¡i¿escherichia coli¡/i¿ using multi-branch-cnn and attention. *mSystems*, 8(4):e00345–23, 2023. doi: 10.1128/msystems.00345-23. URL https://journals.asm.org/doi/abs/10.1128/msystems.00345-23.

Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, et al. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.

Y. Yang, K. Zha, Y. Chen, H. Wang, and D. Katabi. Delving into deep imbalanced regression. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pp. 11842–11851, 2021. URL https://proceedings.mlr.press/v139/yang21m.html.

M. Yasir, M. D. P. Willcox, and D. Dutta. Action of antimicrobial peptides against bacterial biofilms. *Materials*, 11(12):2468, 2018. doi: 10.3390/ma11122468.

Z. Yuan, P. Ouyang, K. Gu, et al. The antibacterial mechanism of oridonin against methicillin-resistant *Staphylococcus aureus* (mrsa). *Pharmaceutical Biology*, 57(1):710–716, 2019. doi: 10.1080/13880209.2019.1674342.

Y. Zhang, J. Lin, L. Zhao, et al. A novel antibacterial peptide recognition algorithm based on bert. *Briefings in Bioinformatics*, 22:bbab200, 2021. doi: 10.1093/bib/bbab200.

Gen Zhou, Sugitha Janarthanan, Lianghong Chen, and Pingzhao Hu. MAC-AMP: A closed-loop multi-agent collaboration system for multi-objective antimicrobial peptide design, 2026. URL https://arxiv.org/abs/2602.14926.

APPENDIX

A DATA PREPROCESSING AND PARTITIONING

During data preprocessing, we implemented stringent filtering and cleaning protocols on the raw data to ensure that the model learned accurate, highly generalizable, and biologically meaningful sequence–activity mapping relationships. The specific data screening pipeline was as follows: (1) only monomeric peptide sequences were retained, excluding multimers or peptides with complex topologies; (2) samples containing noncanonical or chemically modified amino acids were eliminated; (3) entries containing unknown amino acid residues (denoted by "X") were filtered out; and (4) records lacking explicit MIC values were removed.

For the valid sequences obtained after cleaning, we standardized the activity labels. Initially, all target MIC values were uniformly converted to micromolar (μM) units. In real-world experimental data, a single peptide often has multiple MIC records against the same target bacterial species because of differences in the reference literature or measurement batches. To mitigate experimental noise and ensure label consistency, we aggregated these duplicate entries by calculating the arithmetic mean of the corresponding MIC measurements, thereby yielding a single consensus MIC value for each "peptide-target species" pair. Finally, considering the extreme magnitude span and severely right-skewed long-tail distribution of the raw MIC values, we applied a base-10 logarithmic transformation, $y = \log_{10}(\text{MIC})$, to the aggregated activity values. This operation maps the exponentially spanning concentrations into a more constrained and normally distributed space, which is crucial for stabilizing gradients during model training and mitigating numerical instability.

In biological sequence prediction tasks, traditional random splitting often leads to the presence of highly homologous sequences between the training and test sets. This overlap causes data leakage and artificially inflates model performance metrics, leading to severe misjudgment of the model's true generalization ability. To address this issue, we employed the graph-based splitting engine from the QMAP benchmark toolkit to perform strict, homology-aware data splitting (Lavertu et al., 2026). Specifically, we followed the original QMAP protocol by setting the sequence identity threshold to 60% and enabling the postfiltering mechanism, thereby ensuring that the homology between any sequence in the test set and those in the training set remained strictly below 60%. Under these conditions, the dataset for each species was partitioned into a training set and an independent test set at a ratio as close to 8:2 as possible (minor deviations in exact proportions might have occurred because of the stochastic nature of the graph-based splitting engine). Furthermore, to guarantee the rigor and reproducibility of our experimental results, we adhered to the QMAP evaluation protocol by employing five fixed random seeds (1, 3, 7, 12, and 404) to independently generate five sets of data splits, thus ensuring robust model performance evaluation. By implementing this low-homology, deterministic splitting strategy, we have not only eliminated the evaluation fluctuations caused by underlying random engines but also established a reliable and balanced foundation for verifying the model's true predictive capability on unseen sequences.

