RESEARCH ARTICLE

Screening of Antimicrobial Peptides from Duck Farm Soil Metagenomic Sequences Based on Deep Learning and Molecular Dynamics Simulation

Shu Wang[1,*]

[1] School of Light Industry Science and Engineering, Beijing Technology and Business University, Beijing 100048, China

Abstract

Antimicrobial peptides (AMPs) are promising candidates in the fight against antibiotic resistance. However, traditional screening methods are often complex and costly. Metagenome offers a powerful approach for discovering novel AMPs from environmental samples. In this study, the deep learning model AMP-CLIP was employed to screen for novel AMPs from metagenomes of livestock farm soil and feces, which are rich in microbial diversity. The model achieved a high prediction accuracy of 99%.To enhance model interpretability, SHAP analysis was performed, revealing that amino acids contributing most to antimicrobial activity predictions were predominantly positively charged residues, suggesting that the model captures biologically meaningful features. Based on model scores and additional screening using Hemopred and ToxinPred, four candidate peptides were selected.The chemically synthesized peptides exhibited inhibitory activity against Klebsiella pneumoniae, Escherichia coli, and Staphylococcus aureus. Among them, Peptide 5 (GLGMGPVHLVGW) and Peptide 3 (SLGHAAGALAAHAR) were further analyzed using molecular dynamics (MD) simulations. Peptide–membrane complex systems were constructed with POPE:POPG (1:3) representing Gram-positive bacteria, POPE:POPG (2:1) representing Gram-negative bacteria, and DPPC for mammalian cell membranes. MD simulations revealed that both peptides inserted deeper and bound more stably in bacterial membranes, while exhibiting weaker interactions with mammalian membranes, indicating selective antimicrobial potential targeting bacterial membranes. Given their activity against foodborne pathogens and selective membrane interaction, these peptides hold promise as natural antimicrobial agents for improving food safety and shelf life.

Keywords: antimicrobial peptides, metagenomics, explainable deep learning, molecular dynamics, membrane–peptide systems.

1 Introduction

In recent years, with the overuse and abuse of antibiotics, microbial resistance has gradually

Submitted: 17 January 2026
Accepted: 23 March 2026
Published: 26 March 2026

Vol. 3, **No.** 1, 2026.
10.62762/ASFP.2026.132931

***Corresponding author:**
✉Shu Wang
wangshu0926@163.com

Citation
Wang, S. (2026). Screening of Antimicrobial Peptides from Duck Farm Soil Metagenomic Sequences Based on Deep Learning and Molecular Dynamics Simulation. *Agricultural Science and Food Processing*, 3(1), 25–48.

© 2026 by the Author. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

become a serious problem affecting human health and becoming a major threat to global health. At present, antimicrobial peptides have broad application prospects, and their unique antimicrobial activity can effectively overcome antibiotic resistance [1]. At the same time, some studies have shown that antimicrobial peptides also have certain anti-proliferation and apoptosis-inducing effects on tumor cells [2], Antimicrobial peptides not only have a wide range of biological activities and potential medicinal value, but also have a wide range of application prospects in food, and become an effective tool to extend the shelf life of food and ensure food safety. Because it can target bacterial cell membranes and biological macromolecules, it has a broad spectrum of antimicrobial activity, and can effectively inhibit the growth of a variety of food rot bacteria and pathogenic bacteria, such as Escherichia coli and Staphylococcus aureus, thereby extending the shelf life of food [3]; Antimicrobial peptides as natural preservatives, compared with traditional chemical preservatives, has higher safety and environmental protection, easy to be degraded by the human body, reduce the potential risk to the human body [4]. According to statistics and trend analysis, the global demand for antimicrobial peptides (AMPs) continues to rise. Therefore, developing more efficient, accurate, and time- and cost-saving methods for screening antimicrobial peptides has become an urgent task to effectively control and eliminate pathogens in food, thereby protecting human health [5]. A global concern of antibiotic resistance indicates a vital necessity to discover or design new antimicrobial peptides and implement already known AMPs [6, 7].

Compared with traditional antibiotics that target specific biosynthetic pathways (such as protein and cell wall synthesis), AMP disrupts the stability and physical integrity of bacterial membranes, thereby circumvent classical drug resistance mechanisms [8].

