Microbiology and Infectious Disease
Computational and Systems Biology

Exploring the repository of *de novo* designed bifunctional antimicrobial peptides through deep learning

Ruihan Dong, Rongrong Liu, Ziyu Liu, Yangang Liu, Gaomei Zhao, Honglei Li, Shiyuan Hou, Xiaohan Ma, Huarui Kang, Jing Liu, Fei Guo, Ping Zhao, Junping Wang, Cheng Wang, Xingan Wu, Sheng Ye, Cheng Zhu

Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin Key Laboratory of Function and Application of Biological Macromolecular Structures, School of Life Sciences, Faculty of Medicine, Tianjin University, Tianjin, China • Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China • Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China • Department of Microbiology, Second Military Medical University, Shanghai, China • State Key Laboratory of Trauma and Chemical Poisoning, Institute of Combined Injury of PLA, College of Preventive Medicine, Third Military Medical University (Army Medical University), Chongqing, China • Tianjin Cancer Hospital Airport Hospital, Tianjin, China • School of Computer Science and Engineering, Central South University, Changsha, China

Version of Record
March 13, 2025

Reviewed Preprint
v2 • March 3, 2025
Revised by authors

Reviewed Preprint
v1 • May 28, 2024

eLife Assessment

This study presents a **useful** pipeline for de novo design of antimicrobial peptides active both against bacteria and viruses. The method is based on deep learning, using a GAN generator and a regression tasked to predict antimicrobial activity. The experimental evidence supporting the conclusions is **solid**, with 24 validated peptides, although some additional justifications of the computational strategy would be a plus. This work will be of interest to the community working on machine learning for biomedical applications and specifically on antimicrobial peptides.

https://doi.org/10.7554/eLife.97330.2.sa4

Abstract

Antimicrobial peptides (AMPs) are attractive candidates to combat antibiotic resistance for their capability to target bio-membranes and restrict a wide range of pathogens. It is a daunting challenge to discover novel AMPs due to their sparse distributions in a vast peptide universe, especially for peptides that demonstrate potencies for both bacterial membranes and viral envelopes. Here we establish a *de novo* AMP design framework by bridging a deep generative module and a graph-encoding activity regressor. The generative module learns hidden 'grammars' of AMP features and produces candidates sequentially pass antimicrobial predictor and antiviral classifiers. We discover 16 bifunctional AMPs and experimentally validated their abilities to inhibit a spectrum of pathogens *in vitro* and in animal models. Notably, P076 is a highly potent bactericide with the minimal inhibitory concentration of 0.21

μM against multidrug-resistant *A. baumannii*, while P002 broadly inhibits five enveloped viruses. Our study provides feasible means to uncover sequences that simultaneously encode antimicrobial and antiviral activities, thus bolstering the function spectra of AMPs to combat a wide range of drug-resistant infections.

Introduction

In the post-antibiotic era, microbes rapidly evolve to become antimicrobial-resistant under the pressure of antibiotic abuse and through horizontal transfer of resistance genes.[1] Due to increasing cases of untreatable bacterial infections, the World Health Organization (WHO) has projected a looming health threat of 10 million casualties per year by 2050.[2] Specifically, the WHO has identified six antimicrobial-resistant pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) as ESKAPE. The ESKAPE pathogens have severely limited the treatment options for serious infections and impaired the conditions of immunodeficient patients.[3] In sharp contrast, the majority of clinically available antibiotics were developed more than 30 years ago.[4] Hence, there are urgent needs for targeted generation and screening of novel molecules with desired therapeutic properties for the next-generation broad-spectrum antimicrobial agents.

As a naturally evolved defense line of innate immune system, antimicrobial peptides (AMPs) are cationic macromolecules (usually 10-60 amino acids) widely present in all living organisms.[5] AMPs are attractive alternatives to conventional antibiotic treatment because they effectively inhibit a wide range of pathogens, such as the Gram-positive and Gram-negative bacteria, fungi, and viruses.[5] The majorities of AMPs featured amphiphilic patterns of amino acid sequences and enriched positive charges (e.g., lysine and arginine residues), which allowed AMPs to disrupt the membrane components and exert antimicrobial effects.[6] Due to the relative constant compositions of biological membranes as the essential AMP targets, development of resistance to AMPs were rare events even after 100 passages of lab evolutions.[7, 8]

Notably, the naturally evolved and rationally designed AMP sequences (~20 000) are housed in several databases (APD,[9] DBAASP,[10] DRAMP,[11] etc.) and paired with experimentally validated activities (minimal inhibitory concentration, or MIC), thus facilitating data mining endeavors for more diverse sequences and more effective peptides.[12] Indeed, deep learning models have led to the discovery of highly efficient AMPs (e.g., MIC of 2 μM against *E. coli*). Specifically, predictive models like Deep-AMPEP30,[13] AMPlify,[14] TransImbAMP,[15] and sAMPpred-GAT[16] utilized deep convolutional, recurrent, attention-based, and graph neural networks to learn and recognize latent representations of meaningful peptides. These models recently enabled comprehensive mining of the human gut microbiome,[17] modern and ancient human proteomes,[18, 19] or even the entire sequence space of hexapeptides.[8] Generative models like PepVAE,[20] CLaSS,[21] and HydrAMP[22] reconstructed the sequences from latent space according to the encoder-decoder architecture of variational autoencoders (VAE), while PepGAN[23] and AMPGAN v2[24] utilized the generative adversarial network (GAN) to approximate the antimicrobial sequences by deceiving the discriminator.

Despite the recent advances of data-driven AMP classifiers and generators, the exploration of a vast AMPs repertoire is still hindered by several factors: (1) The sparsity of AMPs in the whole peptide space, which demands accurate prediction of MIC values for efficient sampling and filtering of AMP analogues.[22, 25] The current machine learning predictive models usually perform the binary classification task and label the generated sequences as active or not, instead of giving specific activity values. To overcome this shortcoming, a popular pipeline combines independent modules of filtering, classification, and regression, but it faces risks of step-wise bias accumulation.[8] (2) The impact of different phospholipid species on the efficacy of AMPs. Due to the data incompleteness of antimicrobial measurements, a majority of AMP classifiers or MIC

predictions were specific to *E. coli* (Gram-negative). Nevertheless, these models were applied to find peptides that inhibit other Gram-positive bacteria (e.g., *S. aureus*), based on the assumption that the shared membrane components are potential targets of AMPs. More importantly, the enveloped virus also possesses a targetable membrane, which closely mimic the phospholipid components of cellular organelles. Even though the AMP database recorded antiviral activities of at least 2 000 antiviral peptides (AVPs), whether the current models can readily generate antiviral peptides remain elusive. (3) Lack of rigorous pre-clinical models. To advance the designed AMPs towards clinical trials, the experiments should reflect the efficacy of AMPs in tissue samples and require the determination of safe doses of a particular sequence before practical applications. To our knowledge, more than 120 AMPs have been suggested by deep learning models in the past three years, only a fraction of these studies demonstrated animal results.

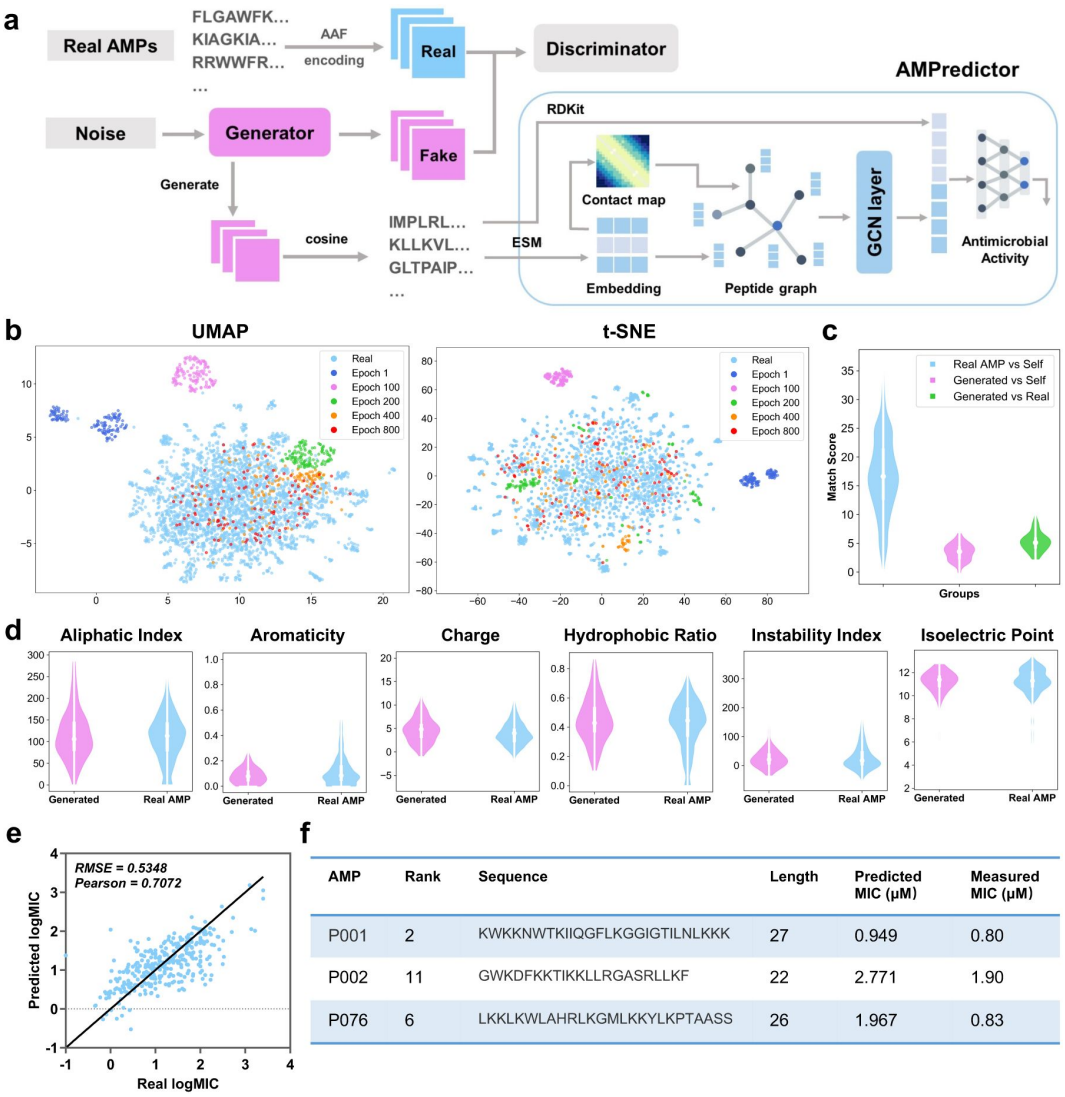
In the current study, we addressed these shortcomings and presented a new deep learning-based *de novo* AMP design framework, consisting of a GAN generator and an antimicrobial activity regressor named AMPredictor. The graph convolution network (GCN) of AMPredictor allowed us to effectively train a regression model for the task of MIC predictions. Through in-depth investigations of generated sequences (e.g., antiviral possibilities), we selected three top-ranking sequences (P001, P002, and P076) and evaluated their efficacies against different pathogens (four drug-resistant bacteria and five enveloped virus). Out of the three candidates, we found P076 is a potent bactericide with MIC of 0.21 μM against multidrug-resistant bacterium. It also outperformed a clinically relevant antibiotic polymyxin B in terms of safety and efficacy, as evaluated in mouse models. Another peptide P002 broadly neutralized viruses in cellular infection models. We challenged our computational models to generate bifunctional AMPs and our results demonstrated that these peptides inhibited a spectrum of pathogens, but with delicate preferences towards bacterial or viral species.

Results

Bridging the generator and regressor for *de novo* AMP design

The overall architecture for designing functional peptides fuses a generative adversarial network and an antimicrobial activity predictor (AMPredictor, **Figure 1a**). We aim to explore beyond the known antimicrobial templates or accessible proteomics space. We trained the GAN module with 3280 AMP sequences to capture the pattern of antimicrobial peptides through few-shot learning. The training goal of GAN is to reach a relative balance between the generator and discriminator (see Methods **Equation 1**). The generator takes random noise as input to produce fake samples, which are fed into the discriminator along with the real data. The discriminator tries to distinguish between the real and fake data, then returns a probability to update the weights of the networks.[26] Here, we employed a low-dimensional vector called amino acid factors (AAF) to encode the peptide sequences and describe their physicochemical features.[27] We monitored the generated sequences at different training stages (epoch 1, 100, 200, 400, 800) and visualized the optimization process using two common dimension reduction approaches, uniform manifold approximation and projection (UMAP) and t-distributed stochastic neighbor embedding (t-SNE). We clearly observed the distributions of newly sampled sequences among the existing AMPs after mapping peptide AAF features onto the two-dimensional space (**Figure 1b**). With iterating and updating, the generator learns the feature distribution of actual data and generates diverse sequences.

We checked the novelty and diversity of generated sequences at epoch 1000 by aligning them with our training set. The sequence similarities were evaluated by global alignment using BLOSUM62 matrix. We found the identity of each generated sequence was around 0.35 to the known database (**Figure 1c**), indicating a low level of similarity to the existing AMPs. With distinct sequence compositions, generated peptides still hold similar physiochemical features to real AMPs, as



AMP	Rank	Sequence	Length	Predicted MIC (μM)	Measured MIC (μM)
P001	2	KWKKNWTKIIQGFLKGGIGTILNLKKK	27	0.949	0.80
P002	11	GWKDFKKTIKKLLRGASRLLKF	22	2.771	1.90
P076	6	LKKLKWLAHRLKGMLKKYLKPTAASS	26	1.967	0.83

Figure 1.

Deep learning-based design of novel antimicrobial peptides.

a. The overall framework of generative adversarial network (GAN) generator and AMPredictor.
b. Visualization of GAN training process via UMAP and t-SNE. The newly generated sequences fit the distribution of known AMP space gradually.
c. Global alignment match scores between the generated sequences and itself and AMPs in the training set.
d. Six physiochemistry properties of generated peptides and real AMP sets.
e. Regression results of AMPredictor model on its test set.
f. The sequences of three selected peptides as well as their predicted and experimentally validated MIC values against *E. coli* ATCC 25922.

evidenced by six descriptors of aliphatic index, aromaticity, net charge, the ratio of hydrophobic amino acids, instability index, and isoelectric point (**Figure 1d**).[28] The distribution of these descriptors closely resembled the physiochemical features of AMPs in the training set. Hence, adopting residue features as the input of GAN generator can capture the functional properties of actual AMPs.