B PERFORMANCE ON DBAASP *P. aeruginosa* DATASET

We further evaluate the effects of GOA-PGN on *P. aeruginosa* subset of DBAASP. Notably, the LLAMP baseline is excluded from this specific evaluation, as the official LLAMP repository exclusively provides the genome for *E. coli*, we omitted its ability to reproduce on other strains to prevent potential genome mismatches and ensure a rigorous, fair comparison. The results for both the full dataset and the high-potency subset are summarized in Table 3.

As presented in the upper panel of Table 3, the GOA-PGN maintains a substantial predictive advantage across the full *P. aeruginosa* dataset, achieving the best performance in all the evaluation metrics. This highlights the broad generalizability of capturing explicit side-chain topological orientations across different bacterial strains.

However, evaluating the high-potency regime (lower panel of Table 3) reveals a distinct biophysical challenge compared with *E. coli*. Owing to the complex outer-membrane permeability barrier and specific lipopolysaccharide (LPS) interactions characteristic of *P. aeruginosa*, the overall ranking correlations suffer noticeable degradation across all the evaluated architectures. Despite this chal-

Table 3: Performance comparison on the DBAASP *P. aeruginosa* dataset. The upper panel presents the results evaluated on the full dataset, while the lower panel details the performance on the high-potency subset (MIC < 10 μM). The best results in each category are highlighted in **bold**.

Architecture	Spearman ↑	Pearson ↑	Kendall's τ ↑	RMSE ↓	MAE ↓	MSE ↓
Panel A: Full Dataset						
Witten Witten & Witten (2019)	0.371 ± 0.052	0.362 ± 0.043	0.256 ± 0.040	0.662 ± 0.071	0.517 ± 0.038	0.443 ± 0.096
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.288 ± 0.049	0.297 ± 0.027	0.197 ± 0.034	0.670 ± 0.080	0.528 ± 0.057	0.454 ± 0.106
ESM2-Linear Lin et al. (2023)	0.270 ± 0.059	0.250 ± 0.053	0.183 ± 0.041	0.922 ± 0.034	0.696 ± 0.024	0.852 ± 0.062
MBC attention Yan et al. (2023)	0.378 ± 0.056	0.366 ± 0.054	0.259 ± 0.040	0.664 ± 0.081	0.521 ± 0.050	0.446 ± 0.109
BERT-AMPEP60-ES Cai et al. (2025)	0.268 ± 0.109	0.283 ± 0.089	0.186 ± 0.077	0.682 ± 0.073	0.547 ± 0.055	0.469 ± 0.096
ANIA Chiu et al. (2026)	0.369 ± 0.071	0.371 ± 0.048	0.251 ± 0.048	0.678 ± 0.064	0.535 ± 0.039	0.463 ± 0.087
GOA-PGN (Ours)	$\mathbf{0.436 \pm 0.062}$	$\mathbf{0.419 \pm 0.071}$	$\mathbf{0.304 \pm 0.045}$	$\mathbf{0.651 \pm 0.079}$	$\mathbf{0.502 \pm 0.042}$	$\mathbf{0.428 \pm 0.106}$
Panel B: High-Potency Subset (MIC < 10 μM)						
Witten Witten & Witten (2019)	0.163 ± 0.116	0.119 ± 0.233	0.112 ± 0.075	0.811 ± 0.151	0.657 ± 0.092	0.676 ± 0.259
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.196 ± 0.067	$\mathbf{0.193 \pm 0.097}$	0.134 ± 0.047	0.838 ± 0.109	0.712 ± 0.069	0.711 ± 0.194
ESM2-Linear Lin et al. (2023)	0.116 ± 0.155	0.099 ± 0.189	0.079 ± 0.106	0.984 ± 0.097	0.763 ± 0.070	0.976 ± 0.195
MBC-Attention Yan et al. (2023)	$\mathbf{0.201 \pm 0.128}$	0.152 ± 0.215	$\mathbf{0.139 \pm 0.083}$	0.776 ± 0.114	0.610 ± 0.047	0.613 ± 0.183
BERT-AMPEP60-ES Cai et al. (2025)	0.159 ± 0.055	0.136 ± 0.090	0.111 ± 0.035	0.838 ± 0.098	0.723 ± 0.099	0.710 ± 0.169
ANIA Chiu et al. (2026)	0.123 ± 0.127	0.106 ± 0.188	0.084 ± 0.090	0.838 ± 0.077	0.676 ± 0.054	0.707 ± 0.130
GOA-PGN (Ours)	0.186 ± 0.139	0.132 ± 0.235	0.128 ± 0.091	$\mathbf{0.720 \pm 0.145}$	$\mathbf{0.541 \pm 0.070}$	$\mathbf{0.535 \pm 0.230}$