However, the research and application of antimicrobial peptides still face huge challenges. The traditional wet experiments for the research of antimicrobial peptides have many shortcomings. They require a lot of time and human resources. The steps from the extraction, purification to activity verification of AMPs take a long time and the sample size processed is small. By enabling the direct analysis of genetic material from environmental samples, metagenomics has revolutionized how microbial communities are studied, overcoming the limitations of traditional culture-based methods [9]. With the significant growth of genomic data and the development of bioinformatics technology in recent decades [10], the mining methods of antimicrobial peptides based on machine learning and deep learning provide great convenience for laboratory research. Machine learning-based methods in the prediction and discovery of antimicrobial peptides (AMPs) usually rely on protein descriptors [11], extract digital features from protein sequences and use them to characterize the physicochemical properties, structural characteristics and functional attributes of proteins as input; Many machine learning tools based on protein descriptors have been developed for the prediction of antimicrobial peptides, such as CAMPR3 [12], iAMPpred [13], AMPScanner [14]. Deep learning is a machine learning method based on artificial neural networks, which can automatically learn features from data without relying on artificially designed descriptors. Common models include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers. Examples of such models applied to antimicrobial peptide prediction are AMP-BERT [15], AMPFinder [16], DeepPeptide [17].

Molecular dynamics simulation (MD) can study molecular behavior at the atomic scale, track the interactions between molecules and their dynamic changes. This method has high spatiotemporal resolution and is capable of revealing the details of molecular mechanisms and interactions, which is difficult to be fully achieved by other experimental or computational methods [18]. Therefore, it can provide valuable supplements to the experimental information regarding the details of the interaction between peptides and biofilms.

The antimicrobial mechanism of antimicrobial peptides is very complex. Usually, they inhibit and kill bacteria by targeting the protein synthesis, nucleic acid synthesis of bacteria or disrupting the metabolic processes of cells through multiple pathways. However, the main mechanism of action of most antimicrobial peptides is by destroying the integrity of cell membranes [19]. In addition to their microbicidal activity, several cationic AMPs have been proven to have cytotoxic effects on cancer cells. The anti-cancer activity of AMP can usually occur through membrane dissolution or non-membrane dissolution mechanisms [20]. In fact, the presence of anionic components in cancer cell membranes (such as phosphatidylserine, O-glycosylated mucin, sialylated ganglioside and heparan sulfate) enables cationic anti-cancer peptides to target and bind to these cells,