We then developed an independent module AMPredictor to estimate the MIC values of arbitrary peptide sequences. AMPredictor is a GCN-based regression model, equivalent to the 'score function' in virtual drug screen. Unlike the low-dimensional descriptor adopted by GAN generator, we designed distinct representations of peptides in AMPredictor for model accuracy and orthogonal verification of GAN-generated sequences. We chose graph as the representation of peptides, with their amino acid residues as graph nodes and the predicted contact maps as graph edges. The node features were embeddings from the protein language model ESM.[29] The peptide graphs were then fed into a three-layer GCN to learn the hidden 'grammars' of AMP features. In addition to the sequence-level and structural-level features, we transformed the input amino acid sequences into Morgan fingerprints to encode the chemical features at each residual position. This atomic-level information was concatenated with GCN representations before entering a three-layer fully-connected networks to regress on MIC datasets (**Figure 1a**). AMPredictor obtained a low root mean squared error (RMSE) of 0.5348 and a high Pearson correlation coefficient (PCC) of 0.7072 on the test set (10% size of the dataset) (**Figure 1b**). The output of AMPredictor is the logMIC value of an AMP sequence.

By integrating the representations of residual-level, sequence-level, and protein structural level information, AMPredictor obtained the best performance in comparison to its ablating versions (Table S1). The ablation study indicated that the context embeddings from ESM play vital roles in predicting MICs, and the contact map assisted in improvement despite it being a simplified representation of protein structure. We also compared the performance of AMPredictor with five typical deep neural networks as baselines, including the multilayer perceptron (MLP), convolutional neural network (CNN), gated recurrent unit (GRU), long short-term memory (LSTM), and Transformer. AMPredictor outperforms these methods in terms of accuracy at the task of predicting antimicrobial activity values (scatter plots in Figure S2). Hence, a well-trained generator could produce a variety of AMP candidates and then AMPredictor ranked the candidates by MIC values. Together with coarse to fine representations of AMP samples, the GAN and AMPredictor facilitated *de novo* design and *in silico* evaluation of AMPs.

Designing bifunctional antimicrobial peptides

Our study was based on the hypothesis that both bacterial membranes and viral envelopes were potential targets of AMPs, and the dual-function of antibacterial and antiviral could be encoded into one peptide sequence. We directly compared the outcome of two different training sets for two versions of the GAN generator: with or without adding AVP sequences, for their efficiencies of designing bifunctional peptides (**Figure 2a**). The generator v1 was trained solely on collected AMP sequences, and generator v2 was trained on the AMP set and known 1788 AVPs from AVPdb. To establish a sequential workflow and filter out antimicrobial peptides with antiviral capabilities, we then considered assembling several AVP classifiers subsequent to the AMPredictor. As the source data, the AVPdb provided ~2000 sequences with IC_{50} labels.[30] However, data curation resulted in only 759 unique AVPs. We determined that the quantitative data of AVPs were insufficient for training another GCN-based regression model. Instead, binary classification should exhibit robustness against noise of a relatively small dataset. Hence, we merely made yes/no decisions using a collective of five published AVP classification tools (AVPpred,[31] DeepAVP,[32] Deep-AVPpred,[33] ENNAVIA,[34] and AI4AVP,[35] details in Table S2). We obtained 1024 new peptides from each generator and kept the sequences with a predicted MIC value of less than 10 μM by AMPredictor. More than one hundred peptides passed the ensemble filter of all five AVP classifiers. Interestingly, generator v2 did not feature a higher passing ratio, although it was trained with AVPs. Based on their hemolysis and other physicochemical properties, we selected 12

peptides from each generator for experimental validation. We also checked their novelty by BLAST with the training set or DRAMP database and all their E-values are greater than 0.1 or no hits found (Table S3).

All 24 peptides exhibited effective antibacterial capability in inhibiting six different strains involving three multidrug-resistant species (**Figure 2b**). Several peptides obtained remarkable MIC values of less than 1 μM, such as P026 and P036, indicating the effectiveness of our generator for *de novo* designing AMPs. Then we used a stricter rule of 'antibacterial' to define bifunctional peptides. If the MIC value of a peptide was less than 10 μM against at least one strain, we described it as antibacterial. We also tested whether these peptides can inhibit Herpes simplex virus 1 (HSV-1) at the concentration of 10 μM. The 'antiviral' feature of a peptide was checked when significant inhibition was observed in immunofluorescence experiments. There were six and seven bifunctional AMPs out of 12 from generators v1 and v2, respectively (**Figure 2c**). Given that we did not screen on more virus species, the success rate of being antiviral is relatively lower than being antibacterial. This result indicates the capability of our design method for bifunctional AMPs. Through the comparison between generators v1 and v2, we demonstrated that the antiviral feature was 'naturally' encoded in AMP sequences.

We then paired small-scale design with more comprehensive experimental characterization, using four drug-resistant bacteria, five enveloped viruses, and animal infection models, focusing on generator v1. We acquired 104 novel peptides via the *de novo* design framework. Among them, 42 and 26 were predicted to own MIC values below 10 μM and 5 μM, respectively (supplemental spread sheet). The small-scale design procedure did not sacrifice the efficiency of discovering novel peptides and featured a comparable success rate with AMPredictor. We regarded a peptide as an antiviral candidate if four or more positive votes were recorded and 32 peptides met our criteria.

We also conducted a comprehensive evaluation of efficacy and safety *in silico*, to select novel AMPs before experimental validations. Considering the membrane corruption mechanism of most AMPs, we predicted their hemolytic activity as a quick measure of cellular toxicity.[36] Through the ranking of high activity and low toxicity, we picked three potent peptides P001, P002, and P076 (**Figure 1f**), all of which featured amphiphilic signatures of sequence and enriched cationic regions (e.g., helical wheel projections in Figure S3), for *in vitro* and *in vivo* experimental assessment.

In vitro antibacterial and hemolytic assays

We examined the antibacterial activity of P001, P002, and P076 against one Gram-positive (*S. aureus*) and three Gram-negative (*A. baumannii*, *E. coli*, and *P. aeruginosa*) bacteria, including three ESKAPE strains. For each species, we cultured a standard strain and a multidrug-resistant (MDR) strain, and measured the MIC values of individual peptides using the broth microdilution method. All the peptides exhibited inspiring bactericidal properties with MICs ranging from 0.20 to 15.18 μM (corresponding to 0.625 to 40 μg/mL, **Figure 3a**). P001 showed broad-spectrum antibacterial capability with MICs less than 4 μM across all the eight strains of four bacteria species. P002 featured lower MIC values in inhibiting Gram-positive bacteria *S. aureus*, while P076 mainly inhibits Gram-negative bacteria *A. baumannii*, *E. coli*, and *P. aeruginosa*. Notably, the experimentally measured MICs against *E. coli* agreed well with the predictions of AMPredictor (**Figure 1f**), indicating the accuracy of our regression models.

The hemolytic analysis revealed that both P001 and P002 displayed dose-dependent toxicity to mouse erythrocytes (**Figure 3b**), with average hemolysis higher than 50% at high concentrations (20 μM, or 100 μg/mL). Nevertheless, the hemolysis of P076 was lower than 1% at concentrations up to 70 μM. We also inspected the apparent cytotoxicity (CC_{50}) of three peptides on three mammalian cell lines (Vero-E6, A549, Huh-7) using CCK8 assays. For the cell cultures, CC_{50} values of all three peptides were well above 30 μM (Figure S8a). The CC_{50} of P076, specifically,

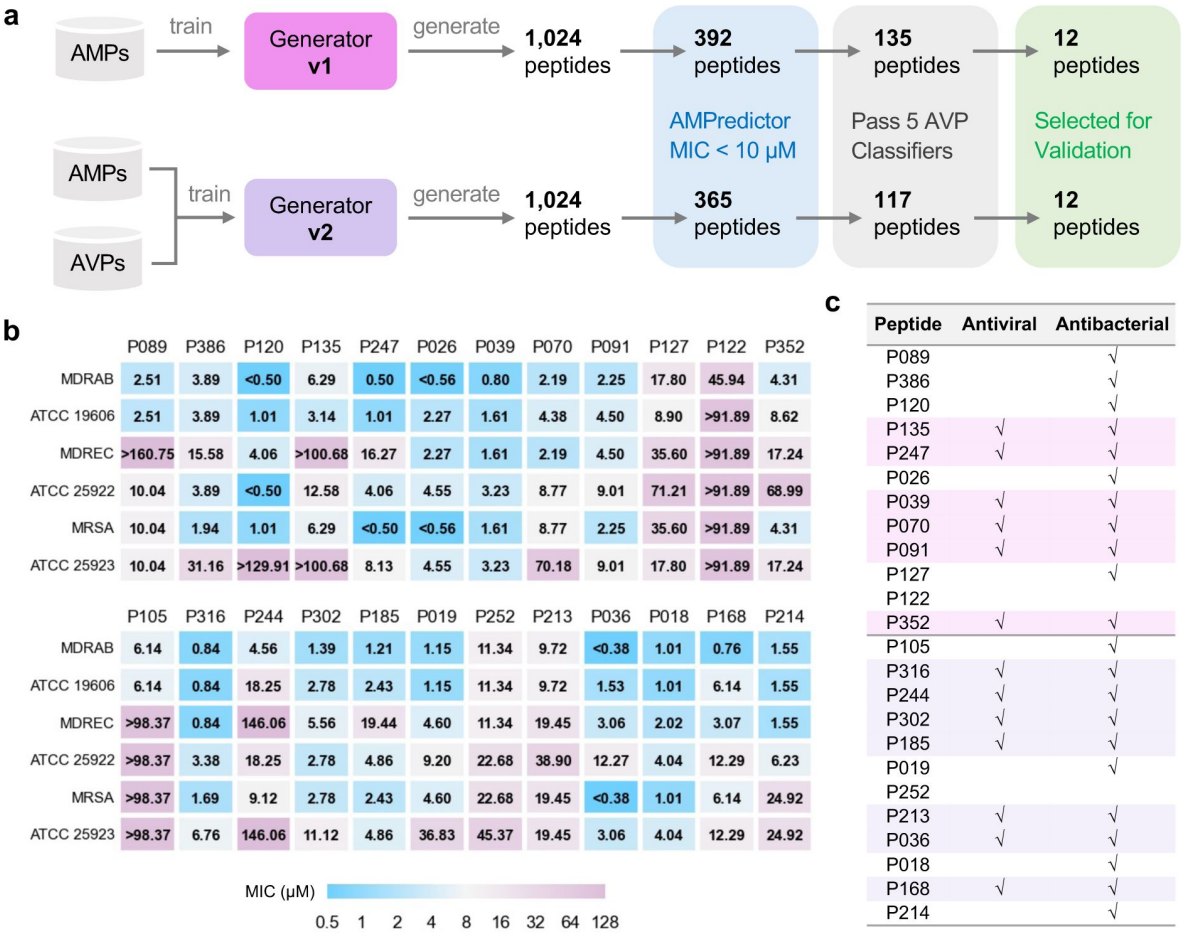


Figure 2.

Large-scale design and validation of bifunctional AMPs.

a. Schemes of designing bifunctional AMPs from two versions of generators. 12 designed peptides from each generator were selected for experimental validation.
b. MIC test of 24 designed peptides against six bacterial strains. The first group included 12 peptides from generator v1, and the other 12 peptides were from generator v2.
c. Antiviral and antibacterial results for bifunctional peptides. The rows with pink or purple backgrounds are bifunctional AMPs acquired from generators v1 or v2, respectively.

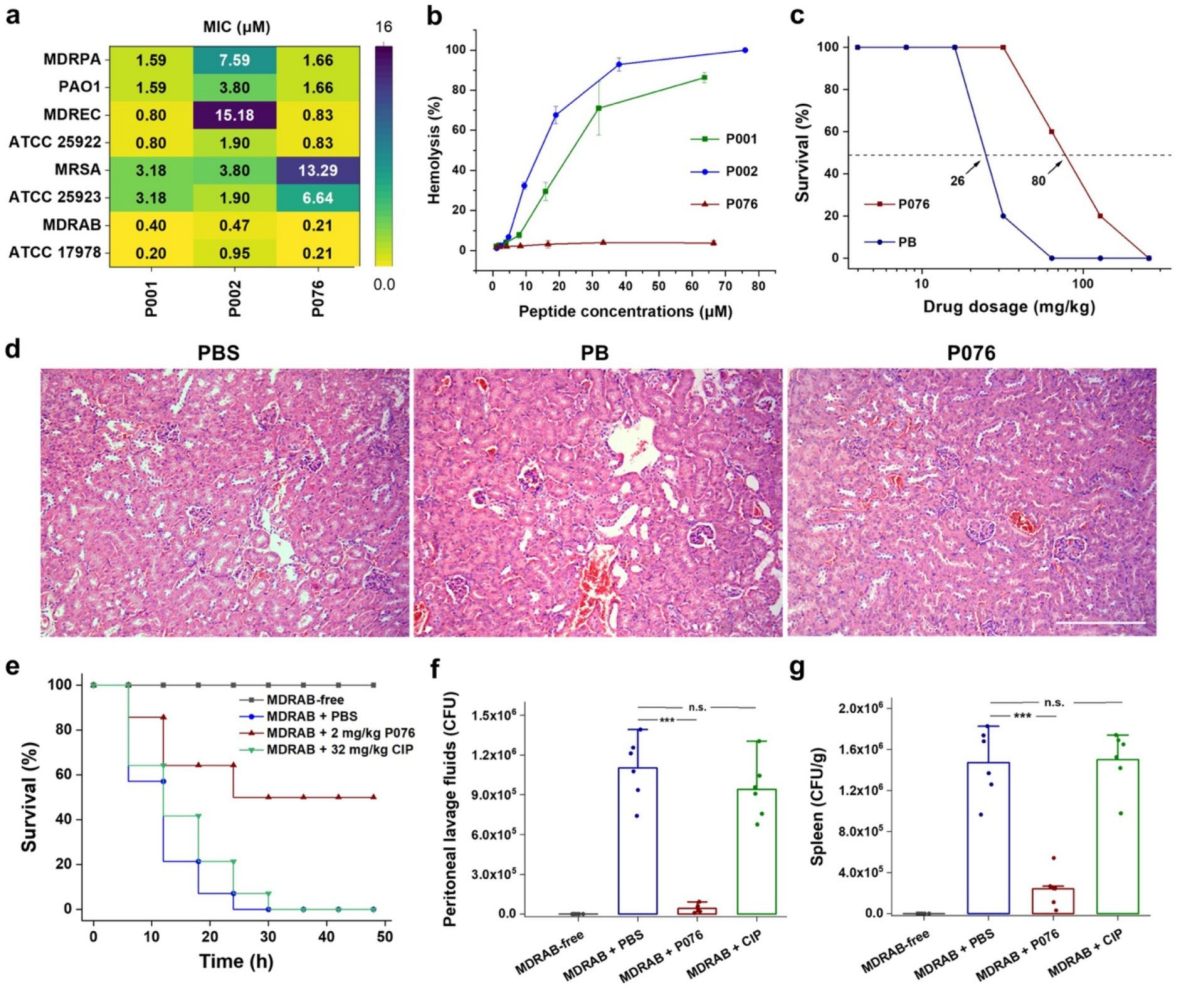


Figure 3.

