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

Ruihan Dong[1,2#], Rongrong Liu[3#], Ziyu Liu[3#], Yangang Liu[4#], Gaomei Zhao[5], Honglei Li[6], Shiyuan Hou[3], Xiaohan Ma[3], Huarui Kang[3], Jing Liu[3], Fei Guo[7], Ping Zhao[4], Junping Wang[5], Cheng Wang[5,*], Xingan Wu[3,*], Sheng Ye[1,*], Cheng Zhu[1,*]

[1] Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin Key Laboratory of Function and Application of Biological Macromolecular Structures, School of Life Sciences, Tianjin University, Tianjin 300072, China

[2] Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[3] Department of Microbiology, School of Basic Medicine, Fourth Military Medical University, Xi'an 710000, Shaanxi, China

[4] Department of Microbiology, Second Military Medical University, Shanghai 200433, China

[5] 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 400038, China

[6] Tianjin Cancer Hospital Airport Hospital, Tianjin 300308, China

[7] School of Computer Science and Engineering, Central South University, Changsha 410083, Hunan, China

[#] These authors contributed equally to this work.

[*] Correspondence: wangctmmu@126.com (C.W.), wuxingan@fmmu.edu.cn (X.W.), sye@tju.edu.cn (S.Y.), cheng_zhu@tju.edu.cn (C.Z.)

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 three 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.

Keywords

Antimicrobial peptides, Antiviral peptides, Deep learning, Dual-function, Generative model, Sequence design

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 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

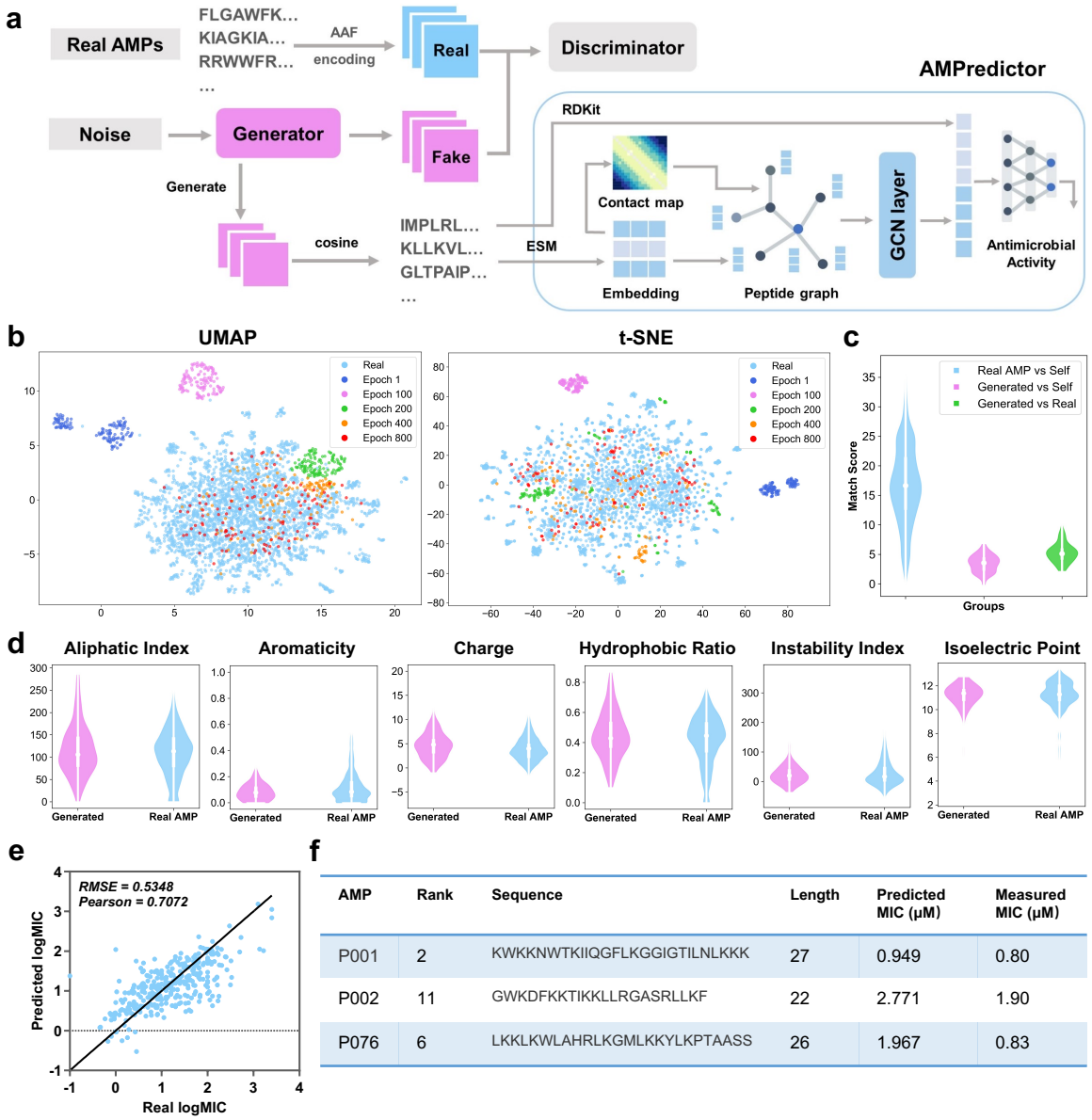
The overall architecture for designing functional peptides fuses a generative adversarial network and an antimicrobial activity predictor (AMPredictor, Figure 1a). We aim to explore beyond the known antimicrobial templates or accessible proteomics space. We trained the GAN module with 3 280 AMP sequences 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 1 000 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 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 AMPredictior 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 S2a). Hence, a well-trained generator could produce a variety of AMP candidates and then AMPredcitor 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.