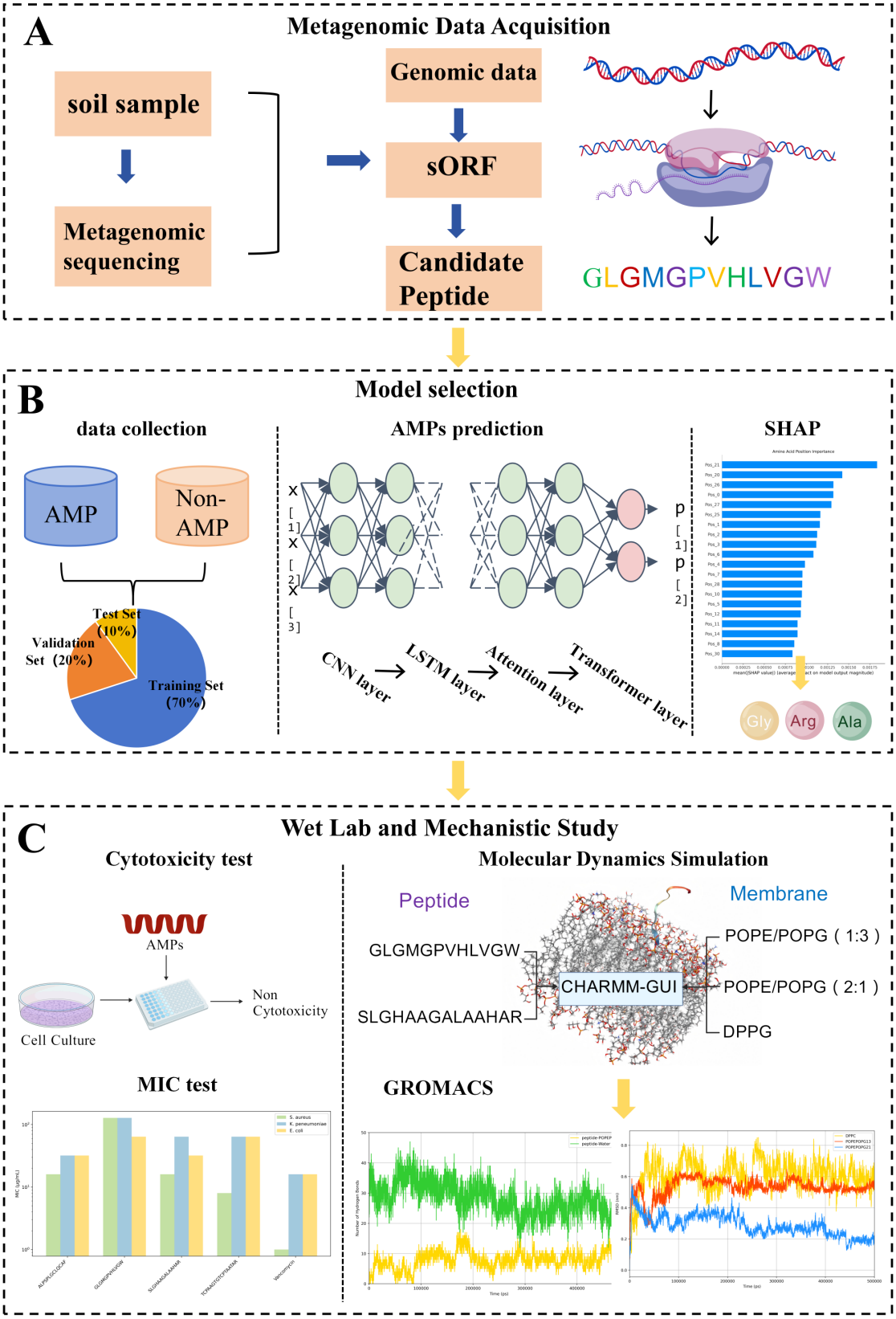

Figure 1. Research framework of the study. (A) Metagenomic sequencing of soil samples, ORF prediction to obtain a candidate peptide dataset, followed by data preprocessing. (B) Construction of training, validation, and test datasets, screening of antimicrobial peptides using machine learning models, and SHAP-based interpretability analysis of the models. (C) Experimental validation of the selected peptides, and molecular dynamics simulations of different membrane–peptide systems to investigate their transmembrane mechanisms.

while normal mammalian cells are generally neutral in charge due to amphoteric phosphatidylcholine and sphinolipin [21]. This leads to the leakage of cell contents and eventually causes the death of bacteria. Due to the crucial role of the cell membrane in the survival of bacteria, antimicrobial peptides interact with the membrane to cause phenomena such as membrane rupture and pore formation, thereby affecting the structure and function of the cell membrane. Therefore, in-depth research on the interaction between antimicrobial peptides and cell membranes is the key to understanding their antimicrobial mechanisms.

However, traditional machine learning methods still face various challenges in the high-throughput screening process of antimicrobial peptides (AMPs), such as insufficient training data, reliance on experience for feature extraction, and inaccurate processing of uncertainties. Although deep learning methods have been widely used in the prediction of AMPs in recent years, their complexity often leads to model overfitting, and as a "black box model", it is difficult to explain the prediction logic, and the hyperparameter tuning process is also rather complex [5]. Therefore, in view of the incomplete process of "prediction - verification - mechanism analysis" in AMPs research, this study innovatively constructed a three-in-one antimicrobial peptide screening process of "bioinformatics + membrane action molecular dynamics simulation + experimental verification". We took the soil-manure samples from duck farms as the original materials, extracted and analyzed their metagenomic data, predicted potential AMP candidate peptides through multi-layer bioinformatics methods, selected the candidate peptides with high scores for in vitro antimicrobial activity experiments and hemolytic toxicity detection, and innovatively introduced the molecular dynamics simulation strategy of membrane peptide interactions. Evaluate its binding ability with the membrane, embedding depth and the influence on membrane stability. Different from previous studies, in this study, membrane simulation analysis, as a mechanism deepening step after experimental verification, is helpful to reveal the functional mechanism of AMPs from the structural and dynamic levels. The discovery of such membrane-active and selective AMPs not only contributes to antimicrobial drug development, but also provides new candidates for safe and effective antimicrobial agents in food preservation. The overall research framework of this study is illustrated in Figure 1, which outlines the three-stage workflow encompassing bioinformatics screening, experimental validation, and molecular dynamics simulation-based mechanism analysis.