Antibacterial and toxicological evaluations of P076 peptide.

a. *In vitro* antibacterial assessment of peptides by determination of MICs.
b. Hemolysis of increasing concentrations of peptides. Results are shown as the means ± standard deviation (SDs).
c. Survival curves of mice given P076 and PB. Arrows indicate the drug concentration inducing a 50% death.
d. H&E staining of mouse kidneys. The scale bar indicates 200 μm.
e. Survival of infected mice treated with P076 and CIP.
f. g. Bacterial colonization of MDRAB in mouse peritoneal lavage fluid and spleen. Results are shown as the means ± SDs. ***, P<0.001. n.s., no significance.

were in the range of 35.5 μM (for A549 cells from lung tissues) to 66.8 μM (for Huh-7 cells from liver tissues). The *in vitro* safety assays pointed out thresholds of concentration for administrating these peptides, which were adapted in subsequent animal experiments for estimation of safe dosage.

Our data suggest that all three peptides are potential candidates for inhibiting the clinically relevant *A. baumannii*, which is a top-ranking pathogen with critical priority for development of new antibiotics, as defined by WHO.[3] The MIC of P076 at eliminating the persistent MDRAB reached 0.21 μM, a value half of P001 and P002 (0.40 and 0.47 μM, respectively). Meanwhile, P076 was comparable with P001 but more efficient than P002 against *E. coli* and *P. aeruginosa*. Despite a slight inferiority in killing *S. aureus*, P076 owned a better biosafety. Taken together, the sequence of P076 likely leveraged the balance between efficacy and safety, we therefore focused the following *in vivo* evaluation and animal infection models on P076.

In vivo antibacterial and toxicological evaluations

We first determined a safe dosage for administrating P076 on mouse models. A total of 140 C57 mice (randomly divided into 14 groups) were given increasing concentrations of P076 through intraperitoneal injection, and the cyclic peptide antibiotic polymyxin B (PB) was employed as a positive control. In spite of nephrotoxicity, PB is one of the polymyxin antibiotics that is considered “the last ditch” for the treatment of infections caused by extensively drug-resistant Gram-negative bacteria in clinic.[37] Animal survival was monitored for 48 h after a single drug administration. The P076 concentration to induce a 50% death was calculated to be 80 mg/kg, approximately 3-fold higher than that of PB (26 mg/kg, **Figure 3c**). Hematoxylin-eosin (HE) staining further verified that no pathological changes were observed in the mice liver, spleen and kidneys of mice given 32 mg/kg P076 after 12 h, whereas an apparent renal tubular injury was discovered in PB-treated mouse (**Figure 3c** and S5a).

Encouraged by the biocompatibility of P076, we further conducted an *in vivo* antibacterial assay using a mouse peritoneal infection model.[38] A total of 3×10^7 CFU of MDRAB was intraperitoneally injected in C57 mouse to establish infection symptoms. P076 was administered at 2 mg/kg at 0.5 and 2.5 h after bacterial invasion. As the positive control treatment, we employed a single injection of 32 mg/kg ciprofloxacin (CIP) at 0.5 h after infection. The mice given PBS (blank buffer) or CIP all deceased after 48 h, while 50% of the mice treated with P076 remained alive (**Figure 3e**). Additional investigations revealed that the bacterial colonization in the peritoneal lavage fluid of PBS-treated mouse was averagely 1.1×10^6 CFU, which was 25 times higher than that in the mice treated with P076 (4.2×10^4 CFU, $P<0.001$), whereas CIP featured less of effect on the number of bacteria (**Figure 3f**). Similarly, P076 dramatically decreased the bacterial loads in the mouse spleen (about seven times lower, **Figure 3g** and S5b), thus demonstrating the prospect of P076 to be developed into a new antibiotic.

Broad-spectrum antiviral assays

A cohort of traditional antimicrobial peptides have been rediscovered with antiviral abilities.[39] One significant reason is that their membrane-disrupting capacity can adapt to both bacterial membranes and viral envelopes, which may be captured by AMPredictor and AVP classifier filters. Thus, we tested the antiviral activities of P001, P002, and P076, which passed our dual filters. Our experiments included five typical enveloped viruses. Chikungunya virus (CHIKV), Hantaan virus (HTNV), Dengue virus 2 (DENV-2), Herpes simplex virus 1 (HSV-1), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, and the BA.2. omicron variant) are all pathogens of the past or ongoing epidemics, such as the chikungunya fever, hemorrhagic fever with renal syndrome, dengue, encephalitis, and COVID-19. Hence, the peptide drugs may be repurposed as preventions against viral infections.

We incubated the peptides with different viruses before they invaded the cultured cells and self-amplified. We analyzed the potential antiviral effects of P001, P002, P076 at the mRNA and protein levels, respectively. At the mRNA level, the quantitative real-time PCR suggested dose-dependent inhibition of viral-specific gene expressions (E1 gene for CHIKV, S gene for HTNV, VP16 gene for HSV-1, NS5a gene for DENV-2, ORF for SARS-CoV-2, **Figure 4a** and Figure S6b). At the protein level, we applied fluorescence-conjugated antibodies to visualize trace amounts of virus populations. Corroborating with the RT-PCR results, P002 nearly abolished all viral infections at 12.5 μM (**Figure 4b**, Figure S6a and Figure S7). To quantify the immunofluorescence results, we plotted the log-scale fluorescent focus forming unit (FFU) against varying peptide concentrations (Figure S8b). Among the three peptides, P002 exhibited the highest potential with EC_{50} of 0.37 μM (against CHIKV) to 2.08 μM (against HTNV) (**Figure 4b**, Table S4). The EC_{50} of P076 and P001 were in the range of 1.62 to 2.67 μM (except for inhibiting DENV-2).

Additionally, we tested the antiviral effects of P135 and P244 from our second round of generation. Both the immunofluorescence and qRT-PCR results showed that these AMPs were active broad-spectrum antiviral peptides against four viruses (Figure S9, S10). In particular, these two AMPs showed a significant inhibition on DENV-2, for which P135 obtained an EC_{50} of 0.09 μM (Table S3). Hence, we demonstrated the effectiveness of GAN and AMPredictor combined pipeline on discovering new antiviral peptides.

Combining the EC_{50} and CC_{50} calculations, we calculated the selectivity index ($SI = CC_{50}/ EC_{50}$) as the experimental guides of safety versus efficacy (Table S5). The SI of P002 against four viruses are 149.21, 31.75, 51.05, and 59.01, three of which outperformed the SI of P001 and P076, especially for the case of HTNV. Hence, we further observed the performance of P002 with transmission electron microscopy. For the control set, the viruses were allowed to propagate in cells for days and identified as assembled particles in cytosol (**Figure 5a, b**). When the viruses were mixed with 10 μM P002 before addition to the cells, the viral particles became absent while the cell morphology kept intact, indicating the broad-spectrum inhibition effects of P002 (**Figure 5c, d**).

Because we observed a gradient of antimicrobial and antiviral activities from P001, P002, and P076 with their individual preferences against bacteria and viruses, we aimed to probe the interactions between peptides and different phospholipid assemblies, thus deriving feasible means judging their underlying preferences. We performed all-atom molecular dynamics simulations with molecular models of bacterial membranes and viral envelopes, placed each peptide adjacent to the phospholipid bilayers, and simulated them attaching to the membrane (Figure S11a, b). For the model of Gram-negative bacteria systems, the MM/PBSA calculations and ΔG_{bind} distinguished P002 from P076 and P001 (Table S10). Evaluated by ΔG_{bind} (-465.79 ± 11.39 kJ/mol), P076 would associate with the membrane of Gram-negative bacteria tighter than others. For the model of viral envelopes, all binding energies were weaker than that to bacteria. The ranking of $\Delta G_{nonpolar}$ corroborated with the antiviral potential of P002. Combining this energy term with a geometrical parameter (average distance between the COM of peptide and phosphorus atoms in the top leaflet on z-axis) may predict the behavior of P002 better than $\Delta G_{nonpolar}$ alone (Figure S11c).

Membrane-penetration mechanisms

We hypothesized that a common target among bacteria and viruses was the membrane components. To shed light on the possible membrane-penetration abilities of designed peptides, we applied quantitative measurements, first by focusing on signature components of Gram-negative bacterial outer membranes. We measured the binding of three peptides to lipopolysaccharide (LPS) via biolayer interferometry (BLI). More P076 and P001 peptides were recruited to immobilized LPS than that of P002 at 500 nM (**Figure 6b**). Further experiments demonstrated that P076 bound to lipid A, the innermost part of LPS, at an equilibrium dissociation constant (K_D) of 56.1 nM. This binding affinity of P076 to lipid A was slightly higher than that of P001 (65.7 nM) and was approximately 2.4-fold less than that of P002 (135 nM, **Figure 6a**). Then we directly investigated the membrane permeability by 1-N-phenylnaphthylamine (NPN) uptake

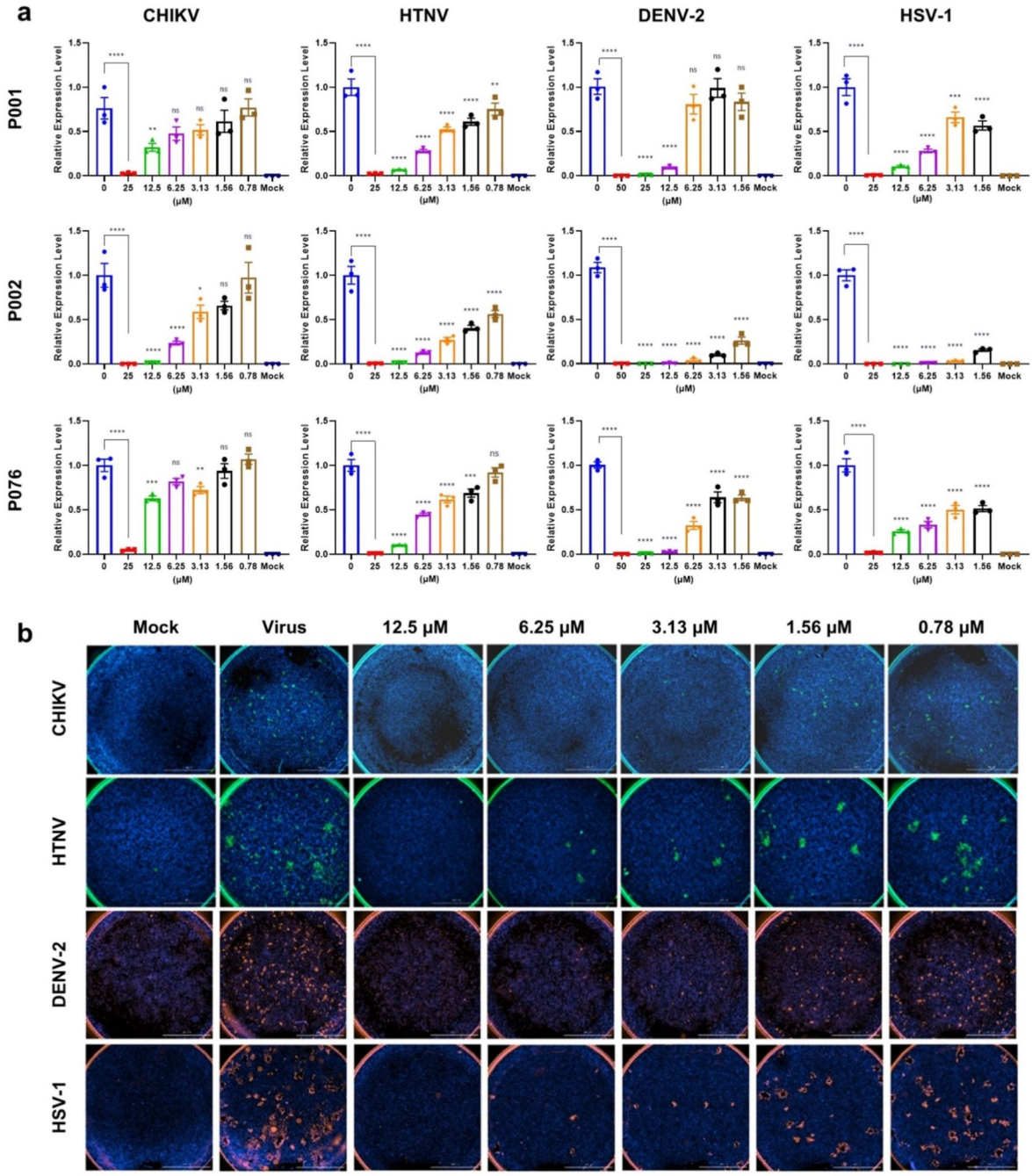


Figure 4.

Antiviral assays of AMPs against CHIKV, HTNV, DENV-2, and HSV-1.

a. Quantitative real-time PCR of virus RNA with gradient concentrations of P001, P002, and P076.
b. Immunofluorescence results of four viruses with gradient concentrations of P002.
All experiments are performed in triplicates. Column bars are means ± SEMs. * P<0.05, ** P<0.01, *** P<0.001, **** P<0.0001. ns, no significance.