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 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.

191 f. The sequences of three selected peptides as well as their predicted and experimentally validated
192 MIC values against *E. coli* ATCC 25922.
193
194
195 **Designing bifunctional antimicrobial peptides *in silico***
196
197 We acquired 104 novel peptides via the *de novo* design framework. Among them, 42 and 26 were
198 predicted to own MIC values below 10 μM and 5 μM, respectively (supplemental spread sheet).
199 To establish a sequential workflow and filter out antimicrobial peptides with antiviral capabilities,
200 we then considered assembling several AVP classifiers subsequent to the AMPredictor. As the
201 source data, the AVPdb provided ~2 000 sequences with IC_{50} labels.[30] However, data curation
202 resulted in only 759 unique AVPs. We determined that the quantitative data of AVPs were
203 insufficient for training another GCN-based regression model. Instead, binary classification should
204 exhibit robustness against noise of a relatively small dataset. Hence, we merely made yes/no
205 decisions using a collective of five published AVP classification tools (AVPpred,[31] DeepAVP,[32]
206 Deep-AVPpred,[33] ENNAVIA,[34] and AI4AVP,[35] details in Table S2). We regarded a peptide as an
207 antiviral candidate if four or more positive votes were recorded and 32 peptides met our criteria.
208
209 We conducted a comprehensive evaluation of efficacy and safety, to select novel AMPs before
210 experimental validations. Considering the membrane corruption mechanism of most AMPs, we
211 predicted their hemolytic activity as a quick measure of cellular toxicity.[36] Through the ranking of
212 high activity and low toxicity, we picked three potent peptides P001, P002, and P076 (Figure 1f),
213 all of which featured amphiphilic signatures of sequence and enriched cationic regions (e.g.,
214 helical wheel projections in Figure S3), for *in vitro* and *in vivo* experimental assessment.
215
216
217 ***In vitro* antibacterial and hemolytic assays**
218
219 We examined the antibacterial activity of P001, P002, and P076 against one Gram-positive (*S.*
220 *aureus*) and three Gram-negative (*A. baumannii*, *E. coli*, and *P. aeruginosa*) bacteria, including
221 three ESKAPE strains. For each species, we cultured a standard strain and a multidrug-resistant

222 (MDR) strain, and measured the MIC values of individual peptides using the broth microdilution
223 method. All the peptides exhibited inspiring bactericidal properties with MICs ranging from 0.20
224 to 15.18 μM (corresponding to 0.625 to 40 μg/mL, Figure 2a). P001 showed broad-spectrum
225 antibacterial capability with MICs less than 4 μM across all the eight strains of four bacteria species.
226 P002 featured lower MIC values in inhibiting Gram-positive bacteria *S. aureus*, while P076 mainly
227 inhibits Gram-negative bacteria *A. baumannii*, *E. coli*, and *P. aeruginosa*. Notably, the
228 experimentally measured MICs against *E. coli* agreed well with the predictions of AMPredictor
229 (Figure 1f), indicating the accuracy of our regression models.
230
231 The hemolytic analysis revealed that both P001 and P002 displayed dose-dependent toxicity to
232 mouse erythrocytes (Figure 2b), with average hemolysis higher than 50% at high concentrations
233 (20 μM, or 100 μg/mL). Nevertheless, the hemolysis of P076 was lower than 1% at concentrations
234 up to 70 μM. We also inspected the apparent cytotoxicity (CC_{50}) of three peptides on three
235 mammalian cell lines (Vero-E6, A549, Huh-7) using CCK8 assays. For the cell cultures, CC_{50}
236 values of all three peptides were well above 30 μM (Figure S8a). The CC_{50} of P076, specifically,
237 were in the range of 35.5 μM (for A549 cells from lung tissues) to 66.8 μM (for Huh-7 cells from
238 liver tissues). The *in vitro* safety assays pointed out thresholds of concentration for administrating
239 these peptides, which were adapted in subsequent animal experiments for estimation of safe dosage.
240
241 Our data suggest that all three peptides are potential candidates for inhibiting the clinically relevant
242 *A. baumannii*, which is a top-ranking pathogen with critical priority for R&D of new antibiotics,
243 as defined by WHO.[3] The MIC of P076 at eliminating the persistent MDRAB reached 0.21 μM, a
244 value half of P001 and P002 (0.40 and 0.47 μM, respectively). Meanwhile, P076 was comparable
245 with P001 but more efficient than P002 against *E. coli* and *P. aeruginosa*. Despite a slight
246 inferiority in killing *S. aureus*, P076 owned a better biosafety. Taken together, the sequence of
247 P076 likely leveraged the balance between efficacy and safety, we therefore focused the following
248 *in vivo* evaluation and animal infection models on P076.
249
250