lenging noise regime, GOA-PGN demonstrates exceptional resilience because it achieves the lowest prediction errors (RMSE, MAE, and MSE) by a substantial margin. While purely sequence-based or statistical baselines exhibit erratic prediction variance, our topological graph acts as a rigorous biophysical constraint. By anchoring predictions to fundamental amphipathic orientations, GOA-PGN successfully prevents catastrophic magnitude mispredictions, confirming that distance-agnostic topological priors provide robust stability even under highly strain-specific biological perturbations.

Table 4: Performance comparison on the DBAASP *S. aureus* dataset. The upper panel presents the results evaluated on the full dataset, while the lower panel details the performance on the high-potency subset (MIC < 10 μM). The best results in each category are highlighted in **bold**.

Architecture	Spearman ↑	Pearson ↑	Kendall's τ ↑	RMSE ↓	MAE ↓	MSE ↓
Panel A: Full Dataset						
Witten Witten & Witten (2019)	0.333 ± 0.045	0.359 ± 0.032	0.227 ± 0.032	0.726 ± 0.026	0.585 ± 0.024	0.528 ± 0.038
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.247 ± 0.039	0.240 ± 0.043	0.168 ± 0.027	0.743 ± 0.026	0.599 ± 0.014	0.553 ± 0.038
ESM2-Linear Lin et al. (2023)	0.319 ± 0.019	0.330 ± 0.028	0.217 ± 0.014	0.848 ± 0.042	0.662 ± 0.025	0.721 ± 0.073
MBC-Attention Yan et al. (2023)	0.398 ± 0.049	0.418 ± 0.053	0.274 ± 0.035	0.694 ± 0.017	0.559 ± 0.018	0.482 ± 0.024
BERT-AMPEP60-ES Cai et al. (2025)	0.248 ± 0.058	0.261 ± 0.057	0.172 ± 0.041	0.747 ± 0.014	0.611 ± 0.011	0.557 ± 0.021
ANIA Chiu et al. (2026)	0.348 ± 0.031	0.369 ± 0.032	0.237 ± 0.022	0.751 ± 0.040	0.601 ± 0.027	0.565 ± 0.061
GOA-PGN (Ours)	$\mathbf{0.445 \pm 0.023}$	$\mathbf{0.451 \pm 0.043}$	$\mathbf{0.307 \pm 0.018}$	$\mathbf{0.690 \pm 0.022}$	$\mathbf{0.548 \pm 0.017}$	$\mathbf{0.476 \pm 0.030}$
Panel B: High-Potency Subset (MIC < 10 μM)						
Witten Witten & Witten (2019)	0.080 ± 0.081	0.086 ± 0.067	0.054 ± 0.054	0.805 ± 0.052	0.655 ± 0.048	0.651 ± 0.080
AAk-mer + XGBoost ValizadehAslani et al. (2020)	0.071 ± 0.051	0.090 ± 0.072	0.047 ± 0.034	0.862 ± 0.019	0.737 ± 0.023	0.744 ± 0.032
ESM2-Linear Lin et al. (2023)	0.082 ± 0.059	0.080 ± 0.067	0.056 ± 0.040	0.933 ± 0.067	0.745 ± 0.058	0.874 ± 0.128
MBC-Attention Yan et al. (2023)	0.112 ± 0.036	$\mathbf{0.137 \pm 0.025}$	0.076 ± 0.024	0.787 ± 0.033	0.640 ± 0.035	0.620 ± 0.052
BERT-AMPEP60-ES Cai et al. (2025)	$\mathbf{0.198 \pm 0.070}$	0.113 ± 0.082	$\mathbf{0.137 \pm 0.048}$	0.801 ± 0.099	0.681 ± 0.109	0.649 ± 0.157
ANIA Chiu et al. (2026)	0.117 ± 0.038	0.090 ± 0.038	0.079 ± 0.025	0.877 ± 0.137	0.712 ± 0.121	0.785 ± 0.226
GOA-PGN (Ours)	0.115 ± 0.056	0.133 ± 0.054	0.078 ± 0.037	$\mathbf{0.743 \pm 0.035}$	$\mathbf{0.578 \pm 0.021}$	$\mathbf{0.553 \pm 0.053}$