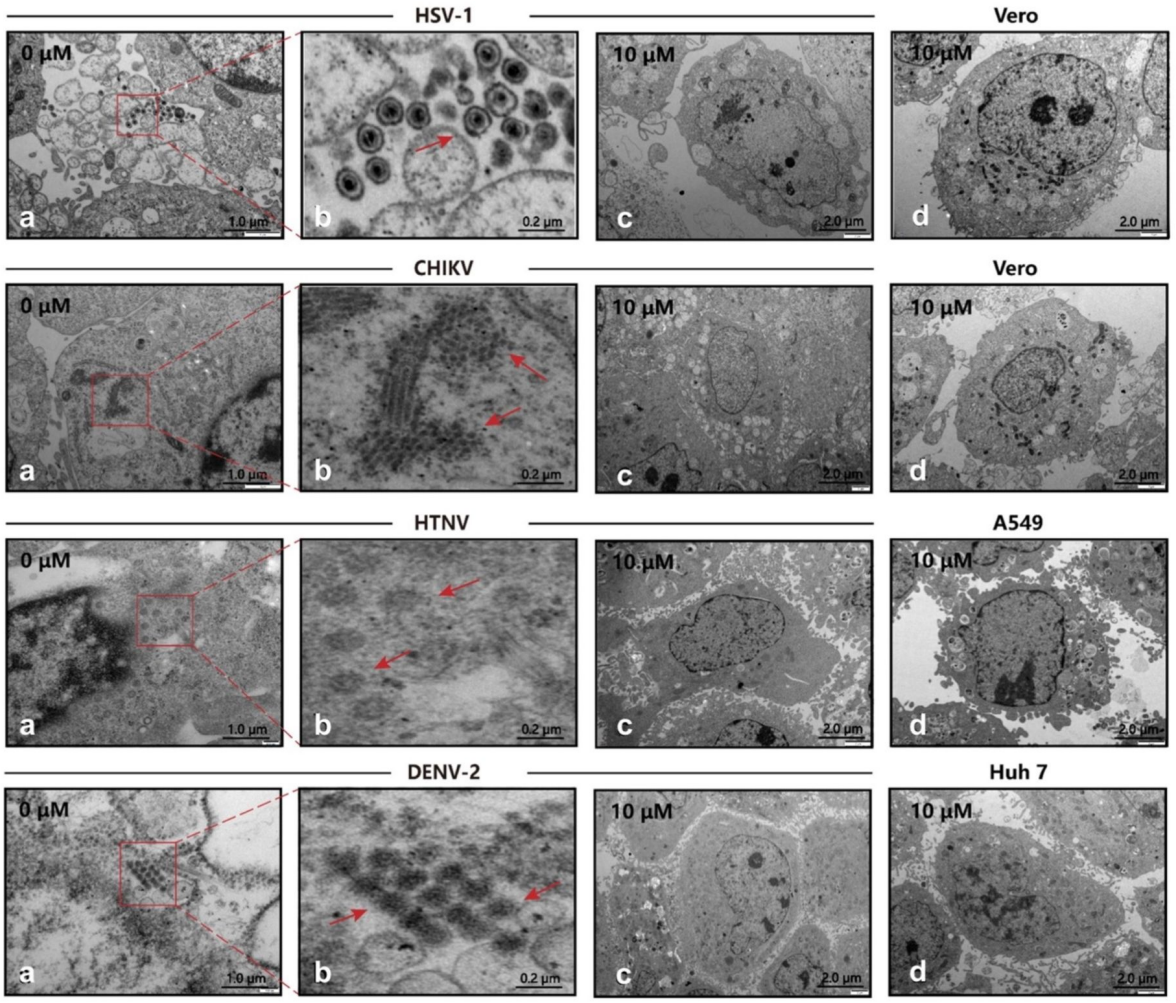


Figure 5.

Transmission electron microscopy (TEM) of viral infection with P002.

a. b. Cells infected with virus, where the red arrows mark the viral particles.
c. Cells are treated with 10 μM P002 and infected with virus (MOI=1), where no viral proliferation is observed.
d. Control of intact cells with 10 μM P002. All experiments are performed in triplicates.

assay. NPN is a fluorescent dye for detecting the impaired outer membrane of Gram-negative bacteria. P076 elicited a stronger membrane penetration relative to P001 and P002 (wavelength scanning in **Figure 6c**). Together, our results suggested that P076 could efficiently bind to and damage bacterial membranes.

To gain a more comprehensive understanding of AMP-membrane interactions with fine content of amphiphilic assemblies (both outer membranes and inner membranes), we also observed the membrane attachment and partial insertion of AMPs via all-atom molecular dynamics simulations. The top leaflet of Gram-negative bacterial outer membranes was built as simplified *E. coli* LPS (lipid A and R1 core). With an initial equilibrated position above the membrane surface, the peptide started to approach the oligosaccharides of LPS and then bind to them. During the simulation processes, P002 and P076 gradually inserted into the outer leaflet, and represented poses binding to LPS at 100 ns (**Figure 6e**). We also measured the distance between the COM of peptide and phosphorus atoms of lipid A on z-axis in the latter 50 ns of trajectories (**Figure 6d**). Compared with the initial value about 4 nm (showing as dashed line), this metric of P002 and P076 declined to less than 2 nm. Nevertheless, P001 performed weaker interactions with an increasing distance, which was consistent with the membrane-disrupting experiment in **Figure 6c**. Moreover, three AMPs tended to attach to the membrane in the MD simulations with lipid bilayer system mimicking the inner membranes of Gram-negative bacteria (Figure S11b). Collectively, experimental characterizations along with computational simulations indicate the membrane-acting mechanisms of our AMPs when playing antimicrobial roles. Further explanation of membrane-disrupting details or other potential biomolecular targets remains to be explored in the future.

Discussion

The traditional 'lock and key' model for therapeutic molecule design requires knowledge of target structures.[40] When the targets are biological membranes, the means of docking and ranking by binding affinities (K_D) can easily become futile. The rapid development of deep neural networks, on the other hand, has facilitated self-supervised learning of meaningful molecule representations in hidden dimensions.[29] Combined with vast data of known AMPs, numerous deep learning models have enabled mining of new antibiotics.[18], [19] In the current study, we deployed a new combination of descriptors to extract representations of biological sequences at residual and protein levels. First, we transformed input amino acid sequences into Morgan fingerprints to encode the chemical features of each residue. Second, the scaffolds of AMPs are closely relative to their antimicrobial function, but rarely considered at the structural level in recently reported models (except for sAMPpred-GAT[16]). Instead of using limited amounts of crystal structures, we introduced predicted contact maps of peptides as the edge encoding of peptide graphs. Finally, we added the embedding from ESM as the residue node features to present the associations of a single peptide sequence in the large protein repertoire. ESM was trained on millions of natural protein sequences via mask attention mechanism and assisted in few-shot learning via transferring the related knowledge of protein sequences.[29] The resulting AMPredictor accurately predicted the MICs of three *de novo* design AMPs in our study. As an independent test on 28 new AMPs discovered by different models, AMPredictor also scored well (Table S6).

Despite the ability of deep learning model to guide targeted generation and screen of functional peptides, it lacks the merit of molecular docking program to explicitly consider subtle environmental features. For a particular sequence, we found it was difficult to tell its preference towards bacterial membranes or viral envelopes, due to slight differences in their lipid compositions. Consequently, we observed broad antimicrobial activity for P001, broad antiviral activity for P002, while P076 demonstrated superior capacity on *A. baumannii*. As a control, we

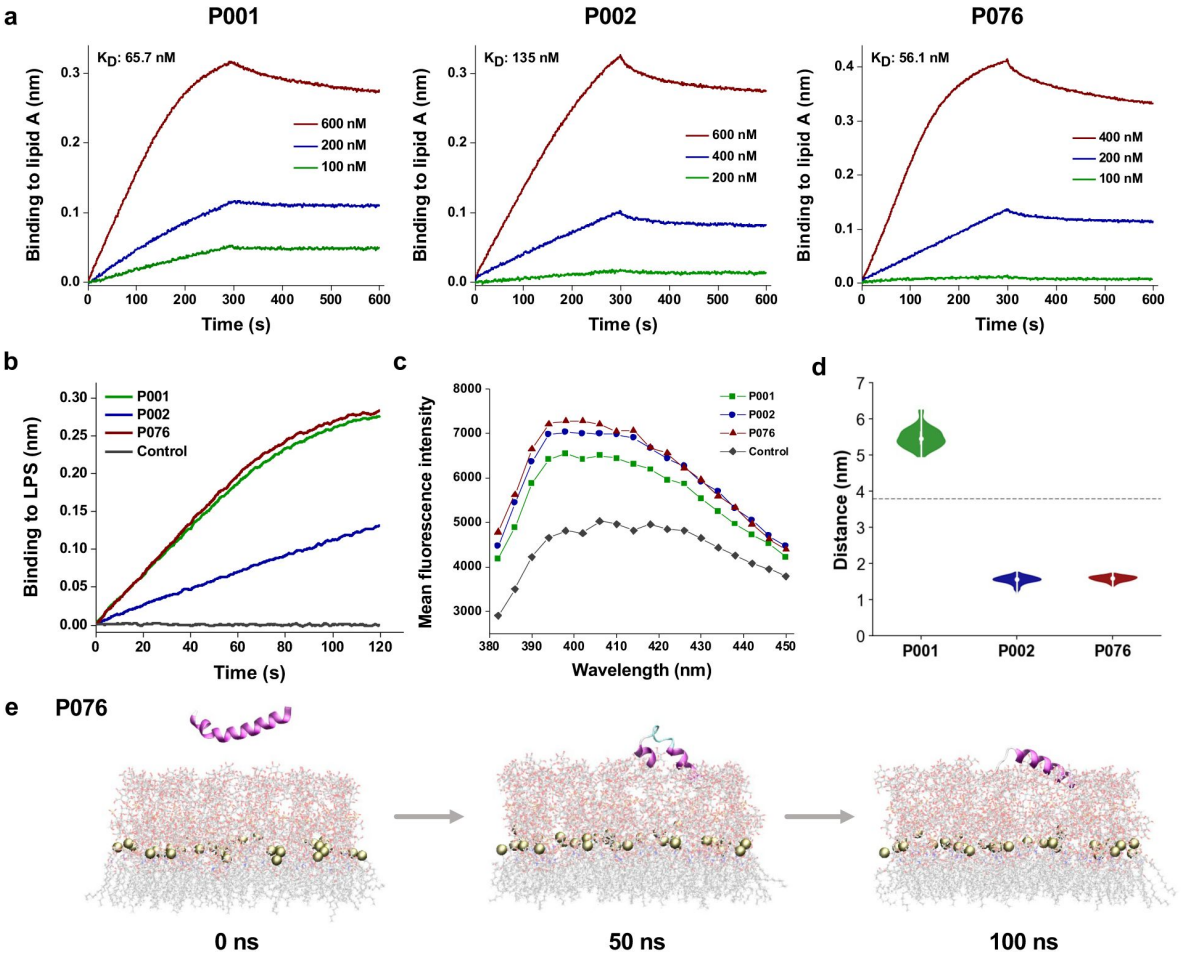


Figure 6.

Bacterial membrane-acting assessment of AMPs.

a. Binding kinetics of three AMPs to lipid A measured by biolayer interferometry.
b. Binding thicknesses of 500 nM AMPs to immobilized LPS within 120 seconds.
c. Fluorescence intensity of NPN after excitation at 350 nm. The maximum emission appeared at approximately 396 nm, where P001, P002, and P076 were detected higher intensity than control.
d. The distance between COM of peptides and phosphorus atoms of lipid A (around the midplane of G-bacterial outer membranes) during the last 50 ns of the MD simulation trajectories. The gray dashed line displays the initial distance at 0 ns.
e. Snapshots of P076 with G-bacterial outer membranes at 0 ns, 50 ns, and 100 ns. Only the top leaflet (LPS) is shown. Tan balls represent the phosphorus atoms of lipid A molecules.

made a tiny change of P076 C-terminal with amide modification but observed loss of functions (Table S7). The change of a few atoms on an intact sequence can lead to dramatic functional consequences, which may be challenging for neural networks to capture.

Recent studies indicated unexpected side effects of traditional antibiotics. Instead of restraining the growth of carbapenem-resistant *Enterobacteriaceae* in the intestine environment, broad-spectrum antibiotics reduced gut microbial populations and depleted beneficial metabolites, thus facilitated colonization of *Enterobacteriaceae* species.[41] With more limitations of conventional small molecule antibiotics discovered, AMPs are attractive alternatives to conventional antibiotic treatment. Their wide bioactivity spectra can expand to virus, thus minimizing the cost of screen multiple drugs. Our results suggested efficient *in silico* screen tools plus rigorous experimental validations could uncover highly potent peptide drugs without comprising on safety, and large-scale animal experiment could facilitate translation towards clinical research. In summary, we believe that a mature workflow involving sequence generation and evaluation can promote the discovery of ideal AMPs to eventually alleviate the hazard of drug-resistant diseases.

Methods

Data Curation

Our AMP dataset is collected from PepVAE,[20] containing 3280 antimicrobial peptide sequences with their minimal inhibitory concentrations (MICs) in logarithmic form. To integrate various AMP sequence and bioactivity data, Witten et al. developed the Giant Repository of AMP Activity (GRAMPA).[42] Since the MIC values change when targeting different microbial strains, only AMPs inhibiting *E. coli* are selected. Also, to simplify the input and reduce accidental effect, peptides without modifications are filtered so that no cysteine appears (amino acid composition in Figure S1a). The sequence length of AMPs are less than 40 amino acids, and the majority of lengths (~ 60%) are between 10 and 20 (Figure S1a). Labels, i.e., logMIC values, are distributed from -1 to 4. During the training process, the dataset is divided into the training, validation, and testing sets in a ratio of 8:1:1 randomly. The entire AMP dataset (3280 sequences) is used for training the generator v1. The AVP dataset for training the generator v2 is collected from AVPdb (accessed in Aug. 2023). AVPs with lengths of more than 60 amino acids are excluded. After removing overlaps with the AMP dataset, 1788 AVPs are kept and added to the training set of generator v2.

Generator

The generative adversarial network (GAN) is composed of a generator network and a discriminator. Through the co-training strategy, the generator evolves to produce ideal sequences from random noises z, fitting the distribution of real samples and faking the discriminator.

$$min_G max_D V(G,D) = E_{x\sim p_{real(x)}}[logD(x)] + E_{z\sim p_{fake(z)}}\left[log\left(1 - D\left(G(z)\right)\right)\right] \quad (1)$$

For the combined training of GAN generator and discriminator, the AMP sequences were encoded with amino acid factors (AAF). AAF encoding represents five physiochemical values, including the polarity, secondary structures, molecular size, codon diversity, and charges.[27] The GAN architecture was utilized to generate novel AMP sequences, where three layers of 2D convolution were inside both the generator and discriminator.[43] Loss function modified with WGAN-GP strategy was taken to avoid mode collapse, and the optimizer was kept as RMSProp. The learning rate was 1e-4. Cross training ratio between generator and discriminator was set as 1:5. Deep learning-based models were trained on an NVIDIA GeForce RTX 2080 Ti GPU. After training for 1000 epochs, string sequences are calculated via cosine similarity. A batch of sequences were generated with the limited length of 30 amino acids. Redundant alanine (A) at the end of every sequence were removed, therefore leaving sequences with variable length.