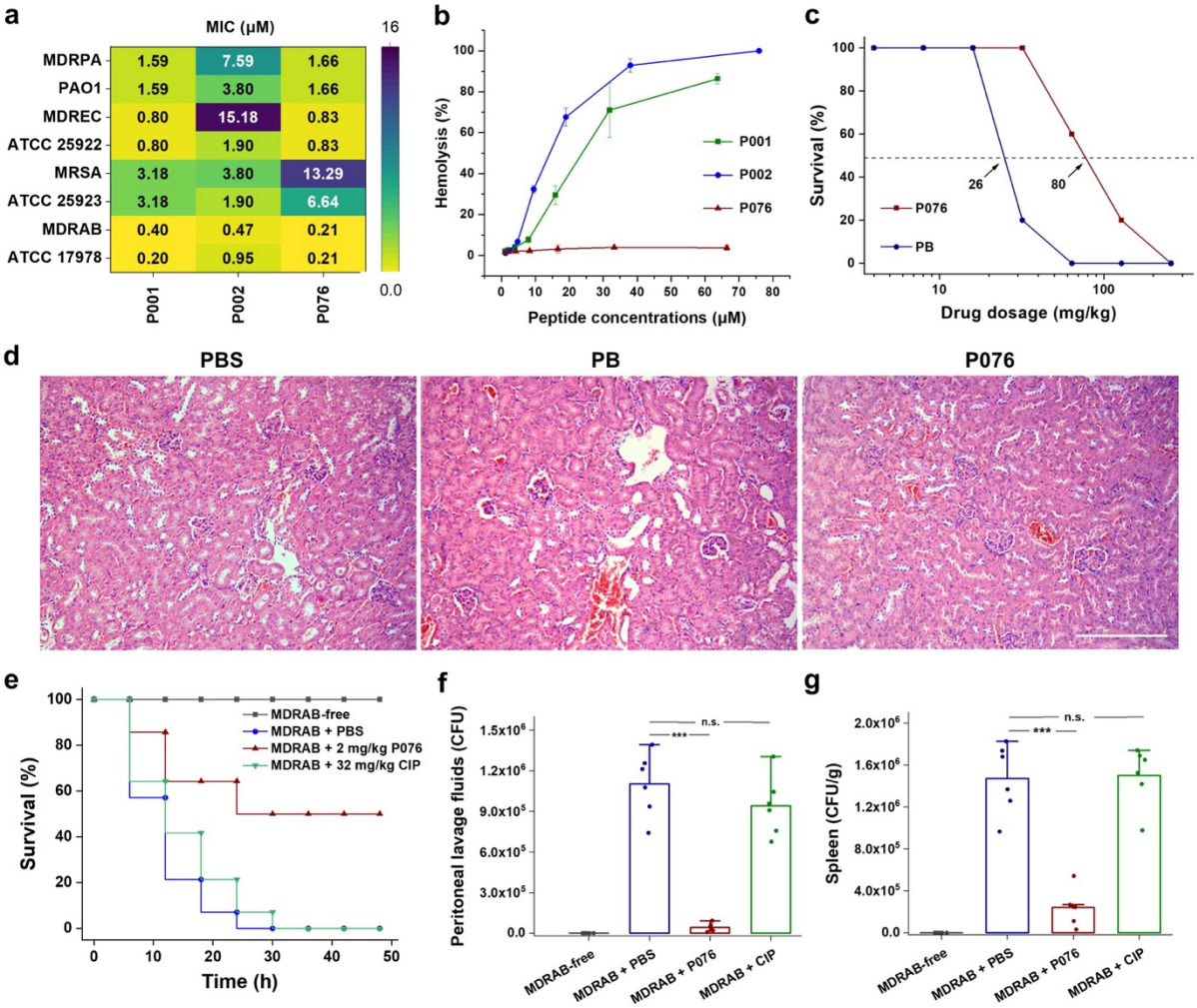


Figure 2. Antibacterial and toxicological evaluations of P076 peptide.

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.

264 ***In vivo* antibacterial and toxicological evaluations**

265

266 We first determined a safe dosage for administrating P076 on mouse models. A total of 140 C57
267 mice (randomly divided into 14 groups) were given increasing concentrations of P076 through
268 intraperitoneal injection, and the cyclic peptide antibiotic polymyxin B (PB) was employed as a
269 positive control. In spite of nephrotoxicity, PB is one of the polymyxin antibiotics that is
270 considered "the last ditch" for the treatment of infections caused by extensively drug-resistant
271 Gram-negative bacteria in clinic.[37] Animal survival was monitored for 48 h after a single drug
272 administration. The P076 concentration to induce a 50% death was calculated to be 80 mg/kg,
273 approximately 3-fold higher than that of PB (26 mg/kg, Figure 2c). Hematoxylin-eosin (HE)
274 staining further verified that no pathological changes were observed in the mice liver, spleen and
275 kidneys of mice given 32 mg/kg P076 after 12 h, whereas an apparent renal tubular injury was
276 discovered in PB-treated mouse (Figure 2c and S5a).