C Performance on DBAASP *S. aureus* Dataset

We further evaluate the effects of the GOA-PGN on *S. aureus* subset of DBAASP, with results on the full dataset and a high-potency subset (MIC < 10 μM) summarized in Table 4.

As shown in the upper panel of Table 4, GOA-PGN consistently achieves state-of-the-art results across the full dataset, demonstrating its strong generalizability to gram-positive pathogens.

Interestingly, the high-potency regime reveals a striking decoupling between the ranking correlation and absolute prediction error. Unlike our findings in gram-negative bacteria, purely sequence-based architectures (e.g., BERT-AMPEP60-ES) surpass GOA-PGN in terms of rank-based correlations. Biologically, the gram-positive *S. aureus* lacks an outer lipid membrane and instead features a thick peptidoglycan cell wall. Initial target interaction and cell wall penetration here may be predominantly dictated by specific local sequence motifs, which NLP-based models excel at capturing for relative sorting.

However, despite this shift in correlation dominance, the GOA-PGN still robustly maintains the lowest absolute errors (RMSE, MAE, and MSE). This highlights a crucial dichotomy: while 1D sequence patterns may sufficiently guide relative efficacy ranking in gram-positive bacteria, modelling explicit topological amphipathicity acts as a rigorous biophysical anchor, preventing catastrophic magnitude deviations in absolute activity prediction.

D Distance-Relaxed Variant and Implementation Details

To adapt the Global Amphipathic Graph Network for smaller, structurally rigid peptide distributions such as the CAMP dataset, we introduce a distance-relaxed variant of the Topology Constructor. This variant formally reintroduces spatial coordinates into the graph formulation by incorporating an explicit distance scaling factor to modulate the message-passing adjacency matrix.

D.1 Distance-Relaxed Topological Construction

Instead of relying purely on the coordinate-free side chain angular orientation, this formulation computes the Euclidean distance $D_{i,j}$ between the spatial coordinates (e.g., C_α atoms) of residue i and residue j. To guarantee numerical stability during the backwards pass, a small epsilon value (10^{-6}) is added prior to the square root calculation. The modified adjacency matrix element $A_{i,j}$ is defined as follows:

$$A_{i,j} = \exp\left(-\frac{D_{i,j}}{\tau}\right) \cdot \text{ReLU}\left(\frac{v_i \cdot v_j}{\|v_i\|\|v_j\|}\right). \tag{14}$$

In our specific implementation optimized for the CAMP dataset, the distance decay constant τ is explicitly set to 8.0. This specific calibration serves a dual purpose: it successfully integrates the critical spatial features inherent to the CAMP dataset while simultaneously preventing a precipitous decay in the graph's receptive field that would otherwise hinder global message passing.