Predictor

Antimicrobial peptide activity predictor (AMPredictor) was built based on PyTorch and PyTorch Geometric library. All input sequences were treated with padded zeros to reach unified lengths of 60. For chemical fingerprint, RDKit (*www.rdkit.org*) was used to transform each amino acid into 2048-bit Morgan fingerprint in peptide sequences. For each peptide, the fingerprint vectors were concatenated and contrasted into a 60-dimensional vector by average.[44] Pretrained language model esm1b_t33_650M_UR50S was used to provide the embeddings and the contact maps.

During the training process, the optimizer was Adam, and loss function was the mean squared loss. Batch size and learning rate were decided upon grid-search, whose values were 128 and 0.001, respectively (Figure S1b). Evaluation metrics MSE, RMSE, PCC, and CI are calculated as following:

$$MSE = \frac{1}{n}\sum_{1}^{n}(b_i - a_i)^2 \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n}(b_i - a_i)^2} \quad (3)$$

$$PCC = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2 \sum_{i=1}^{n}(b_i - \bar{b})^2}} \quad (4)$$

$$CI = \frac{1}{Z}\sum_{a_i > a_j} h(b_i - b_j) \quad (5)$$

where a_i is the real activity label of peptide i, and b_i is the predicted activity value, Z is the scale constant, and $h(b)$ is the step function.

For comparisons in Figure S2, five baseline models were implemented by DeepPurpose package.[45] The amino acid composition encoding was used as the input of MLP, while one hot encoding was for CNN, CNN-GRU, and CNN-LSTM. Fragment partition fingerprint of peptides was fed into the Transformer encoder.

In ablation study, three encoding approaches (chemical fingerprint, contact matrix, and ESM) and their different combinations were tested individually. It was noted that the peptide graph exists only when both the node embedding and contact matrix are present together. Therefore, a CNN with the same number of layers and output dimension was adopted instead of the GCN in ablation test.

Peptide synthesis

All the peptides were synthesized by DGpeptide Co., ltd. with 95% purity via solid-phase peptide synthesis procedures. Mass spectrometry was used to verify the molecular weights (Figure S4, S12-S35).

MIC detection

The minimum inhibitory concentrations (MICs) of peptides against bacteria including *A. baumannii* (ATCC 19606), multidrug-resistant *A. baumannii* (MDRAB),[46] *S. aureus* (ATCC 25923), methicillin-resistant *S. aureus* (MRSA, ATCC 43300), *E. coli* (ATCC 25922), multidrug-resistant *E. coli* (MDREC),[47] *P. aeruginosa* (PAO1), and multidrug-resistant *P. aeruginosa* (MDRPA),[48] were determined using the broth microdilution method recommended by the Clinical and Laboratory

Standards Institute (CLSI) with modifications reported by the Hancock laboratory.[49] Peptides were prepared in 0.01% acetic acid containing 0.2% BSA at concentrations of 6.25, 12.5, 25, 50, 100, 200, 400, 800, 1600, and 3200 μg/mL. Bacteria grown to the mid-logarithmic phase in Meller-Hinton broth (MHB) were diluted to 2×10^5 CFU/mL. Aliquot of 100 μL of the bacterial suspension was then co-incubated with 11 μL of the peptides at 37°C in a sterile polypropylene microtiter plate (3365, Corning, Shanghai, China). An M2e microplate reader (Molecular Devices, Silicon Valley, CA, USA) was applied to detect the bacterial absorbance at 600 nm after 18 h. The recorded MIC was the lowest concentration of peptides causing a 70% reduction in bacterial turbidity.

Hemolytic assay

The hemolysis of peptides was evaluated as previously described in Ref.[50] Briefly, the fresh mouse blood was centrifuged at 1000 g for 10 min and washed with cold PBS three times. The erythrocytes were collected and diluted in saline solution to prepare a 2% (v/v) suspension. A total of 75 μL of the erythrocyte suspension was co-incubated with 75 μL of the peptides (6.25, 12.5, 25, 50, 100, 200, and 400 μg/mL) at 37°C for 1 h. After centrifuging at 1000 g for 10 min, 60 μL supernatant was obtained, and the absorbance was determined at 414 nm. This experiment was conducted in triplicate and repeated twice.

In vivo antibacterial evaluation

The therapeutic efficacy of P076 peptide against drug-resistant bacteria was analyzed using a mouse peritoneal infection model.[38] The 8-week-old female C57 mice (n = 10) was intraperitoneally administered with 3×10^5 CFU/mL of MDRAB. P076 (2 mg/kg) was given by intraperitoneal injection at 0.5 and 2.5 h after bacterial infection. Ciprofloxacin (CIP, C129896, Aladdin, 32 mg/kg) was employed as the control. Mouse survival was monitored for 48 h. For detection of the bacterial colonization, another 24 mice were randomly divided into four groups (n = 6 for each group). The mice were euthanized 12 h after MDRAB infection, and the peritoneal cavity was rinsed with 2 mL of PBS. Additionally, mouse liver and spleen were obtained by surgery and homogenized in 1 mL of PBS. The lavage fluid and tissue suspension were both diluted for bacterial colony counting using an MHB agar plate as described in MIC detection.

In vivo toxicological evaluation

The 8-week-old female mice (18-22 g) were randomly divided into 14 groups (n = 10 for each group). P076 peptide and polymyxin B (PB, P105490, Aladdin, Shanghai, China) were given to the mice by intraperitoneal injection at 4, 8, 16, 32, 64, and 128 mg/kg. Mouse survival was monitored for 48 h. For pathologic analysis, mouse liver, spleen, and kidneys were obtained by surgery for hematoxylin-eosin (HE) staining after treatment with 64 mg/kg P076 and PB for 12 h. Mice were cared for and treated in accordance with the NIH guidelines for the care and use of laboratory animals (NIH Publication No. 85e23 Rev. 1985). Animal experiments were conducted with approval from the Animal Experimental Ethics Committee of the Army Medical University (AMUWEC20223290).

Cell cultures

Vero-E6 (African green monkey kidney cell line), Huh-7 (Hepatocellular carcinoma cell line), A549 (Human Lung carcinoma cell line) and HeLa-hACE2 were cultured in DMEM (Thermo Fisher Scientific, Waltham, USA) containing 10% fetal bovine serum in a 5% CO_2 incubator at 37°C. HSV-1 and CHIKV were propagated in Vero-E6 cells. DENV-2 and HTNV were propagated in Huh-7 and A549 cells, respectively. And SARS-CoV-2 were propagated in HeLa-hACE2 cells.

Cytotoxicity measurements

The Vero-E6, Huh-7 and A549 cells (2×10^4 cells/well) were inoculated in the 96-well culture plate and grown overnight to 90% confluence at 37°C. The cells were treated with different concentrations of peptides (P001, P002, and P076) for 1 hour, and the same concentration of PBS was used as a control. After 48 hours of cell culture, 10% CCK8 solution was added to each well and incubated at 37°C for 1 hour. The absorbance was measured at 450 nm using a microplate reader. The CC_{50} represents the peptide concentration at the 50% of viability of uninfected cells.

Immunofluorescence assay

The cells were cultured in a 96-well plate (2×10^4 cells/well). Different viruses were mixed with P001, P002, or P076 peptides for 10 min before infecting the cells for 1 h. The infected cells were then cultured with DMEM containing 10% serum for 24-48 h. After discarding the supernatant and washing it three times with PBS, it was permeated with PBS containing 0.2% Triton X-100 (containing 0.1% BSA) at room temperature for 30 minutes. Then, the cells were mixed with an antiviral antibody at 4°C overnight. After washing with PBS three times, the cells were incubated with fluorescence-conjugated secondary antibodies at room temperature for 1 hour. Cell nuclei were stained using 4',6'-diamidino-2-phenylindole (DAPI) in the dark for 5 min, and images were obtained by fluorescence microscope.

Real-time PCR

The cells were cultured in a 6-well plate (2×10^5 cells/well) following the same cell culture procedures as in the immunofluorescence assay. Cellular RNAs were extracted and purified using a viral RNA extraction kit (Qiagen, Dusseldorf, Germany), and then cDNA was synthesized and monitored with SuperScript-III kit (Takara, Dalian, China). qRT–PCR analysis was performed using SYBR Green (BIO-RAD, California, USA) according to the manufacturer's protocol. Viral RNA expression was calculated using the $2^{-\Delta\Delta C.T.}$ (cycle threshold) method normalized to GAPDH expression. The qRT–PCR primers sequences for HTNV, CHIKV, HSV-1 and DENV-2 are listed in the Table S8, while the detection kits (Daan gene, Guangzhou, China) were used for SARS-CoV-2 wild-type and BA.2 subtype.

Transmission electron microscopy

The cells were cultured in a 6-well plate (2×10^5 cells/well) following the same cell culture procedures as in the immunofluorescence assay. The virus-infected cells were collected by centrifugation to prepare cell particles and fixed in cacodylate sodium buffer containing 1% glutaraldehyde (0.2 M, pH 7.2). The fixed sample was dehydrated by acetone solution and embedded in epoxy resin. The sample was polymerized at 60°C for 3 days. The resin block was used to prepare ultrathin slices (50-70 nm thick). These sections were supported by copper mesh, negatively stained with uranyl acetate and lead citrate (electron microscopy) and observed with a JEM100SX transmission electron microscope (JEOL, Tokyo, Japan).

Biolayer interferometry

The Octet Red96 BLI platform (Sartorius BioAnalytical Instruments, Bohemia, USA) was utilized to measure the binding kinetics of three peptides to lipid A (L5399, Sigma, Shanghai, China). Peptides were diluted with 75 mM NaCl solution at concentrations of 100, 200, 400, and 600 nM. The amine reactive second-generation (AR2G) biosensors were processed with 1-ehtyl-3-(3-dimethylaminopropy) carbodiimide hydrochloride (EDC, E1769, Sigma) and sulfo-N-hydroxysulfosuccinimide (s-NHS, 56485, Sigma). Subsequently, lipid A was loaded onto the activated AR2G biosensors at 30°C for 5 min, and both the processes of association and disassociation were monitored for 5 min to measure the kinetic constants. The equilibrium dissociation constant (K_D) was defined as the division of dissociation rate constant (K_{off}) and the

association rate constant (K_{on}). In case of LPS, AR2G biosensors immobilized with 15 μg/mL LPS (L2880, Sigma) were incubated with 500 nM peptides for 2 min in 75 mM NaCl solution. The binding thickness was recorded and analyzed using ForteBio Data Analysis 7.0 software.

NPN uptake assay

MDRAB in mid-logarithmic phase was diluted to 1×10^8 CFU/mL in 10 mM sodium phosphate buffer (pH 7.4). Then 10 μL of each peptide was mixed with 190 μL of MDRAB at 37°C for 1 hour. Afterwards, 1-N-phenylnaphthylamine (NPN, 104043, Sigma, Shanghai, China) was added with a final concentration of 10 μM. The fluorescence intensity was assessed with the wavelength ranging from 380 to 450 nm at 2 nm intervals, after excitation at 350 nm using a Tecan Infinite M1000 Pro microplate reader (Mannedorf, Zurich, Switzerland). Each experiment was conducted in duplicate and repeated three times.

Molecular dynamics simulations

The initial structures of peptides were built via AlphaFold2 single sequence version.[51] Lipid bilayer models were constructed by CHARMM-GUI server along with placing parallel peptides approximately 2 nm above the membrane surface.[52] Nine systems were built for three peptides versus three kinds of lipid bilayers (viral envelopes, Gram-negative bacterial inner membranes and outer membranes).[39, 53] Ions of 150 mM NaCl were added to neutralize the system, and the calcium ions were used for LPS in outer leaflet of bacterial outer membranes. Detailed system settings are recorded in Table S9. The standard CHARMM-GUI equilibration protocols including NVT and NPT ensembles were adopted, keeping the temperature at 310K. All-atom MD simulations were performed using GROMACS 2021.2 with the CHARMM36 force field and the TIP3P water model for 100 ns, with a time step of 2 fs.[54, 55] Three independent run was performed for each system. The visual molecular dynamics (VMD) software version 1.9.4 was used for visualization.

The distance between the center of mass (COM) of peptides and phosphorus atoms in top leaflets was calculated via *gmx distance*. Binding free energy was calculated using the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method on average of extracted 20 frames from the last 20 ns. The binding free energy of a peptide-membrane complex is calculated as below:

$$\Delta G_{bind} = G_{complex} - (G_{peptide} + G_{membrane}) \quad (6)$$

$$G_{bind} = G_{polar} + G_{nonpolar} = \left(E_{elec} + G_{psolv}\right) + (E_{vdw} + G_{sasa}) \quad (7)$$

where E_{elec} is the Coulomb potential, E_{vdw} is the van der Waals energy, G_{psolv} and G_{sasa} are polar and nonpolar solvation free energy. Following previous study, the dielectric constant was set as 7 for phospholipids and the entropy term was neglected.[56]

Data Availability

All the python codes of the AMP design model are available at *https://github.com/ruihan-dong/GAN-for-AMP-Design* and *https://github.com/ruihan-dong/AMPredictor*. An online version of AMPredictor can be accessed via *https://huggingface.co/spaces/ruihan-dong/AMPredictor*.

Acknowledgements

We thank Prof. Chen Song from Peking University for helpful discussion in MD. We also thank Qiaozhen Meng, Yuqian Pu and Nan Song from Tianjin University. This work was supported by the National Key Research and Development Program of China (grant no. 2020YFA0908500), the National Natural Science Foundation of China (grant no. 22007071, 31971127, 81971563, 82272330, 32100122, 62322215), Shanghai Sailing Program (21YF1457800), and Independent Research Project of State Key Laboratory of Trauma and Chemical Poisoning (SKLYQ202101).

Additional information

Author Contributions

C.Z. and R.D. devised the project. C.Z., S.Y., and H.L. conceived the main conceptual ideas and were in charge of overall direction and planning. R.D. and F.G. developed the computational model. J.W., C.W., and G.Z. measured antimicrobial activities and performed animal experiments. X.W., R.L., Z.L., S.H., X.M., H.K., and J.L. carried out antiviral evaluations. Y.L. and P.Z. carried out anti-SARS-CoV-2 evaluations. C.Z. and R.D. analyzed the data and wrote the paper with input from all authors.