2 Materials and Methods

2.1 Genetic Sequencing and Data Analysis

Metagenomic sequencing mainly includes four steps: sample preparation, library construction, sequencing and data analysis [22]; Samples were collected from surrounding breeding farms. Soil from different areas of the chicken farm was selected and mixed for sampling. 200-500g of each sample was collected. 3-5 duplicate samples were taken, mixed evenly and then packaged separately. Take 5-10g and pass it through a 2mm sieve. Homogenization treatment: Take 0.2-0.5g of feces into a sterile centrifuge tube, add 1mL of PBS (pH7,4), vortex for 5 minutes, centrifuge at 5000×g for 5 minutes, discard the supernatant, collect the precipitate, and repeat 2-3 times until the supernatant becomes clear. To remove a large amount of the inhibitor from the sample: add 2%CTAB+0.7M NaCl (preheat to 65°C), vortex for 5 minutes, add an equal volume of chloroform: isopentanol (24:1), shake vigorously for 5 minutes, centrifuge at 10000×g for 10 minutes, and take the supernatant. Sample homogenization: add 0.5g of glass beads (0.3mm), use Bead Beater(6000rpm,5min), and ice bath for 2min. Repeat this process three times to ensure complete microbial disruption. Total DNA was extracted using the commercial DNA extraction kit provided by Novogene Co., Ltd. After extraction, the DNA samples were constructed into libraries and double-ended sequencing was performed using the Illumina platform. The sequencing data were output in FASTQ format.

2.2 Construction of the Peptide Dataset

Following quality control and assembly of the metagenomic sequencing data, open reading frames (ORFs) were predicted from the assembled DNA sequences (FASTA format) using the EMBOSS getORF tool [1], resulting in the corresponding amino acid sequences. Subsequently, the NCBI BLAST tool (Basic Local Alignment Search Tool) was utilized to align these sequences and filter out non-protein-coding sequences. To construct the peptide segment dataset, python scripts were written by. The Biopython module was used to conduct in vitro simulated enzymatic hydrolysis of these sequences. The simulated digestive

[1] https://bioinformatics.nl/emboss-explorer/

enzymes included trypsin, pepsin I and pepsin II [23], Generate peptide segments of different lengths. The CD-HIT tool was used to cluster and remove redundancies for all generated peptide segments, filter out repetitive sequences with similarity thresholds, and optimize the structure of the data set. Finally, all non-redundant enzymatically hydrolyzed peptide segments were integrated as the basic data set for subsequent peptide analysis.

2.3 Data Collection

According to previous research, dataset quality significantly affects the accuracy of model construction and is a critical component of the modeling process [24]. As of March 2025, AMP data were collected from publicly available databases, including APD3 [25], CAMPR3 [12], DRAMP [26]. The AMP datasets obtained from the three databases were integrated together and repetitive sequences were removed. A total of 3,600 refined AMP sequences were obtained, with lengths ranging from 13 to 154 amino acids. During data preprocessing, 433 sequences longer than 50 amino acids were removed to reduce redundancy and potential noise in the dataset. Eventually, 3,167 non-repetitive AMP sequences were obtained, which will be used for subsequent analysis and model construction. The non-antimicrobial peptide (non-AMP) data were derived from the public protein database UniProt. Keywords unrelated to "antimicrobial activity" were selected, amino acid sequences with a length of less than 50 were downloaded, and incomplete or invalid sequences were removed to maintain consistency with the AMP dataset. A total of 3,167 sequences were randomly selected from the above non-AMP data for subsequent analysis. In this study, the dataset was randomly split into three parts: 70% for training, 20% for validation, and 10% for testing. Positive and negative samples were proportionally distributed across all subsets.

2.4 Model Composition

This study is based on the deep learning framework AMP-Potency-Prediction-and-EvoGradient, originally proposed by Wang et al. [27] and implemented using PyTorch, with several innovative improvements. We introduced modules to enhance model interpretability, while optimizing the model architecture and training strategies to improve the screening accuracy of antimicrobial peptides against multi-drug resistant pathogens. Built upon the AMP-CLIP classification model, which integrates CNN, LSTM, Attention, and Transformer architectures, we refined both the structure and training procedures to further boost the identification precision of antimicrobial peptides. Finally, peptides classified as antimicrobial by the improved AMP-CLIP model were selected for subsequent validation.