To construct the valid message-passing matrix, we apply a strict masking strategy. The adjacency matrix is subjected to a Boolean mask that simultaneously zeros out connections involving padded spatial regions and explicitly removes self-loops via an identity matrix constraint. The masked matrix is subsequently normalized by its rowwise degree. To prevent division by zero during normalization, the node degree sum is clamped to a minimum threshold of 10^{-6}.

D.2 Network Dimensionality and PConv1D Configuration

The feature refinement process within the definitive adapter architecture operates within a bottleneck dimension of 256. For the structural channel partitions of the 1D PConv block, the representative subset $H^{(1)}$ is allocated a specific dimension $d_{cp} = 64$, and the remaining untouched subset $H^{(2)}$ comprises the subsequent 192 channels. The 1D convolution that operates exclusively on $H^{(1)}$ uses a kernel size $K = 3$ without a bias term, which is subsequently processed by batch normalization and GELU activation. To integrate the physical graph representation, we employ a zero-initialized gating mechanism via the residual scalar γ. Specifically, γ is defined as a learnable parameter strictly initialized to 0.0. This initialization guarantees that the module temporarily acts as an identity mapping at the start of training, ensuring gradient stability while the network organically learns to intertwine the sequential motifs with the distance-relaxed 3D topologies.

Table 5: Ablation study of GOA-PGN on the DBAASP *E. coli* dataset. The upper panel presents the results evaluated on the full dataset, while the lower panel details the performance on the high-potency subset (MIC < 10 μM). The best results in each category are highlighted in **bold**.

Model Variant	Spearman ↑	Pearson ↑	Kendall's τ ↑	RMSE ↓	MAE ↓	MSE ↓
Panel A: Full Dataset						
w/o PConv	0.526 ± 0.052	0.540 ± 0.040	0.373 ± 0.039	0.656 ± 0.014	0.511 ± 0.009	0.431 ± 0.018
w/o PhysGraph	0.526 ± 0.048	0.546 ± 0.039	0.373 ± 0.038	0.660 ± 0.018	0.512 ± 0.015	0.436 ± 0.023
w/o AttentionPool	0.538 ± 0.051	0.554 ± 0.036	0.383 ± 0.039	0.649 ± 0.016	0.505 ± 0.011	0.421 ± 0.020
w/o SE-Head	0.515 ± 0.049	0.535 ± 0.045	0.365 ± 0.038	0.663 ± 0.015	0.517 ± 0.006	0.440 ± 0.020
w/o Two-Stage	0.525 ± 0.047	0.542 ± 0.040	0.372 ± 0.036	0.658 ± 0.010	0.513 ± 0.012	0.434 ± 0.013
w/o Multi-Task	0.524 ± 0.051	0.546 ± 0.042	0.371 ± 0.037	0.658 ± 0.008	0.510 ± 0.009	0.433 ± 0.011
GOA-PGN (Full)	$\mathbf{0.549 \pm 0.042}$	$\mathbf{0.567 \pm 0.032}$	$\mathbf{0.391 \pm 0.033}$	$\mathbf{0.646 \pm 0.013}$	$\mathbf{0.501 \pm 0.005}$	$\mathbf{0.417 \pm 0.017}$
Panel B: High-Potency Subset (MIC < 10 μM)						
w/o PConv	0.294 ± 0.040	0.270 ± 0.063	0.202 ± 0.028	0.695 ± 0.056	0.522 ± 0.051	0.486 ± 0.077
w/o PhysGraph	0.299 ± 0.065	0.297 ± 0.097	0.206 ± 0.045	0.695 ± 0.037	$\mathbf{0.519 \pm 0.019}$	0.483 ± 0.052
w/o AttentionPool	0.307 ± 0.057	0.305 ± 0.089	0.212 ± 0.039	$\mathbf{0.687 \pm 0.044}$	0.520 ± 0.043	$\mathbf{0.474 \pm 0.059}$
w/o SE-Head	0.282 ± 0.063	0.284 ± 0.094	0.194 ± 0.043	0.723 ± 0.035	0.546 ± 0.011	0.524 ± 0.052
w/o Two-Stage	0.288 ± 0.069	0.273 ± 0.081	0.198 ± 0.048	0.708 ± 0.015	0.538 ± 0.026	0.502 ± 0.021
w/o Multi-Task	0.279 ± 0.063	0.298 ± 0.087	0.191 ± 0.044	0.734 ± 0.017	0.563 ± 0.013	0.539 ± 0.025
GOA-PGN (Full)	$\mathbf{0.313 \pm 0.056}$	$\mathbf{0.315 \pm 0.072}$	$\mathbf{0.216 \pm 0.037}$	0.697 ± 0.034	0.524 ± 0.020	0.486 ± 0.047