Additional files

Supporting information

Supplemental spreadsheet

References

1. Lewies A., Du Plessis L. H., Wentzel J. F. 2019) **Antimicrobial Peptides: the Achilles' Heel of Antibiotic Resistance?** *Probiotics Antimicrob Proteins* **11**:370–381

2. Antimicrobial Resistance C. 2022) **Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis** *Lancet* **399**:629–655

3. De Oliveira D. M. P., Forde B. M., Kidd T. J., Harris P. N. A., Schembri M. A., Beatson S. A., Paterson D. L., Walker M. J. 2020) **Antimicrobial Resistance in ESKAPE Pathogens** *Clin Microbiol Rev* **33**

4. Hutchings M. I., Truman A. W., Wilkinson B. 2019) **Antibiotics: past, present and future** *Curr Opin Microbiol* **51**:72–80

5. Huan Y., Kong Q., Mou H., Yi H. 2020) **Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields** *Front Microbiol* **11**:582779

6. Mahlapuu M., Hakansson J., Ringstad L., Bjorn C. 2016) **Antimicrobial Peptides: An Emerging Category of Therapeutic Agents** *Front Cell Infect Microbiol* **6**:194

7. Spohn R., Daruka L., Lazar V., Martins A., Vidovics F., Grezal G., Mehi O., Kintses B., Szamel M., Jangir P. K., Csorgo B., Gyorkei A., Bodi Z., Farago A., Bodai L., Foldesi I., Kata D., Maroti G., Pap B., Wirth R., Papp B., Pal C. 2019) **Integrated evolutionary analysis reveals antimicrobial peptides with limited resistance** *Nat Commun* **10**:4538

8. Huang J., Xu Y., Xue Y., Huang Y., Li X., Chen X., Xu Y., Zhang D., Zhang P., Zhao J., Ji J. 2023) **Identification of potent antimicrobial peptides via a machine-learning pipeline that mines the entire space of peptide sequences** *Nat Biomed Eng* **7**:797–810

9. Wang G., Li X., Wang Z. 2016) **APD3: the antimicrobial peptide database as a tool for research and education** *Nucleic Acids Res* **44**:D1087–1093

10. Pirtskhalava M., Amstrong A. A., Grigolava M., Chubinidze M., Alimbarashvili E., Vishnepolsky B., Gabrielian A., Rosenthal A., Hurt D. E., Tartakovsky M. 2021) **DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics** *Nucleic Acids Res* **49**:D288–D297

11. Shi G., Kang X., Dong F., Liu Y., Zhu N., Hu Y., Xu H., Lao X., Zheng H. 2022) **DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides** *Nucleic Acids Res* **50**:D488–D496

12. Xu J., Li F., Leier A., Xiang D., Shen H. H., Marquez Lago T. T., Li J., Yu D. J., Song J. 2021) **Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides** *Brief Bioinform* **22**

13. Yan J., Bhadra P., Li A., Sethiya P., Qin L., Tai H. K., Wong K. H., Siu S. W. I. 2020) **Deep-AmPEP30: Improve Short Antimicrobial Peptides Prediction with Deep Learning** *Mol Ther Nucleic Acids* **20**:882–894

14. Li C., Sutherland D., Hammond S. A., Yang C., Taho F., Bergman L., Houston S., Warren R. L., Wong T., Hoang L. M. N., Cameron C. E., Helbing C. C., Birol I. 2022) **AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens** *BMC Genomics* **23**

15. Pang Y., Yao L., Xu J., Wang Z., Lee T. Y. 2022) **Integrating transformer and imbalanced multi-label learning to identify antimicrobial peptides and their functional activities** *Bioinformatics* **38**:5368–5374

16. Yan K., Lv H., Guo Y., Peng W., Liu B. 2023) **sAMPpred-GAT: prediction of antimicrobial peptide by graph attention network and predicted peptide structure** *Bioinformatics* **39**

17. Ma Y., Guo Z., Xia B., Zhang Y., Liu X., Yu Y., Tang N., Tong X., Wang M., Ye X., Feng J., Chen Y., Wang J. 2022) **Identification of antimicrobial peptides from the human gut microbiome using deep learning** *Nat Biotechnol* **40**:921–931

18. Maasch J., Torres M. D. T., Melo M. C. R., de la Fuente-Nunez C. 2023) **Molecular de-extinction of ancient antimicrobial peptides enabled by machine learning** *Cell Host Microbe* **31**:1260–1274

19. Torres M. D. T., Melo M. C. R., Flowers L., Crescenzi O., Notomista E., de la Fuente-Nunez C. 2022) **Mining for encrypted peptide antibiotics in the human proteome** *Nat Biomed Eng* **6**:67–75

20. Dean S. N., Alvarez J. A. E., Zabetakis D., Walper S. A., Malanoski A. P. 2021) **PepVAE: Variational Autoencoder Framework for Antimicrobial Peptide Generation and Activity Prediction** *Front Microbiol* **12**:725727

21. Das P., Sercu T., Wadhawan K., Padhi I., Gehrmann S., Cipcigan F., Chenthamarakshan V., Strobelt H., Dos Santos C., Chen P. Y., Yang Y. Y., Tan J. P. K., Hedrick J., Crain J., Mojsilovic A. 2021) **Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations** *Nat Biomed Eng* **5**:613–623

22. Szymczak P., Mozejko M., Grzegorzek T., Jurczak R., Bauer M., Neubauer D., Sikora K., Michalski M., Sroka J., Setny P., Kamysz W., Szczurek E. 2023) **Discovering highly potent antimicrobial peptides with deep generative model HydrAMP** *Nat Commun* **14**:1453

23. Tucs A., Tran D. P., Yumoto A., Ito Y., Uzawa T., Tsuda K. 2020) **Generating Ampicillin-Level Antimicrobial Peptides with Activity-Aware Generative Adversarial Networks** *ACS Omega* **5**:22847–22851

24. Van Oort C. M., Ferrell J. B., Remington J. M., Wshah S., Li J. 2021) **AMPGAN v2: Machine Learning-Guided Design of Antimicrobial Peptides** *J Chem Inf Model* **61**:2198–2207

25. Pandi A., Adam D., Zare A., Trinh V. T., Schaefer S. L., Burt M., Klabunde B., Bobkova E., Kushwaha M., Foroughijabbari Y., Braun P., Spahn C., Preusser C., Pogge von Strandmann E., Bode H. B., von Buttlar H., Bertrams W., Jung A. L., Abendroth F., Schmeck B., Hummer G., Vazquez O., Erb T. J. 2023) **Cell-free biosynthesis combined with deep learning accelerates de novo-development of antimicrobial peptides** *Nat Commun* **14**:7197

26. Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. 2014) **Generative Adversarial Nets** *Adv Neur In* **27**:2672–2680

27. Atchley W. R., Zhao J., Fernandes A. D., Druke T. 2005) **Solving the protein sequence metric problem** *Proc Natl Acad Sci U S A* **102**:6395–6400

28. Muller A. T., Gabernet G., Hiss J. A., Schneider G. 2017) **modlAMP: Python for antimicrobial peptides** *Bioinformatics* **33**:2753–2755

29. Rives A., Meier J., Sercu T., Goyal S., Lin Z., Liu J., Guo D., Ott M., Zitnick C. L., Ma J., Fergus R. 2021) **Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences** *Proc Natl Acad Sci U S A* **118**

30. Qureshi A., Tandon H., Kumar M. 2015) **AVP-IC50 Pred: Multiple machine learning techniques-based prediction of peptide antiviral activity in terms of half maximal inhibitory concentration (IC50)** *Biopolymers* **104**:753–763

31. Thakur N., Qureshi A., Kumar M. 2012) **AVPpred: collection and prediction of highly effective antiviral peptides** *Nucleic Acids Res* **40**:W199–204

32. Li J., Pu Y., Tang J., Zou Q., Guo F. 2020) **DeepAVP: A Dual-Channel Deep Neural Network for Identifying Variable-Length Antiviral Peptides** *IEEE J Biomed Health Inform* **24**:3012–3019

33. Sharma R., Shrivastava S., Singh S. K., Kumar A., Singh A. K., Saxena S. 2022) **Deep-AVPpred: Artificial Intelligence Driven Discovery of Peptide Drugs for Viral Infections** *IEEE J Biomed Health Inform* **26**:5067–5074

34. Timmons P. B., Hewage C. M. 2021) **ENNAVIA is a novel method which employs neural networks for antiviral and anti-coronavirus activity prediction for therapeutic peptides** *Brief Bioinform* **22**

35. Lin T. T., Sun Y. Y., Wang C. T., Cheng W. C., Lu I. H., Lin C. Y., Chen S. H. 2022) **AI4AVP: an antiviral peptides predictor in deep learning approach with generative adversarial network data augmentation** *Bioinform Adv* **2**:vbac080

36. Timmons P. B., Hewage C. M. 2020) **HAPPENN is a novel tool for hemolytic activity prediction for therapeutic peptides which employs neural networks** *Sci Rep-Uk* **10**

37. Chung J. H., Bhat A., Kim C. J., Yong D., Ryu C. M. 2016) **Combination therapy with polymyxin B and netropsin against clinical isolates of multidrug-resistant Acinetobacter baumannii** *Sci Rep* **6**:28168

38. Zhao G., Jia C., Zhu C., Fang M., Li C., Chen Y., He Y., Han S., He Y., Gao J., Wang T., Wang C., Wang J. 2022) **, gamma-Core Guided Antibiotic Design Based on Human Enteric Defensin 5** *Membranes (Basel)* **13**

39. Liu R., Liu Z., Peng H., Lv Y., Feng Y., Kang J., Lu N., Ma R., Hou S., Sun W., Ying Q., Wang F., Gao Q., Zhao P., Zhu C., Wang Y., Wu X. 2022) **Bomidin: An Optimized Antimicrobial Peptide With Broad Antiviral Activity Against Enveloped Viruses** *Front Immunol* **13**:851642

40. Kitchen D. B., Decornez H., Furr J. R., Bajorath J. 2004) **Docking and scoring in virtual screening for drug discovery: methods and applications** *Nat Rev Drug Discov* **3**:935–949

41. Yip A. Y. G., King O. G., Omelchenko O., Kurkimat S., Horrocks V., Mostyn P., Danckert N., Ghani R., Satta G., Jauneikaite E., Davies F. J., Clarke T. B., Mullish B. H., Marchesi J. R., McDonald J. A. K. 2023) **Antibiotics promote intestinal growth of carbapenem-resistant Enterobacteriaceae by enriching nutrients and depleting microbial metabolites** *Nat Commun* **14**:5094

42. Jacob W., Zack W. 2019) **Deep learning regression model for antimicrobial peptide design** *bioRxiv* :692681 https://doi.org/10.1101/692681

43. Tzu-Tang L., Li-Yen Y., Ching-Tien W., Ga-Wen L., Chi-Fong K., Yang-Hsin S., Shu-Hwa C., Chung-Yen L. 2021) **Discovering Novel Antimicrobial Peptides in Generative Adversarial Network** *bioRxiv* https://doi.org/10.1101/2021.11.22.469634

44. Schissel C. K., Mohapatra S., Wolfe J. M., Fadzen C. M., Bellovoda K., Wu C. L., Wood J. A., Malmberg A. B., Loas A., Gómez-Bombarelli R., Pentelute B. L. 2021) **Deep learning to design nuclear-targeting abiotic miniproteins** *Nat Chem* **13**:992

45. Huang K. X., Fu T. F., Glass L. M., Zitnik M., Xiao C., Sun J. M. 2020) **DeepPurpose: a deep learning library for drug-target interaction prediction** *Bioinformatics* **36**:5545–5547

46. Wang C., Zhao G., Wang S., Chen Y., Gong Y., Chen S., Xu Y., Hu M., Wang X., Zeng H., Wang A., Liu D., Su Y., Cheng T., Chen F., Wang J. 2018) **A Simplified Derivative of Human Defensin 5 with Potent and Efficient Activity against Multidrug-Resistant Acinetobacter baumannii** *Antimicrob Agents Chemother* **62**

47. Zhao G., Chen Y., He Y., Chen F., Gong Y., Chen S., Xu Y., Su Y., Wang C., Wang J. 2019) **Succinylated casein-coated peptide-mesoporous silica nanoparticles as an antibiotic against intestinal bacterial infection** *Biomater Sci* **7**:2440–2451

48. Gao C., Wang Y., Zhang Y., Wei J., Cheng X., Zhang J., Zou Q., Gu J. 2020) **Complete Genome Sequence of Pseudomonas aeruginosa XN-1, Isolated from the Sputum of a Severe Pneumonia Patient** *Microbiol Resour Announc* **9**

49. Wu M., Hancock R. E. 1999) **Interaction of the cyclic antimicrobial cationic peptide bactenecin with the outer and cytoplasmic membrane** *J Biol Chem* **274**:29–35

50. Oddo A., Hansen P. R. 2017) **Hemolytic Activity of Antimicrobial Peptides** *Methods Mol Biol* **1548**:427–435

51. Mirdita M., Schutze K., Moriwaki Y., Heo L., Ovchinnikov S., Steinegger M. 2022) **ColabFold: making protein folding accessible to all** *Nat Methods* **19**:679–682

52. Wu E. L., Cheng X., Jo S., Rui H., Song K. C., Davila-Contreras E. M., Qi Y., Lee J., Monje-Galvan V., Venable R. M., Klauda J. B., Im W. 2014) **CHARMM-GUI Membrane Builder toward realistic biological membrane simulations** *J Comput Chem* **35**:1997–2004

53. Allsopp R., Pavlova A., Cline T., Salyapongse A. M., Gillilan R. E., Di Y. P., Deslouches B., Klauda J. B., Gumbart J. C., Tristram-Nagle S. 2022) **Antimicrobial Peptide Mechanism Studied by Scattering-Guided Molecular Dynamics Simulation** *J Phys Chem B* **126**:6922–6935

54. Best R. B., Zhu X., Shim J., Lopes P. E., Mittal J., Feig M., Mackerell A. D. 2012) **Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles** *J Chem Theory Comput* **8**:3257–3273

55. Klauda J. B., Venable R. M., Freites J. A., O'Connor J. W., Tobias D. J., Mondragon-Ramirez C., Vorobyov I., MacKerell A. D., Pastor R. W. 2010) **Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types** *J Phys Chem B* **114**:7830–7843

56. Lee J., Jung S. W., Cho A. E. 2016) **Molecular Insights into the Adsorption Mechanism of Human β-Defensin-3 on Bacterial Membranes** *Langmuir* **32**:1782–1790

Author information

Ruihan Dong[#]
Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin Key Laboratory of Function and Application of Biological Macromolecular Structures, School of Life Sciences, Faculty of Medicine, Tianjin University, Tianjin, China, Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China
ORCID iD: 0009-0001-5862-8410

[#]These authors contributed equally to this work

Rongrong Liu[#]
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

[#]These authors contributed equally to this work

Ziyu Liu[#]
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

[#]These authors contributed equally to this work

Yangang Liu[#]
Department of Microbiology, Second Military Medical University, Shanghai, China