277

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

290

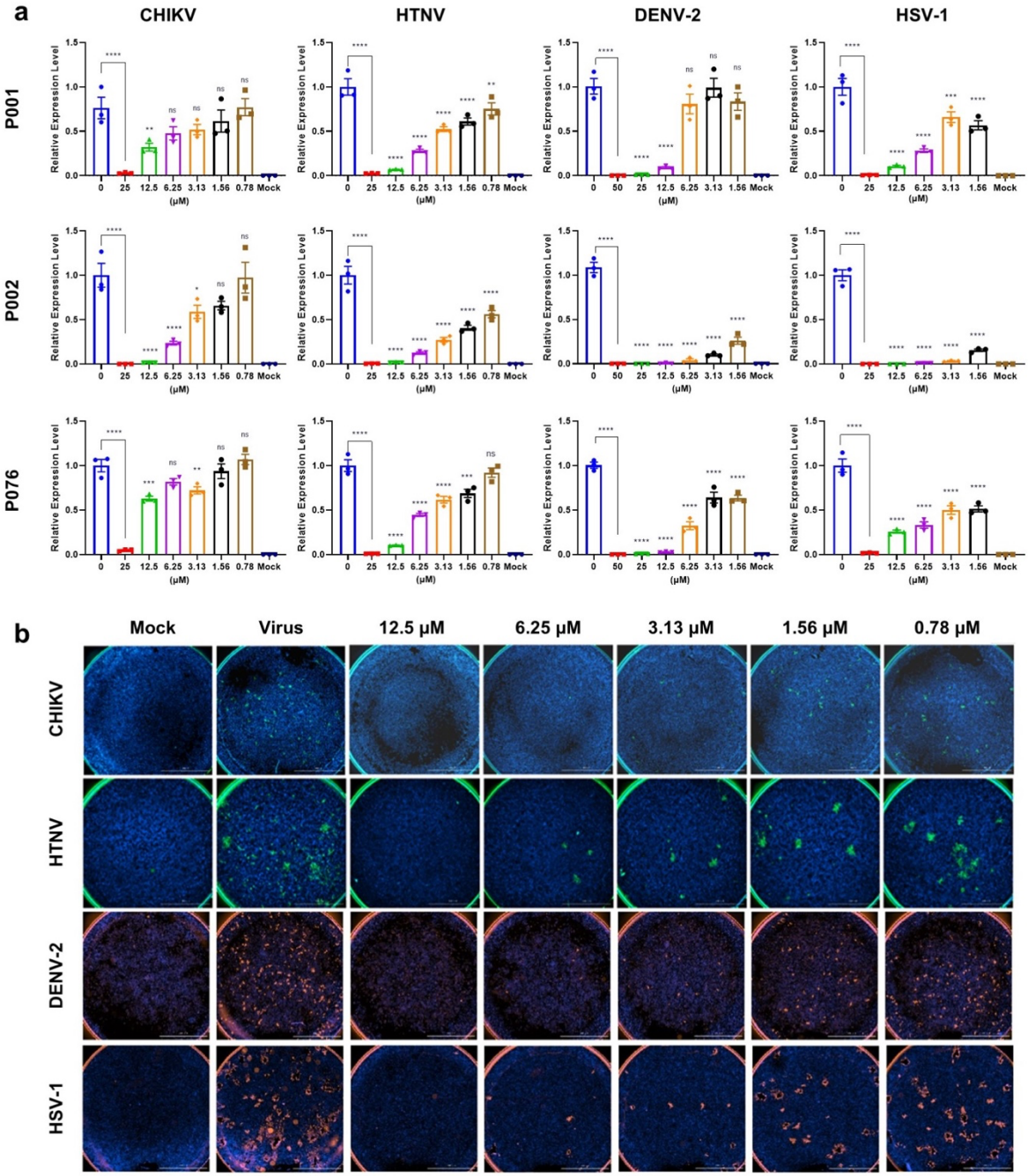


Figure 3. Antiviral assays of AMPs against CHIKV, HTNV, DENV-2, and HSV-1.

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.

298 **Broad-spectrum antiviral assays**

299

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

310

311 We incubated the peptides with different viruses before they invaded the cultured cells and self-
312 amplified. We analyzed the potential antiviral effects of P001, P002, P076 at the mRNA and protein
313 levels, respectively. At the mRNA level, the quantitative real-time PCR suggested dose-dependent
314 inhibition of viral-specific gene expressions (E1 gene for CHIKV, S gene for HTNV, VP16 gene
315 for HSV-1, NS5a gene for DENV-2, ORF for SARS-CoV-2, Figure 3a and Figure S6b). At the
316 protein level, we applied fluorescence-conjugated antibodies to visualize trace amounts of virus
317 populations. Corroborating with the RT-PCR results, P002 nearly abolished all viral infections at
318 12.5 μM (Figure 3b, Figure S6a and Figure S7). To quantify the immunofluorescence results, we
319 plotted the log-scale fluorescent focus forming unit (FFU) against varying peptide concentrations
320 (Figure S8b). Among the three peptides, P002 exhibited the highest potential with EC_{50} of 0.37
321 μM (against CHIKV) to 2.08 μM (against HTNV) (Figure 3b). The EC_{50} of P076 and P001 were
322 in the range of 1.62 to 2.67 μM (except for inhibiting DENV-2).

323

324 Combining the EC_{50} and CC_{50} calculations, we defined a selectivity index ($SI = CC_{50}/ EC_{50}$) as the
325 experimental guides of safety versus efficacy (Table S3). The SI of P002 against four viruses are
326 149.21, 31.75, 51.05, and 59.01, three of which outperformed the SI of P001 and P076, especially
327 for the case of HTNV. Hence, we further observed the performance of P002 with transmission
328 electron microscopy. For the control set, the viruses were allowed to propagate in cells for days

329 and identified as assembled particles in cytosol (Figure 4a, b). When the viruses were mixed with
330 10 μM P002 before addition to the cells, the viral particles became absent while the cell
331 morphology kept intact, indicating the broad-spectrum inhibition effects of P002 (Figure 4c, d).
332
333 Because we observed a gradient of antimicrobial and antiviral activities from P001, P002, and
334 P076 with their individual preferences against bacteria and viruses, we aimed to probe the
335 interactions between peptides and different phospholipid assemblies, thus deriving feasible means
336 judging their underlying preferences. We performed all-atom molecular dynamics simulations with
337 molecular models of bacterial membranes and viral envelopes, placed each peptide adjacent to the
338 phospholipid bilayers, and simulated them attaching to the membrane (Figure S9a, b). For the
339 model of Gram-negative bacteria systems, the MM/PBSA calculations and ΔG_{bind} distinguished
340 P002 from P076 and P001 (Table S8). Evaluated by ΔG_{bind} (-465.79 ± 11.39 kJ/mol), P076 would
341 associate with the membrane of Gram-negative bacteria tighter than others. For the model of viral
342 envelopes, all binding energies were weaker than that to bacteria. The ranking of $\Delta G_{nonpolar}$
343 corroborated with the antiviral potential of P002. Combining this energy term with a geometrical
344 parameter (average distance between the COM of peptide and phosphorus atoms in the top leaflet
345 on z-axis) may predict the behavior of P002 better than $\Delta G_{nonpolar}$ alone (Figure S9c).
346