E Ablation Analysis

Complementing the high-level ablation summary provided in the main text, this section details the quantitative results and mechanistic insights driving the performance of the GOA-PGN on the DBAASP *E. coli* dataset (Table 5).

Observing the full dataset performance (Panel A), the integration of local sequential contexts and channelwise recalibration proves foundational. Stripping the ZeroParamPhysGraph (w/o PhysGraph) deprives the model of explicit 3D spatial information exchange, causing the Spearman correlation to decrease noticeably to 0.526 and the RMSE to inflate to 0.660. This quantitative degradation confirms that pure 1D convolutions alone are insufficient; explicit topological amphipathicity modelled via side-chain orientations provides indispensable biophysical cues. Furthermore, replacing our adaptive sequence-level attention pooling with a standard mask-aware GAP (w/o AttentionPool) results in suboptimal performance, confirming that the network inherently benefits from allocating varying attention weights to structurally critical anchor residues rather than aggregating all positions uniformly.

Beyond the spatial architecture, Table 5 highlights the mechanical necessity of our training paradigms. Bypassing the two-stage strategy (w/o two-stage) forces the deeply pretrained ESM-2 backbone to coadapt with randomly initialized downstream adapter weights from the very first epoch. This abrupt joint optimization inevitably disrupts the fragile, pretrained evolutionary representations, leading to systematic metric degradation. Similarly, discarding the auxiliary physical predictions (w/o Multi-Task) reduces the model to a naive single-target regressor. We hypothesize that in the absence of the biophysical regularizations imposed by simultaneous charge and hydrophobicity predictions, the network struggles to internalize the true driving forces of membrane disruption, falling back on superficial sequence patterns.

Evaluating these ablations on the high-potency subset (Panel B) further quantifies the rank-error trade-off introduced in the main text. In this constrained elite regime, the full GOA-PGN architecture decisively secures the optimal ranking capability (Spearman 0.313 ± 0.056 and Pearson 0.315 ± 0.072). While omitting specific modules such as AttentionPool or PhysGraph yields marginally lower absolute errors (e.g., w/o AttentionPool achieves an RMSE of 0.687 ± 0.044), this fractional gain in concentration prediction exacts a severe penalty on ranking accuracy. Given that virtual screening pipelines prioritize the correct sorting of true-positive candidates over marginal absolute error minimization, the behavior of the full model is biologically highly actionable. Notably, the Multi-

Task objective proves to be the ultimate anchor in this elite regime: its removal triggers a catastrophic error spike (RMSE escalating to 0.734 ± 0.017). This stark contrast explicitly demonstrates that guiding the network with biophysical priors is the most effective defense against overfitting to sequence biases when highly active, long-tail AMPs are being evaluated.