[#]These authors contributed equally to this work

Gaomei Zhao
State Key Laboratory of Trauma and Chemical Poisoning, Institute of Combined Injury of PLA, College of Preventive Medicine, Third Military Medical University (Army Medical University), Chongqing, China

Honglei Li
Tianjin Cancer Hospital Airport Hospital, Tianjin, China

Shiyuan Hou
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

Xiaohan Ma
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

Huarui Kang
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

Jing Liu
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

Fei Guo
School of Computer Science and Engineering, Central South University, Changsha, China

Ping Zhao
Department of Microbiology, Second Military Medical University, Shanghai, China

Junping Wang
State Key Laboratory of Trauma and Chemical Poisoning, Institute of Combined Injury of PLA, College of Preventive Medicine, Third Military Medical University (Army Medical University), Chongqing, China

Cheng Wang
State Key Laboratory of Trauma and Chemical Poisoning, Institute of Combined Injury of PLA, College of Preventive Medicine, Third Military Medical University (Army Medical University), Chongqing, China

For correspondence: wangctmmu@126.com

Xingan Wu
Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an, China

For correspondence: wuxingan@fmmu.edu.cn

Sheng Ye
Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin Key Laboratory of Function and Application of Biological Macromolecular Structures, School of Life Sciences, Faculty of Medicine, Tianjin University, Tianjin, China

For correspondence: sye@tju.edu.cn

Cheng Zhu
Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin Key Laboratory of Function and Application of Biological Macromolecular Structures, School of Life Sciences, Faculty of Medicine, Tianjin University, Tianjin, China
ORCID iD: 0000-0003-0260-6287

For correspondence: cheng_zhu@tju.edu.cn

Editors

Reviewing Editor
Anne-Florence Bitbol
Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland

Senior Editor
Aleksandra Walczak
École Normale Supérieure - PSL, Paris, France

Reviewer #1 (Public review):

This manuscript presents a pipeline incorporating a deep generative model and peptide property predictors for the de novo design of peptide sequences with dual antimicrobial/antiviral functions. The authors synthesized and experimentally validated three peptides designed by the pipeline, demonstrating antimicrobial and antiviral activities, with one leading peptide exhibiting antimicrobial efficacy in animal models.

Overall, the authors have addressed each major comment through new experiments, particularly by validating 24 peptides, clarifying alignment methods, and demonstrating sequence novelty. These additions have strengthened the manuscript. To further refine the work, it would be helpful to briefly describe any steps taken to mitigate GAN pathologies (such as mode collapse), provide a short rationale for the use of five AVP classifiers and how they complement each other, and clearly present the expanded experimental data (including MIC values and antiviral results) in the main text. Finally, the authors should also compare their approach with recently described deep-learning-enabled antibiotic discovery methods.

https://doi.org/10.7554/eLife.97330.2.sa3

Reviewer #2 (Public review):

Summary:

This study marks a noteworthy advance in the targeted design of AMPs, leveraging a pioneering deep learning framework to generate potent bifunctional peptides with specificity against both bacteria and viruses. The introduction of a GAN for generation and a GCN-based AMPredictor for MIC predictions is methodologically robust and a major stride in computational biology. Experimental validation in vitro and in animal models, notably with the highly potent P076 against a multidrug-resistant bacterium and P002's broad-spectrum viral inhibition, underpins the strength of their evidence. The findings are significant, showcasing not just promising therapeutic candidates, but also demonstrating a replicable means to rapidly develop new antimicrobials against the threat of drug-resistant pathogens.

Strengths:

The de novo AMP design framework combines a generative adversarial network (GAN) with an AMP predictor (AMPredictor), which is a novel approach in the field. The integration of deep generative models and graph-encoding activity regressors for discovering bifunctional AMPs is cutting-edge and addresses the need for new antimicrobial agents against drug-resistant pathogens. The in vitro and in vivo experimental validations of the AMPs provide strong evidence to support the computational predictions. The successful inhibition of a spectrum of pathogens in vitro and in animal models gives credibility to the claims. The discovery of effective peptides, such as P076, which demonstrates potent bactericidal activity against multidrug-resistant A. baumannii with low cytotoxicity, is noteworthy. This could

have far-reaching implications for addressing antibiotic resistance. The demonstrated activity of the peptides against both bacterial and viral pathogens suggests that the discovered AMPs have a wide therapeutic potential and could be effective against a range of pathogens.

Comments on revisions: I have no further comments on revisions.

https://doi.org/10.7554/eLife.97330.2.sa2

Reviewer #3 (Public review):

Summary:

Dong et al. described a deep learning-based framework of antimicrobial (AMP) generator and regressor to design and rank de novo antimicrobial peptides (AMPs). For generated AMPs, they predicted their minimum inhibitory concentration (MIC) using a model that combines the Morgan fingerprint, contact map and ESM language model. For their selected AMPs based on predicted MIC, they also use a combination of antiviral peptide (AVP) prediction models to select AMPs with potential antiviral activity. They experimentally validated 3 candidates for antimicrobial activity against S. aureus, A. baumannii, E. coli, and P. aeruginosa, and their toxicity on mouse blood and three human cell lines. The authors select their most promising AMP (P076) for in vivo experiments in A. baumannii-infected mice. They finally test the antiviral activity of their 3 AMPs against viruses.

Strengths:

- The development of de novo antimicrobial peptides (AMPs) with the novelty of being bifunctional (antimicrobial and antiviral activity).

- Novel, combined approach to AMP activity prediction from their amino acid sequence.

Weaknesses:

- I missed the justification for combined antiviral and antibacterial activities. As the authors responded, less than 10% of the training data has antiviral activity. Therefore, I do not understand how the high percentage of antiviral activities was achieved. Especially reading that the antiviral filtering did not have an influence on the number of antiviral peptides obtained.

- I had difficulty in reading the story because of the use of acronyms without referring to their full name for the first time, and incomplete information annotation in figures and captions.

https://doi.org/10.7554/eLife.97330.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *This manuscript presents a pipeline incorporating a deep generative model and peptide property predictors for the de novo design of peptide sequences with dual antimicrobial/antiviral functions. The authors synthesized and experimentally validated*

> *three peptides designed by the pipeline, demonstrating antimicrobial and antiviral activities, with one leading peptide exhibiting antimicrobial efficacy in animal models. However, the manuscript as it stands, has several major limitations on the computational side.*

Thanks for your comments.

> *Major issues:*
>
> *(1) The choice of GAN as the generative model. There are multiple deep generative frameworks (e.g., language models, VAEs, and diffusion models), and GANs are known for their training difficulty and mode collapse. Could the authors elaborate on the specific rationale behind choosing GANs for this task?*

We thank the reviewer for his/her concern on GAN models. We agree that there are some limitations of GAN itself such as its training difficulty, but we cannot deny its potential in generating biological sequences, especially in AMP generation. GAN and VAE are the two most commonly used generative models in the field of AMP design (Curr Opin Struct Biol 2023, 83:102733). AMPGAN (J Chem Inf Model, 2021, 61, 2198-2207.), Multi-CGAN (J Chem Inf Model 2024, 64, 1, 316–326), PepGAN (ACS Omega, 2020, 5, 22847-22851) and others have verified its application ability on peptide design. Moreover, PandoraGAN (Sn Comput Sci 2023, 4, 607) is one of the few works on AVP generation which is also based on GAN architecture. GAN updates the generator weights on the backpropagation from the discriminator directly rather than manually defined complicated loss function, which alleviates the reliance on input data. Our current results demonstrated that the trained GAN generator could produce novel sequences that featured high antimicrobial activity, both validated *in silico* and *in vitro*.

> *(2) The pipeline is supposed to generate peptides showing dual properties. Why were antiviral peptides not used to train the GAN? Would adding antiviral peptides into the training lead to a higher chance of getting antiviral generations?*

A major mechanism of antimicrobial peptides is to disrupt cell membranes. Thus, some antimicrobial peptides are reported with broad-spectrum antibacterial and antiviral activities, since the virus shares a membrane structure with bacteria, especially the enveloped viruses. In APD3 database, 244 of 3940 AMPs are labeled with antiviral activities. In constrast, most reported antiviral peptides inhibit the viruses by binding to specific targets (proteins and nucleic acids) related to viral proliferation so that they may not have antibacterial effects. Therefore, we trained the GAN with the AMP dataset. We chose this AMP dataset mainly for AMPredictor (with detailed logMIC label against *E.coli*) and then used the same dataset to train a GAN for simplification.

In the revised manuscript, we also tested adding available antiviral peptides from AVPdb to train the GAN model. The number of AVPs is 1,788 after removing overlaps with used AMP dataset. The GAN architecture and hyperparameters remain the same. After generating a batch of sequences with this trained generator, we scored them by AMPredictor and filtered them with five AVP classifiers. As expected, the predicted MIC values shifted to higher performance with 17 sequences < 5 μM and 39 sequences < 10 uM, and previous numbers are 26 and 42 in the manuscript. Among 39 sequences < 10 μM, 13 passed all five AVP classifiers and 17 passed four (33.3% and 43.6%, respectively). Previous ratios are 40.5% and 35.7% (17 and 15 out of 42). Two generators perform roughly the same for generating AVPs (76.9% vs. 76.1%) as evaluated by our rules (4 or more positives), but the generator trained solely with AMPs provided more AVPs with higher possibility (5 positives).

We also experimentally tested dozens of generated peptides from two versions of generators (v1 for training solely on AMPs, v2 for training with AVPs, Figure 2 in revised manuscript). The 'antiviral' feature of a peptide was checked when significant inhibition was observed in immunofluorescence assays against HSV-1 at the concentration of 10 μM. Six and seven antiviral peptides were found out of 12 tested peptides from generators v1 and v2, respectively. Therefore, the success rates for two versions of generators are about 60% (including three reported peptides in the original manuscript) and show no significant difference.

> *(3) For the antimicrobial peptide predictor, where were the contact maps of peptides sourced from?*

The contact maps of AMPs were predicted from ESM, which were obtained at the same time when obtaining the ESM embeddings (Methods section, Page 24, Line 538: Pretrained language model esm1b_t33_650M_UR50S was used to provide the embeddings and the contact maps.)

> *(4) Morgan fingerprint can be used to generate amino acid features. Would it be better to concatenate ESM features with amino acid-level fingerprints and use them as node features of GNN?*

We thank the reviewer for this suggestion. We test using ESM and fingerprint (FP) features on graph nodes and the result is shown in Author response table 1. AMPredictor (ESM on nodes, FP after GNN) still performed slightly better than concatenating FP on node features on four regression metrics.

Author response table 1.

Results of AMPredictor with fingerprint on nodes

Model	RMSE	MSE	Pearson	CI
AMPredictor (ESM on nodes, FP after GNN)	**0.5348**	**0.2860**	**0.7072**	**0.7294**
ESM + FP on nodes	0.5356	0.2869	0.7012	0.7265

> *(5) Although the number of labeled antiviral peptides may be limited, the input features (ESM embeddings) should be predictive enough when coupled with shallow neural networks. Have the authors tried simple GNNs on antiviral prediction and compared the prediction performance to those of existing tools?*

We thank the reviewer for his/her suggestion on AVP predictions. We haven't tried it. An important reason is that we focused on developing regressors instead of binary classifiers. Currently available AVP data with numerical labels did not support training a reliable regressor, for their limited amount as well as heterogenous virus target and experimental assay. Therefore, we decided to use reported AVP classifiers as an additional filter following AMPredictor. Since only using one classifier may lead to bias, we chose five AVP classifiers as ensemble votes.

> *(6) Instead of using global alignment to get match scores, the authors should use local alignment.*

We calculated the match scores by global alignment methods referred to AMPGAN v2 (J Chem Inf Model 2021, 61, 2198−2207), CLaSS (Nat Biomed Eng 2021 5, 613–623), and AMPTrans-lstm (Comput Struct Biotechnol J 2022, 21, 463-471), to check the similarity between the generated sequences and any sequences in the training set. In addition, we also used local alignment to check the novelty of peptides (regarding the next question).

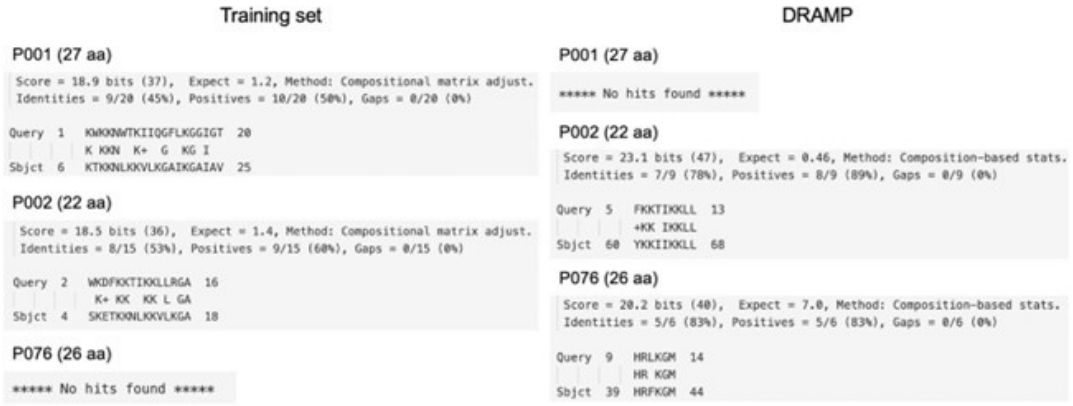
> *(7) How novel are the validated peptides? The authors should run a sequence alignment to get the most similar known AMP for each validated peptide, and analyze whether they are similar.*

We have listed the most similar AMP segments to our generated peptides from the training set and DRAMP database (28,233 sequences after filtering out those containing irregular characters). BLAST parameters were set as CLaSS (Nat Biomed Eng 2021 5, 613–623) for short peptides. The lowest Evalue of P001 aligned with the training set is 1.2, and no hits were found for P001 with DRAMP. Two E-values of P002 are 1.4 and 0.46. P076 had no hits in the training set and got a high E-value of 7.0 with DRAMP. Detailed alignments are shown below. This result indicates that our three validated AMPs are novel.

Since we generated more sequences using two versions of generator for validation, we also checked the BLAST E-value of these validated peptides. The results are listed in Table S3. All sequences obtained E-values > 0.1 and some of them had no hits when aligned with the training set or the DRAMP database.

Author response image 1.

Alignments of three validated peptides.