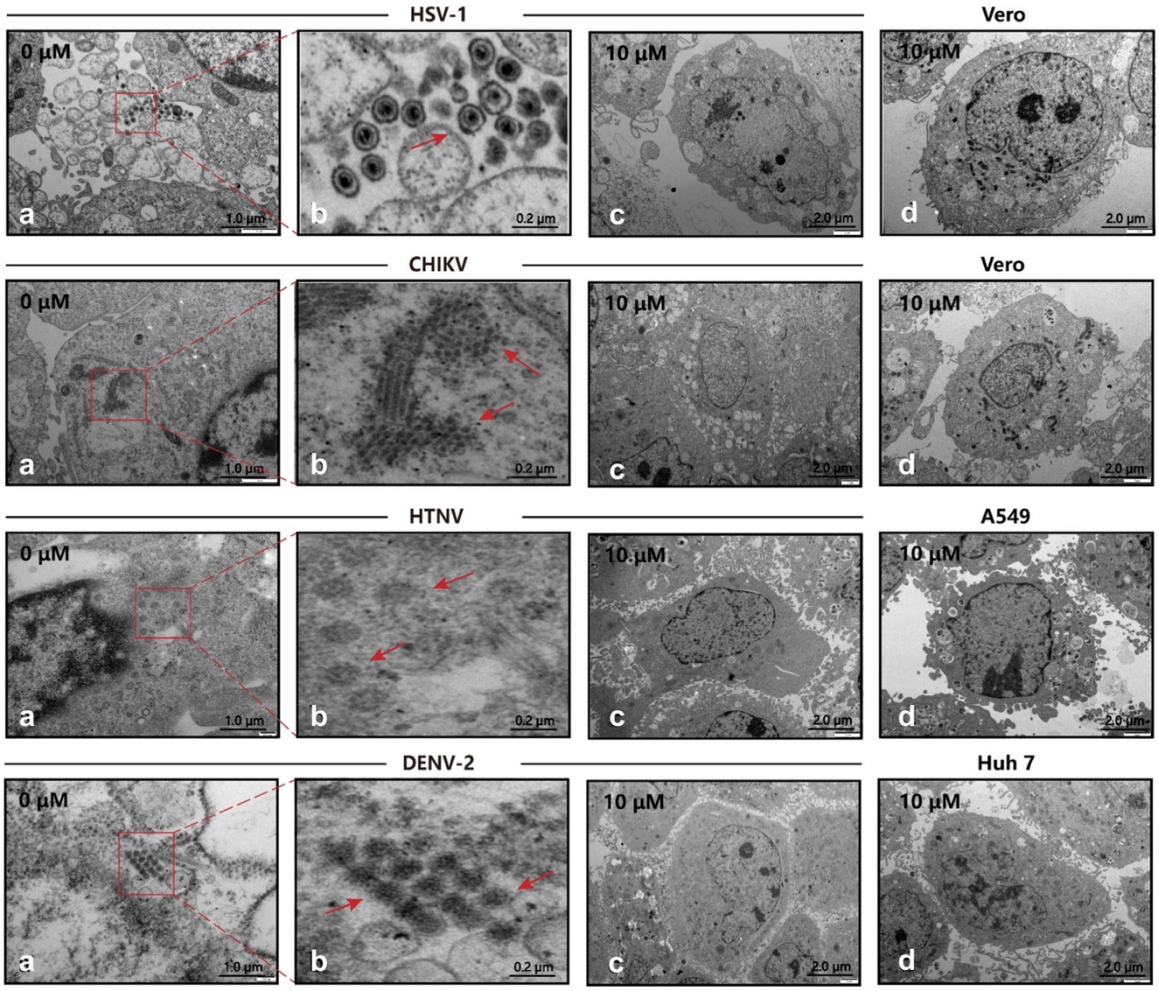


Figure 4. Transmission electron microscopy (TEM) of viral infection with P002.

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.

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 5b). 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 5a). Then we directly investigated the membrane permeability by 1-N-phenylnaphthylamine (NPN) uptake 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 5c). 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 5e). 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 5d). 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 5c. 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 S9b). 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.