AMP-CLIP adopts four DL architectures: CNN, LSTM, Attention and Transformer. To reduce the false positive results where the model wrongly predicts non-AMP as AMP, a conservative loss function is introduced, which is achieved by adding a penalty term based on the difference between the model's prediction and the actual label on the basis of the cross-entropy loss. This penalty mechanism makes the model's prediction more cautious, thereby reducing the FP rate. Furthermore, a multi-model integration strategy is adopted. Only when all four models recognize the peptide as AMP and the prediction probability exceeds 0.5, will it be classified as AMP to further improve the accuracy of the method. This model utilizes an embedding layer to represent the AA sequence. The CNN layer is used to capture local features, aiming to capture the dependencies between adjacent AAs in the peptide sequence. This model utilizes an embedding layer to represent the AA sequence, followed by three 1D Conv layers with different filter sizes to extract local features. These convolutional layers are supplemented by a 1D dense layer as the maximized pooling layer to capture the most prominent information. Finally, the dense layer is adopted as the classification head to make predictions based on the extracted features; The LSTM layer is used to process sequence information and extract long-term dependencies in the sequence. This architecture includes an embedding layer to represent peptides, followed by a double LSTM layer, which captures the forward and backward sequential dependencies. Then the output of the embedding layer is connected with the output of the bi-LSTM layer to enable the model to capture local and global information. The dense layer is then used as a classification head to make predictions based on the learned representations. The Attention mechanism is used to enhance the model's ability to focus on key features, in order to capture long-term relationships and concentrate on the decisive AA. This architecture also includes an embedding layer. The multi-head attention layer allows the model to capture the interdependence between different positions and assign different degrees of importance to a single AA. Furthermore, the output of the attention layer, through the dense layer serving as the classification head, can predict the antimicrobial activity of the

peptide. The Transformer layer is used to capture the complex dependencies between AAs. The Transformer architecture consists of an AA embedding layer and a position embedding layer. This model contains two encoders, each of which includes an attention layer and a feedforward layer. The attention layer captures the relationships between different positions, enabling the model to effectively capture long-term dependencies. Finally, the dense layer is used as the classification head for the final prediction.

In this study, the following equations were used to evaluate the model performance, including parameters such as accuracy, precision, and specificity [28].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (3)$$

Among them, TN represents “true negative”, FN represents “false negative”, TP represents “true positive”, and FP represents “false positive”.

2.5 SHAP(SHapley Additive exPlanations)

Based on the model, this study conducts an interpretability analysis. SHAP (SHapley Additive exPlanations) is a model interpretation method based on the Shapley value in game theory and can be used to explain the predictions of any machine learning model [29]. It achieves the "interpretability analysis" of the model by quantifying the contribution of each feature to the prediction result. The core idea is the result produced by the "cooperation" of various features. Each feature makes a different contribution to this "cooperation", and SHAP quantifies these contributions in a mathematical way. Applying SHAP to explain the predictions of protein sequence models is suitable for reference in sequence tasks such as antimicrobial peptides [30]. The antimicrobial peptide model was globally interpreted using SHAP: which amino acids at certain positions had the greatest impact on the prediction of antimicrobial activity, locally explained the reasons why a certain peptide segment was predicted as an antimicrobial peptide, and determined whether a specific amino acid contributed positively or negatively at a certain position. This study performed SHAP (SHapley Additive exPlanations) interpretability analysis on CNN, LSTM, and Transformer models.

2.6 Solid phase peptide synthesis(SPPS)

In this study, the solid-phase peptide synthesis method was adopted to synthesize the screened antimicrobial peptide sequences, and the synthesis process was based on the Fmoc (9-fluorenylmethoxycarbonyl) protection strategy [31]. Rink amide resin was selected as the solid-phase carrier and was first swollen in DMF (N, N-dimethylformamide) for 30 minutes. Subsequently, the FMOC-protected amino acid was coupled with the resin through the C-terminal, and the condensation reaction was carried out in the presence of N,N’ -diisopropyl carbodiimide (DIC) and 1-hydroxybenzotriazole (HOBt), and N, N-diisopropyl ethylamine (DIPEA) was added to promote the coupling reaction. After the coupling is completed, the Fmoc protecting group is removed using a 20% piperidine /DMF solution to expose the amino terminal for the next round of amino acids. The above-mentioned "coupling → deprotection" cycle proceeds successively until the peptide chain synthesis is completed. Finally, the N-terminal of the peptide chain is blocked by acetic anhydride to avoid non-specific modifications in the subsequent reaction.