> *(8) Only three peptides were synthesized and experimentally validated. This is too few and unacceptable in this field currently. The standard is to synthesize and characterize several dozens of peptides at the very least to have a robust study.*

We thank the reviewer for the suggestion and promoted our models to generate >10 times more peptides in the revised manuscript. We have synthesized and tested more peptides *in vitro* and added these results in the revised manuscript (Figure 2). From two versions of generators (trained with or without AVPs), we selected 24 peptides in total for antibacterial and antiviral validations. All 24 peptides showed antibacterial activity towards at least bacterial strain, and 13 peptides were screened out through the quick antiviral test. This result indicates the capability of our design method for bifunctional AMPs with a notable success rate (60%).

Reviewer #2 (Public Review):

Summary:

This study marks a noteworthy advance in the targeted design of AMPs, leveraging a pioneering deeplearning framework to generate potent bifunctional peptides with specificity against both bacteria and viruses. The introduction of a GAN for generation and a GCN-based AMPredictor for MIC predictions is methodologically robust and a major stride in computational biology. Experimental validation in vitro and in animal models, notably with the highly potent P076 against a multidrug-resistant bacterium and P002's broad-spectrum viral inhibition, underpins the strength of their evidence. The findings are significant, showcasing not just promising therapeutic candidates, but also demonstrating a replicable means to rapidly develop new antimicrobials against the threat of drug-resistant pathogens.

Strengths:

The de novo AMP design framework combines a generative adversarial network (GAN) with an AMP predictor (AMPredictor), which is a novel approach in the field. The integration of deep generative models and graph-encoding activity regressors for discovering bifunctional AMPs is cutting-edge and addresses the need for new antimicrobial agents against drug-resistant pathogens. The in vitro and in vivo experimental validations of the AMPs provide strong evidence to support the computational predictions. The successful inhibition of a spectrum of pathogens in vitro and in animal models gives credibility to the claims. The discovery of effective peptides, such as P076, which demonstrates potent bactericidal activity against multidrug-resistant A. baumannii with low cytotoxicity, is noteworthy. This could have far-reaching implications for addressing antibiotic resistance. The demonstrated activity of the peptides against both bacterial and viral pathogens suggests that the discovered AMPs have a wide therapeutic potential and could be effective against a range of pathogens.

We thank the reviewer for the comments.

Reviewer #3 (Public Review):

Summary:

Dong et al. described a deep learning-based framework of antimicrobial (AMP) generator and regressor to design and rank de novo antimicrobial peptides (AMPs). For generated AMPs, they predicted their minimum inhibitory concentration (MIC) using a model that combines the Morgan fingerprint, contact map, and ESM language model. For their selected AMPs based on predicted MIC, they also use a combination of antiviral peptide (AVP) prediction models to select AMPs with potential antiviral activity. They experimentally validated 3 candidates for antimicrobial activity against S. aureus, A. baumannii, E. coli, and P. aeruginosa, and their toxicity on mouse blood and three human cell lines. The authors select their most promising AMP (P076) for in vivo experiments in A. baumannii-infected mice. They finally test the antiviral activity of their 3 AMPs against viruses.

Strengths:

-The development of de novo antimicrobial peptides (AMPs) with the novelty of being bifunctional (antimicrobial and antiviral activity).

-Novel, combined approach to AMP activity prediction from their amino acid sequence.

> *Weaknesses:*
>
> *(1) I missed justification on why training AMPs without information of their antiviral activity would generate AMPs that could also have antiviral activity with such high frequency (32 out of 104).*

Thanks for your inquiry. A major mechanism of antimicrobial peptides is to disrupt cell membranes. Thus, some antimicrobial peptides are reported with broad-spectrum antibacterial and antiviral activities, since the virus shares a membrane structure with bacteria, especially the enveloped viruses. In APD3 database, 244 of 3940 AMPs are labeled with antiviral activities. However, several reported antiviral peptides inhibit the viruses by binding to specific targets (proteins and nucleic acids) related to viral proliferation so that they may not have antibacterial effects. Therefore, we trained the GAN with the AMP dataset. We chose this AMP dataset mainly for AMPredictor (with detailed logMIC label against *E.coli*) and then used the same dataset to train a GAN for simplification. In addition, it's not 32 antiviral candidates out of 104 but 32 out of 42 peptides with predicted MIC < 10 μM because we did the filtering process stepwise.

In revision, we also tested adding available antiviral peptides from AVPdb to train the GAN model (generator v2). The number of AVPs is 1,788 after removing overlaps with used AMP dataset. The GAN architecture and hyperparameters remain the same. We used generator v2 to obtain a batch of sequences and screened out bifunctional candidates following the same procedure. 30 out of 39 peptides with predicted MIC < 10 μM passed four or five AVP predictors. Therefore, two generators perform roughly the same for generating AVP candidates (76.9% vs. 76.1%).

> *(2) The justification for AMP predictor advantages over previous tools lacks rationale, comparison with previous tools (e.g., with the very successful AMP prediction approach described by Ma et al. 10.1038/s41587-022-01226-0), and proper referencing.*

Thanks for your suggestion. Ma et al. proposed ensemble binary classification models to mine AMPs from metagenomes successfully. However, we concentrated on the development of regression models. As a regressor, AMPredictor predicts the specific logMIC value of the input sequences instead of giving a yes/no answer. Since the training settings and evaluation metrics are different for the classification and regression tasks, we could not compare AMPredictor with Ma et al. directly. Instead, we compared the performance of AMPredictor with some regression baseline models (Figure S2a) and our model outperformed them.

> *(3) Experimental validation of three de novo AMPs is a very low number compared to recent similar studies.*

Thanks for pointing out this shortcoming. We have synthesized and tested more peptides *in vitro* and added these results in the revised manuscript (Figure 2). From two versions of generators (trained with or without AVPs), we selected 24 peptides in total for antibacterial and antiviral validations. All 24 peptides showed antibacterial activity towards at least bacterial strain, and 13 peptides were screened out through the quick antiviral test. This result indicates the capability of our design method for bifunctional AMPs with a notable success rate (60%).

> *(4) I have concerns regarding the in vivo experiments including i) the short period of reported survival compared to recent studies (0.1038/s41587-022-01226-0, 10.1016/j.chom.2023.07.001, 0.1038/s41551-022-00991-2) and ii) although in Figure 2 f and g statistics have been provided, log scale y-axis would provide a better comparative representation of different conditions.*

Thank you for your suggestions.

i) In current study, we monitored the survival of mice with peritoneal bacterial infection for 48 h.

Because abdominal bacterial infection can induce severe sepsis and cause mouse death within 40 h (Sci Adv 2019, 5(7), eaax1946), the 48 h is sufficient to evaluate the therapeutic efficacy of antimicrobial peptides (Nat Biotechnol 2019, 37(10), 1186-1197).

ii) In Figure 2f and 2g (3f and 3g in the revised manuscript), the y-axis has already been in log-scale and tick labels are marked in scientific notation.

> *(5) I had difficulty reading the story because of the use of acronyms without referring to their full name for the first time, and incomplete annotation in figures and captions.*

Thank you for pointing this. We have checked the manuscript carefully and modified the figure captions during revision.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *(1) To validate the generalizability of the model, it would be prudent to include data on AMPs targeting a broader range of bacteria and viruses. This could help ensure that the peptides designed are not narrowly focused on E. coli but are effective against a more extensive set of pathogens.*

Thanks for your suggestions. We just incorporated AMPs with *E. coli* activity labels since it is the most common strain among available AMP databases. As for a regressive model (AMPredictor), the fitting object should be defined concisely, which means limited targeting bacteria. Some other articles also focused on *E. coli* labels as well (Nat Commun 2023, 14, 7197; mSystems 2023, 8, e0034523).

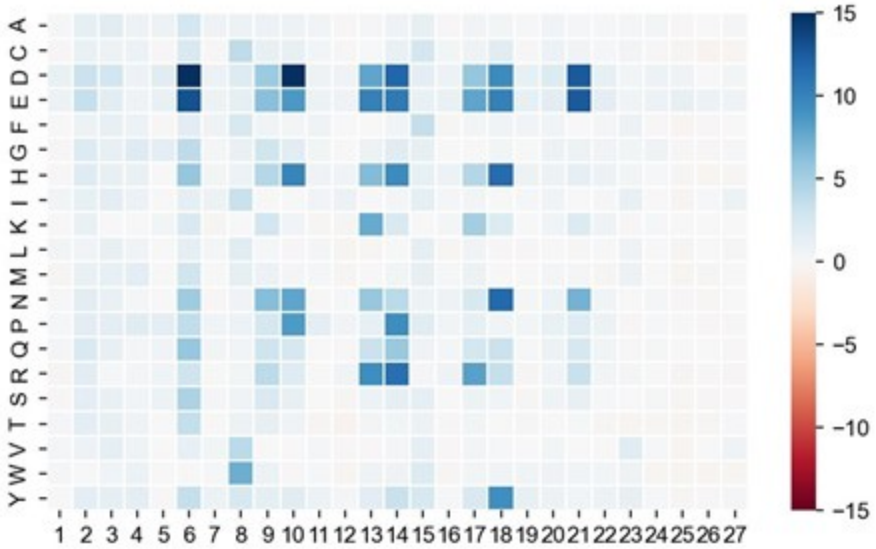
We used the same processed dataset to train the GAN generator for simplification. Most reported AMPs have the potential to target various microbes. We have counted the antimicrobial labels of these peptides in our dataset, shown in Figure S1b. In addition to *E. coli*, some of the peptides target Grampositive *S. aureus*, fungus *C. albicans*, and other bacterial species as well. Our experimental validation also reveals the wide spectrum of designed peptides inhibiting Gram-negative, Gram-positive, drugresistant bacteria, and enveloped viruses. With the expansion of well-curated AMP databases, we expect to update the model with larger scale datasets in the near future.

> *(2) Conduct sensitivity analyses to understand how minor changes in the peptide sequences impact the model's predictions. This will reduce the chances of overlooking potential AMP candidates due to the model's inability to capture subtle changes.*

Thank you for this valuable suggestion. We kept similar known peptide sequences in the training sets regarding that a single mutation may have an impact on their antimicrobial performances. We took P001 as an example to perform the sensitivity analysis by site saturation mutagenesis *in silico*. Author response image 2 represents the change of antimicrobial activity scores as predicted by AMPredictor. Since the predicted MIC of P001 is 0.949 μM (experimentally measured value is 0.80 μM), most single mutations lead to higher scores (i.e., worse performance), especially Asp (D) and Glu (E) residues with negative charges. The largest change value of single amino acid replacement is 25.51 (W6D). Although this value may not reflect the actual changes, it is enough to be distinguished when screening and ranking candidate sequences.

Author response image 2.

Site saturated mutagenesis of P001. Color shows the change of predicted MIC against *E. coli* as predicted by AMPredictor (lower score is better).



> *(3) Given the relatively short length of the peptides, typically ranging from 10 to 20 residues, the authors might consider employing a fully-connected graph in the peptide's graphical representation. This approach could potentially simplify the model without sacrificing the descriptive power due to the limited size of the peptides.*

Thanks for your suggestions. We tested fully-connected graph edge encodings and the results on the test set were shown in Author response table 2 below. We found that AMPredictor with peptide contact map still performed better on Pearson correlation coefficient and CI, while using fully-connected graphs reached a slightly improved RMSE and MSE. Nonetheless, using fully-connected graph demands about 10time memory and more computational costs when processing more complicated message-passing. Therefore, the involvement of structural information is still a preferred choice.

Author response table 2.

Results of AMPredictor with different graph edge encodings

Graph edge encoding	RMSE	MSE	Pearson	CI
Peptide Contact Map	0.5348	0.2860	**0.7072**	**0.7294**
Fully-connected	**0.5319**	**0.2828**	0.6946	0.7250

> *(4) Upon reviewing Table S1, it is apparent that the application of ESM embeddings alone achieves commendable prediction accuracy. It would be intriguing to investigate whether the adoption of the more recent ESM models-specifically the second-generation ESM2*

> *t36_3B, t48_15B, and t33_650Mcould enhance model performance beyond that observed using the esm1b_t33_650M_UR50S model described in the manuscript.*

Thanks for your suggestions. Here, we included various ESM2 models' outputs as our node features and presented the results in Author response table 3. Notably, the dimensions of esm2_t36_3B and esm2_t48_15B are 2560 and 5120, respectively, while both esm2_t33_650M and esm1b_t33_650M are 1280 dimensions.

Interestingly, we found that larger models don't lead to improved performance. ESM-1b version still holds the best metrics in RMSE, MSE, and Pearson correlation coefficient. This indicates that the choice of pretrained model versions depended on specific downstream tasks.

Author response table 3.

Results of AMPredictor with different ESM versions

ESM embedding	RMSE	MSE	Pearson	CI
esm1b_t33_650M_UR50S	**0.5348**	**0.2860**	**0.7072**	0.7294
esm2_t33_650M_UR50D	0.5425	0.2943	0.6944	**0.7563**
esm2_t36_3B_UR50D	0.5581	0.3115	0.6774	0.7494
esm2_t48_15B_UR50D	0.5622	0.3160	0.6503	0.7093

> *(5) It may be pertinent to reevaluate the use of the MM-PBSA approach within the scope of this study. Typically, MM-PBSA is utilized to estimate the free energy differences between the bound and unbound states of solvated molecules. The application of MM-PBSA is to calculate binding energies between proteins and membranes is unconventional and infrequently documented in the literature. Therefore, it is recommended that the authors consider omitting this portion of the manuscript, or provide a robust justification for its inclusion and application in this context.*

Thanks for your comments on MM/PBSA methods. There have been several literatures using this approach to calculate peptide-membrane binding free energy (Langmuir 2016, 32, 1782-1790; J Cell Biochem 2018, 119, 9205-9216; J Chem Inf Model 2019, 59, 3262-3276; Molecular Therapy Oncolytics 2019, 16, 7-19; Microbiology Spectrum 2023, 11, e0320622; J Chem Inf Model 2023, 63, 5823-5833) and we referred to their settings, such as the dielectric constant. All of these works built similar all-atom systems including cationic antimicrobial peptides and membrane bilayers, and utilized MM/PBSA method to describe the absorption process of the peptide from an unbound initial state. The order of magnitude of our calculation results is consistent with other reported works. Additionally, computational results may provide supporting evidence and we discussed that this quantitative energy calculation should be considered along with other observed metrics.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *The weaknesses I mentioned in the Public Review may be addressed by improving the writing and presentation and corrections to the text and figures.*

Thanks for your suggestion. We have carefully checked and improved the presentation of text and figures in the revised manuscript.

https://doi.org/10.7554/eLife.97330.2.sa0