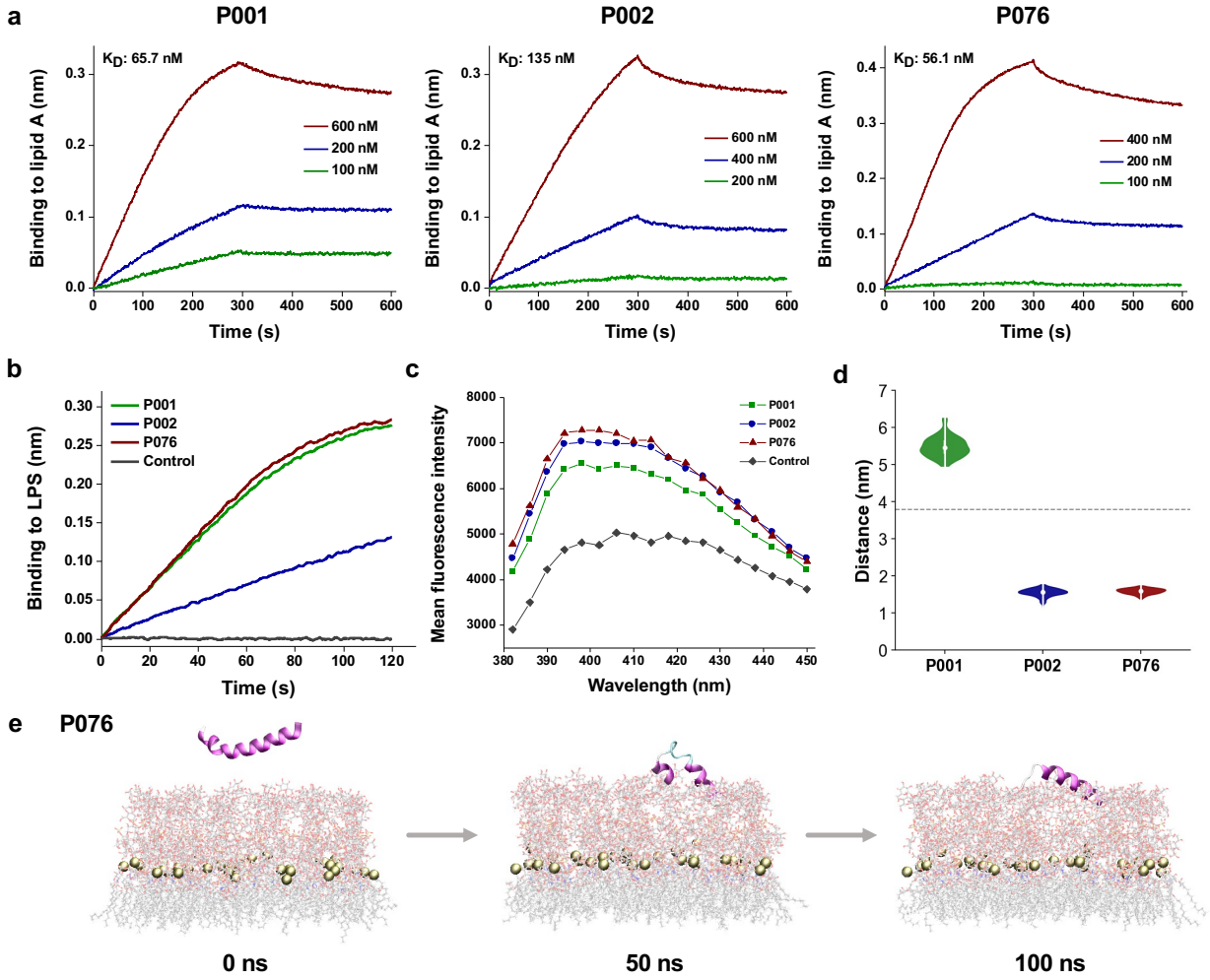


Figure 5. Bacterial membrane-acting assessment of AMPs.

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.

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 (supplementary Table S4).

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 made a tiny change of P076 C-terminal with amide modification but observed loss of functions (Table S5). 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* were 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.

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.

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

For the combined training of GAN generator and discriminator, the 3280 AMP sequences from PepVAE 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 downloaded and 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:

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

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

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

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

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.

502 In ablation study, three encoding approaches (chemical fingerprint, contact matrix, and ESM) and
503 their different combinations were tested individually. It was noted that the peptide graph exists
504 only when both the node embedding and contact matrix are present together. Therefore, a CNN
505 with the same number of layers and output dimension was adopted instead of the GCN in ablation
506 test.
507
508 **Peptide synthesis.** The peptide P001, P002, and P076 were synthesized by DGpeptide Co., ltd.
509 with 95% purity via solid-phase peptide synthesis procedures. Mass spectrometry were used to
510 verify the molecular weights (Figure S4).
511
512 **MIC detection.** The minimum inhibitory concentrations (MICs) of peptides against bacteria
513 including *A. baumannii* (ATCC 19606), multidrug-resistant *A. baumannii* (MDRAB) ,[46] *S.*
514 *aureus* (ATCC 25923), methicillin-resistant *S. aureus* (MRSA, ATCC 43300), *E. coli* (ATCC
515 25922), multidrug-resistant *E. coli* (MDREC),[47] *P. aeruginosa* (PAO1), and multidrug-resistant
516 *P. aeruginosa* (MDRPA),[48] were determined using the broth microdilution method recommended
517 by the Clinical and Laboratory Standards Institute (CLSI) with modifications reported by the
518 Hancock laboratory.[49] P001, P002, and P076 peptides were prepared in 0.01% acetic acid
519 containing 0.2% BSA at concentrations of 6.25, 12.5, 25, 50, 100, 200, 400, 800, 1600, and 3200
520 μg/mL. Bacteria grown to the mid-logarithmic phase in Meller-Hinton broth (MHB) were diluted
521 to 2×10^5 CFU/mL. Aliquot of 100 μL of the bacterial suspension was then co-incubated with 11
522 μL of the peptides at 37°C in a sterile polypropylene microtiter plate (3365, Corning, Shanghai,
523 China). An M2e microplate reader (Molecular Devices, Silicon Valley, CA, USA) was applied to
524 detect the bacterial absorbance at 600 nm after 18 h. The recorded MIC was the lowest
525 concentration of peptides causing a 70% reduction in bacterial turbidity.
526
527 **Hemolytic assay.** The hemolysis of peptides was evaluated as previously described in Ref.[50]
528 Briefly, the fresh mouse blood was centrifuged at 1000 g for 10 min and washed with cold PBS
529 three times. The erythrocytes were collected and diluted in saline solution to prepare a 2% (v/v)
530 suspension. A total of 75 μL of the erythrocyte suspension was co-incubated with 75 μL of the
531 peptides (6.25, 12.5, 25, 50, 100, 200, and 400 μg/mL) at 37°C for 1 h. After centrifuging at 1000