2.7 Circular Dichroism (CD) Spectroscopic Analysis

The CD spectra were collected using a highly sensitive circular dichroism spectrometer (Chirascan v100, Applied Photophysics Ltd., UK). First, prepare the antimicrobial peptide with a concentration of 150 μM using phosphate buffer (PBS, 10 mM, pH 7.4). Then pour the prepared antimicrobial peptide solution into the quartz cuvette. All measurements were carried out at room temperature (25$\pm$1℃). Each sample was subjected to three consecutive scans within the far ultraviolet (UV) wavelength range of 180-260 nm to evaluate the secondary structural characteristics of the synthesized antimicrobial peptide [32].

2.8 Identification of Peptide Sequences by LC-MS/MS

The structure of the antimicrobial peptide components was comprehensively characterized by liquid chromatography-tandem mass spectrometry (LC-MS/MS) [33]. The analysis process adopted an ion trap mass spectrometer (Thermo Scientific LCQ DECA XP MAX). Chromatographic separation was carried out at room temperature using a BioBasic C18 reversed-phase analysis column (Thermo Fisher Scientific, 150 mm × 2.1 mm, 5 μm). The mobile phase system contains two different solutions: Phase A is ultrapure water containing 5% acetonitrile (ACN)

and 0.1% formic acid (FA); Phase B is an aqueous solution containing 95% ACN and 0.1% FA. Gradient elution was adopted, and the flow rate of the mobile phase was a constant flow rate of 200 μL/min. The sheath gas flow rate is maintained at 50arb, the electrospray voltage is 4.0 kV, the capillary voltage is 20 V, and the temperature of the ion transfer capillary is stable at 300°C. Full scan mass spectrometry was collected using the centroid mode, and the mass-to-charge ratio (m/z) range was 150-1500 Da. Tandem mass spectrometry is generated through the DDA mode, preferentially selecting double-charge and triple-charge parent ions for collection-induced dissociation (CID), using 35% standardized collision energy.

2.9 Hemolysis Test

For the evaluation of hemolytic activity, freshly drawn human whole blood anticoagulated with heparin or EDTA was utilized. The blood sample was initially diluted at a 1:4 ratio (v/v) with ice-cold tris-buffered saline (TBS; 10 mM Tris-HCl, 150 mM NaCl, pH 7.2 $\pm$ 0.1) to prevent premature cell lysis. The diluted suspension was then centrifuged at 1500×g for 5 min at 4°C to pellet the erythrocytes. The supernatant, containing plasma and buffy coat, was carefully aspirated and discarded. The harvested human red blood cells (hRBCs) were subjected to three consecutive washing cycles with excess TBS (each wash followed by centrifugation under identical conditions) to ensure complete removal of residual plasma proteins and platelets. Finally, the purified hRBCs were resuspended in TBS to achieve a standardized 5% (v/v) working suspension, as determined by hematocrit measurement.

A serial twofold dilution of antibacterial peptide was prepared across the wells of a sterile 96-well microtiter plate using TBS as the diluent, generating a concentration gradient ranging from 12.5 μg/mL to 200 μg/mL. An equal volume of the 5% hRBC suspension was then dispensed into each well, resulting in a final erythrocyte concentration of 2.5% (v/v). To establish baseline references, negative control wells contained hRBCs suspended in TBS alone (0% hemolysis), while positive control wells were treated with 0.1% (v/v) Triton X-100 in TBS (100% hemolysis).

The plate was sealed with a breathable membrane to prevent evaporation and incubated under static conditions at 37 $\pm$ 0.5°C for 60 min in a humidified incubator. Post-incubation, the plate was gently agitated to resuspend any sedimented cells, and the optical density (OD) of the supernatant was measured at 405 nm using a microplate reader. The percentage of hemolysis was calculated using the following normalized equation:

$$\text{Hemolysis}(\%) = \frac{OD_{\text{peptide}} - OD_{\text{blank}}}{OD_{\text{control}} - OD_{\text{blank}}} \times 100 \quad (4)$$

where OD_{peptide} represents the OD value of hRBCs that were treated with antimicrobial peptides, OD_{control} represents the OD value of the hRBCs that were treated with Triton X-100, and OD_{blank} represents the OD value of TBS solution.

To ensure statistical robustness, all experimental conditions were assayed in triplicate across three biologically independent replicates, with mean values $\pm$ standard deviation (SD) reported.

All experiments were approved by the Ethics Committee of Beijing Technology and Business University. The collection and use of hRBCs were strictly carried out in accordance with relevant regulations. Informed consent was obtained from the donors for this study.

2.10 Cytotoxicity Test

The cytotoxicity test were assessed using a standard MTT (3-(4,5-dimethylthiazol-2-yl) -2,5-diphenyltetrazolium bromide) viability assay with NIH-3T3 cell line. Cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS), 100 U/mL penicillin, and 100 μg/mL streptomycin. Cultures were grown in polystyrene tissue culture flasks under standard incubation conditions (37°C, 5% CO_2, 95% relative humidity) until reaching 70-80% confluence.

For experimental procedures, adherent cells were enzymatically detached using 0.25% trypsin-EDTA solution and subsequently resuspended in complete growth medium. Cell density was adjusted to 5×10^4 cells/mL using a hemocytometer counting method with trypan blue exclusion to ensure >95% viability. Exactly 100 μL of this cell suspension was seeded into each well of a 96-well microplate, achieving a final density of 5×10^3 cells per well. The plate was incubated overnight to allow cell attachment and formation of a monolayer.

Following the attachment period, the culture medium was carefully aspirated and replaced with fresh antibiotic-containing medium. Test compounds were

prepared as a series of two-fold dilutions in culture medium, yielding final concentrations ranging from 12.5 to 200 μg/mL across the plate. After 24 hours of compound exposure at standard culture conditions, 10 μL of MTT reagent (5 mg/mL in phosphate-buffered saline, PBS) was added to each well. The plate was then incubated for 4 hours in complete darkness to allow formazan crystal formation by viable cells.

The reaction was terminated by careful removal of the culture medium. Formazan crystals were solubilized by adding 150 μL of dimethyl sulfoxide (DMSO) to each well, followed by gentle orbital shaking (100 rpm) for 15 minutes at room temperature to ensure complete dissolution. Control wells included: (1) blank controls containing DMSO only for background subtraction, and (2) untreated cell controls for determination of 100% viability.

Absorbance measurements were performed at 570 nm using a microplate reader, with reference readings at 630 nm to account for nonspecific absorption. The percentage cytotoxicity was calculated using the following formula:

$$\text{Cell viability}(\%) = \frac{OD_{\text{peptide}} - OD_{\text{blank}}}{OD_{\text{control}} - OD_{\text{blank}}} \times 100 \quad (5)$$

where OD_{peptide} denotes the absorbance values recorded from experimental wells containing DMEM supplemented with the antimicrobial peptides. OD_{control} corresponds to the baseline absorbance readings obtained from wells cultured in standard DMEM without peptide supplementation, serving as a negative control to account for background microbial growth. OD_{blank} represents the reference absorbance of dimethyl sulfoxide (DMSO) ensuring that any solvent-induced interference in the spectrophotometric measurements was properly normalized [34].

To ensure statistical robustness, all experimental conditions were assayed in triplicate across three biologically independent replicates, with mean values $\pm$ standard deviation (SD) reported.

2.11 MIC Test

To systematically assess the antimicrobial efficacy of the synthesized antimicrobial peptides, minimum inhibitory concentration (MIC) assays were performed against a panel of relevant bacterial strains including *K.peneumoniae*, *E. coli*, and *S. aureus*, with vancomycin serving as the reference antibiotic. Bacterial cultures were initially prepared by inoculating selected species in Luria-Bertani (LB) broth within sterile glass conical flasks, which were subjected to autoclaving at 121°C for 20 minutes to ensure sterility. The inoculated flasks were subsequently incubated at 37°C under constant agitation (200 rpm) in an orbital shaker for 10 hours to facilitate exponential-phase growth [35]. Following incubation, the bacterial suspensions were centrifuged at 4,000 × g for 10 minutes at 4°C, and the resulting pellets were washed twice with phosphate-buffered saline (PBS, pH 7.4) to remove residual medium components. The washed cells were then resuspended in Mueller-Hinton (MH) broth and adjusted to an optical density corresponding to a final concentration of 2×10^5 CFU/mL using spectrophotometric calibration.

For the MIC determination, a 96-well microtiter plate was employed, wherein each well received a serial twofold dilution of the test antimicrobial peptides or vancomycin, with concentrations range from 1 to 128 μg/mL in MH medium. An equivalent volume (100 μL) of the standardized bacterial suspension was introduced into each well, ensuring a homogenous mixture of inoculum and antimicrobial agent. Appropriate controls were incorporated, including wells containing only MH medium (sterility control) and wells with MH medium plus bacterial suspension (growth control). The plate was then incubated under static conditions at 37°C for 9 hours to permit bacterial proliferation in the presence of varying antimicrobial concentrations. Post-incubation, the optical density (OD) of each well was quantified at 600 nm using a microplate reader.