532 g for 10 min, 60 μL supernatant was obtained, and the absorbance was determined at 414 nm. This
533 experiment was conducted in triplicate and repeated twice.
534
535 ***In vivo* antibacterial evaluation.** The therapeutic efficacy of P076 peptide against drug-resistant
536 bacteria was analyzed using a mouse peritoneal infection model.[38] The 8-week-old female C57
537 mice (n = 10) was intraperitoneally administered with 3×10^5 CFU/mL of MDRAB. P076 (2
538 mg/kg) was given by intraperitoneal injection at 0.5 and 2.5 h after bacterial infection.
539 Ciprofloxacin (CIP, C129896, Aladdin, 32 mg/kg) was employed as the control. Mouse survival
540 was monitored for 48 h. For detection of the bacterial colonization, another 24 mice were randomly
541 divided into four groups (n = 6 for each group). The mice were euthanized 12 h after MDRAB
542 infection, and the peritoneal cavity was rinsed with 2 mL of PBS. Additionally, mouse liver and
543 spleen were obtained by surgery and homogenized in 1 mL of PBS. The lavage fluid and tissue
544 suspension were both diluted for bacterial colony counting using an MHB agar plate as described
545 in MIC detection.
546
547 ***In vivo* toxicological evaluation.** The 8-week-old female mice (18-22 g) were randomly divided
548 into 14 groups (n = 10 for each group). P076 peptide and polymyxin B (PB, P105490, Aladdin,
549 Shanghai, China) were given to the mice by intraperitoneal injection at 4, 8, 16, 32, 64, and 128
550 mg/kg. Mouse survival was monitored for 48 h. For pathologic analysis, mouse liver, spleen, and
551 kidneys were obtained by surgery for hematoxylin-eosin (HE) staining after treatment with 64
552 mg/kg P076 and PB for 12 h. Mice were cared for and treated in accordance with the NIH
553 guidelines for the care and use of laboratory animals (NIH Publication No. 85e23 Rev. 1985).
554 Animal experiments were conducted with approval from the Animal Experimental Ethics
555 Committee of the Army Medical University (AMUWEC20223290).
556
557 **Cell cultures**. Vero-E6 (African green monkey kidney cell line), Huh-7 (Hepatocellular carcinoma
558 cell line), A549 (Human Lung carcinoma cell line) and HeLa-hACE2 were cultured in DMEM
559 (Thermo Fisher Scientific, Waltham, USA) containing 10% fetal bovine serum in a 5% CO_2
560 incubator at 37°C. HSV-1 and CHIKV were propagated in Vero-E6 cells. DENV-2 and HTNV
561 were propagated in Huh-7 and A549 cells, respectively. And SARS-CoV-2 were propagated in
562 HeLa-hACE2 cells.

563

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

571

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

581

582 **Real-time PCR.** The cells were cultured in a 6-well plate (2×10^5 cells/well) following the same
583 cell culture procedures as in the immunofluorescence assay. Cellular RNAs were extracted and
584 purified using a viral RNA extraction kit (Qiagen, Dusseldorf, Germany), and then cDNA was
585 synthesized and monitored with SuperScript-III kit (Takara, Dalian, China). qRT–PCR analysis
586 was performed using SYBR Green (BIO-RAD, California, USA) according to the manufacturer's
587 protocol. Viral RNA expression was calculated using the $2^{-\Delta\Delta C.T.}$ (cycle threshold) method
588 normalized to GAPDH expression. The qRT–PCR primers sequences for HTNV, CHIKV, HSV-
589 1 and DENV-2 are listed in the Table S6, while the detection kits (Daan gene, Guangzhou, China)
590 were used for SARS-CoV-2 wild-type and BA.2 subtype.

591

592 **Transmission electron microscopy.** The cells were cultured in a 6-well plate (2×10^5 cells/well)
593 following the same cell culture procedures as in the immunofluorescence assay. The virus-infected

594 cells were collected by centrifugation to prepare cell particles and fixed in cacodylate sodium
595 buffer containing 1% glutaraldehyde (0.2 M, pH 7.2). The fixed sample was dehydrated by acetone
596 solution and embedded in epoxy resin. The sample was polymerized at 60°C for 3 days. The resin
597 block was used to prepare ultrathin slices (50-70 nm thick). These sections were supported by
598 copper mesh, negatively stained with uranyl acetate and lead citrate (electron microscopy) and
599 observed with a JEM100SX transmission electron microscope (JEOL, Tokyo, Japan).
600
601 **Biolayer interferometry**. The Octet Red96 BLI platform (Sartorius BioAnalytical Instruments,
602 Bohemia, USA) was utilized to measure the binding kinetics of three peptides to lipid A (L5399,
603 Sigma, Shanghai, China). Peptides were diluted with 75 mM NaCl solution at concentrations of
604 100, 200, 400, and 600 nM. The amine reactive second-generation (AR2G) biosensors were
605 processed with 1-ehtyl-3-(3-dimethylaminopropy) carbodiimide hydrochloride (EDC, E1769,
606 Sigma) and sulfo-N-hydroxysulfosuccinimide (s-NHS, 56485, Sigma). Subsequently, lipid A was
607 loaded onto the activated AR2G biosensors at 30°C for 5 min, and both the processes of association
608 and disassociation were monitored for 5 min to measure the kinetic constants. The equilibrium
609 dissociation constant (K_D) was defined as the division of dissociation rate constant (K_{off}) and the
610 association rate constant (K_{on}). In case of LPS, AR2G biosensors immobilized with 15 μg/mL LPS
611 (L2880, Sigma) were incubated with 500 nM peptides for 2 min in 75 mM NaCl solution. The
612 binding thickness was recorded and analyzed using ForteBio Data Analysis 7.0 software.
613
614 **NPN uptake assay.** MDRAB in mid-logarithmic phase was diluted to 1×10^8 CFU/mL in 10 mM
615 sodium phosphate buffer (pH 7.4). Then 10 μL of each peptide was mixed with 190 μL of MDRAB
616 at 37°C for 1 hour. Afterwards, 1-N-phenylnaphthylamine (NPN, 104043, Sigma, Shanghai, China)
617 was added with a final concentration of 10 μM. The fluorescence intensity was assessed with the
618 wavelength ranging from 380 to 450 nm at 2 nm intervals, after excitation at 350 nm using a Tecan
619 Infinite M1000 Pro microplate reader (Mannedorf, Zurich, Switzerland). Each experiment was
620 conducted in duplicate and repeated three times.
621
622 **Molecular dynamics simulations.** The initial structures of peptides were built via AlphaFold2
623 single sequence version.[51] Lipid bilayer models were constructed by CHARMM-GUI server along
624 with placing parallel peptides approximately 2 nm above the membrane surface.[52] Nine systems

625 were built for three peptides versus three kinds of lipid bilayers (viral envelopes, Gram-negative
626 bacterial inner membranes and outer membranes).[39, 53] Ions of 150 mM NaCl were added to
627 neutralize the system, and the calcium ions were used for LPS in outer leaflet of bacterial outer
628 membranes. Detailed system settings are recorded in Table S7. The standard CHARMM-GUI
629 equilibration protocols including NVT and NPT ensembles were adopted, keeping the temperature
630 at 310K. All-atom MD simulations were performed using GROMACS 2021.2 with the
631 CHARMM36 force field and the TIP3P water model for 100 ns, with a time step of 2 fs.[54, 55] Three
632 independent run was performed for each system. The visual molecular dynamics (VMD) software
633 version 1.9.4 was used for visualization.
634
635 The distance between the center of mass (COM) of peptides and phosphorus atoms in top leaflets
636 was calculated via *gmx distance*. Binding free energy was calculated using the molecular
637 mechanics Poisson-Boltzmann surface area (MM/PBSA) method on average of extracted 20
638 frames from the last 20 ns. The binding free energy of a peptide-membrane complex is calculated
639 as below:

640 $$\Delta G_{bind} = G_{complex} - (G_{peptide} + G_{membrane}) \quad (6)$$

641 $$G_{bind} = G_{polar} + G_{nonpolar} = \left(E_{elec} + G_{psolv}\right) + (E_{vdw} + G_{sasa}) \quad (7)$$

642 where E_{elec} is the Coulomb potential, E_{vdw} is the van der Waals energy, G_{psolv} and G_{sasa} are polar
643 and nonpolar solvation free energy. Following previous study, the dielectric constant was set as 7
644 for phospholipids and the entropy term was neglected.[56]
645

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).

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.

Conflict of Interest

There are no conflicts to declare.

670 ## References

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

673 2. C. Antimicrobial Resistance, Global burden of bacterial antimicrobial resistance in 2019:
674 a systematic analysis, *Lancet*, 2022, **399**, 629-655.

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

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

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

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

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

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

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

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

697 11. G. Shi, X. Kang, F. Dong, Y. Liu, N. Zhu, Y. Hu, H. Xu, X. Lao and H. Zheng, DRAMP
698 3.0: an enhanced comprehensive data repository of antimicrobial peptides, *Nucleic Acids*
699 *Res*, 2022, **50**, D488-D496.

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

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

14. C. Li, D. Sutherland, S. A. Hammond, C. Yang, F. Taho, L. Bergman, S. Houston, R. L. Warren, T. Wong, L. M. N. Hoang, C. E. Cameron, C. C. Helbing and I. Birol, AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens, *BMC Genomics*, 2022, **23**, 77.

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

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

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

18. J. Maasch, M. D. T. Torres, M. C. R. Melo and C. de la Fuente-Nunez, Molecular de-extinction of ancient antimicrobial peptides enabled by machine learning, *Cell Host Microbe*, 2023, **31**, 1260-1274 e1266.

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

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

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

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

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

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

25. A. Pandi, D. Adam, A. Zare, V. T. Trinh, S. L. Schaefer, M. Burt, B. Klabunde, E. Bobkova, M. Kushwaha, Y. Foroughijabbari, P. Braun, C. Spahn, C. Preusser, E. Pogge von Strandmann, H. B. Bode, H. von Buttlar, W. Bertrams, A. L. Jung, F. Abendroth, B.

745 Schmeck, G. Hummer, O. Vazquez and T. J. Erb, Cell-free biosynthesis combined with
746 deep learning accelerates de novo-development of antimicrobial peptides, *Nat Commun*,
747 2023, **14**, 7197.

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

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

752 28. A. T. Muller, G. Gabernet, J. A. Hiss and G. Schneider, modlAMP: Python for antimicrobial
753 peptides, *Bioinformatics*, 2017, **33**, 2753-2755.

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

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

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

762 32. J. Li, Y. Pu, J. Tang, Q. Zou and F. Guo, DeepAVP: A Dual-Channel Deep Neural Network
763 for Identifying Variable-Length Antiviral Peptides, *IEEE J Biomed Health Inform*, 2020,
764 **24**, 3012-3019.

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

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

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

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

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

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

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

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

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

42. W. Jacob and W. Zack, Deep learning regression model for antimicrobial peptide design, *bioRxiv*, 2019, DOI: 10.1101/692681, 692681.

43. L. Tzu-Tang, Y. Li-Yen, W. Ching-Tien, L. Ga-Wen, K. Chi-Fong, S. Yang-Hsin, C. Shu-Hwa and L. Chung-Yen, Discovering Novel Antimicrobial Peptides in Generative Adversarial Network, *bioRxiv*, 2021, DOI: 10.1101/2021.11.22.469634, 2021.2011.2022.469634.

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

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

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

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

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

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

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

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

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

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

54. R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. Mackerell, Jr., 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*, 2012, **8**, 3257-3273.

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

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