2.12 Molecular Dynamics Simulation

As a common target of antimicrobial peptides, in order to further explore the conformational behavior and mechanism of action of the screened antimicrobial peptides in the membrane environment, in this paper, the candidate peptides were embedded in the lipid bilayer of the model to construct a membrane-peptide complex system, and long-term molecular dynamics simulations were conducted to evaluate their functional potential from aspects such as structure, dynamic behavior, and stability.

The main components of bacterial cell membranes are phosphatidylethanolamine (PE) and phosphatidylglycerol (PG). PG carries a negative charge, making the membrane as a whole negatively charged. The screened peptides are cationic peptides. Take the inner membrane of Escherichia coli as an example. Its composition is approximately 70-80% PE, 20-25% PG and 5%

cardiolipin (CL) [36], Bacterial cell membrane mimics are represented by 1-palmitoyl-2-oil-based phosphatidylethanolamine (POPE) and 1-palmitoyl-2-oil-based phosphatidylglycerol (POPG), which are the main lipid components of bacterial cell membranes and account for approximately 80% and 20% respectively in experimental studies. In this study, POPE:POPG (2:1) membranes were also selected as the model of Gram-negative bacterial membranes because their composition is similar to the characteristics of most Gram-negative bacteria [37], The POPE:POPG (1:3) membrane was selected as the cell membrane of Gram-positive bacteria [38], Pure DPPC membranes were selected as animal cell membranes.

The 3D structures of the peptides were predicted online using the I-TASSER website [2], and saved in pdb format. Two peptides with the highest score coefficients were selected and different membranes were used to construct various membrane peptide systems for simulation using CHARMS-GUI [39–41]; In the POPE/POPG (1/3) membrane system of Gram-positive bacteria, there is 1 peptide molecule and 240 lipid molecules, and each layer contains 30POPE and 90POPG. In the membrane system of Gram-negative bacteria POPE/POPG (2/1), there is 1 peptide molecule and 240 lipid molecules, and each layer contains 80POPE and 40POPG. In the animal membrane DPPC membrane system, there is a total of 1 peptide molecule and 240 lipid molecules, with 120DPPC in the upper and lower layers respectively.

All the simulation systems used rectangular boxes. Each system contained a total of 24,278 atoms and was solvated with 4,454 water molecules. In each system, the peptide molecules were initially placed 20 A above the lipid bilayer surface and simulated in the NPT integrated system using the GROMACS2022.02 simulation package. All simulations were conducted at a temperature of 300K and a pressure of 1bar, with a simulated room temperature of 27 °C [37], CHARMM36m force field [42, 43]. Applied to all membrane peptide systems. To neutralize the system, the neutralizing ions Na^+ Cl^- 0.15M are used [44]. Periodic boundary conditions were applied in all directions of the simulation box, and the TIP3P (Transferable Intermolecular Potential with 3 Points) model was employed to represent water molecules [45].

Firstly, under the condition of positional constraints on the positions of peptide molecules, the system was subjected to a 1ns equilibrium simulation in the NVT ensemble (i.e., constant particle number, constant volume, and constant temperature), allowing the system temperature to gradually stabilize and avoiding the vigorous movement of peptide molecules at the beginning. Then, a 9-nanosecond equilibrium simulation was conducted in the NPT ensemble (i.e., constant particle number, constant pressure, and constant temperature), allowing the volume and pressure of the system to gradually reach a stable state and laying a solid foundation for subsequent production simulations. After completing the above balancing steps, a 500-nanosecond formal production simulation was carried out starting from the current initial configuration and atomic coordinates of the system [46].

In the production simulation stage, the system operates under the NPT ensemble, with a total simulation time of 500 ns. The Newton equation of motion is integrated using the leapfrog algorithm, and the integration step size is 2 fs. The temperature is maintained at 300 K through the velocity-rescaling temperature coupler, with a coupling time constant of 1.0 ps, acting on the solute, membrane and solvent groups respectively. The pressure is controlled at 1 bar through a semi-isotropic Parrinelo-Rahman pressure coupler, with a coupling time constant of 5.0 ps and a compression ratio set at 4.5×10^{-5} bar^{-1} [46], coupling is applied respectively in the xy plane and the z direction to adapt to the anisotropic deformation of the membrane structure [47]. All the bonds involving hydrogen atoms are constrained by the LINCS algorithm [48], thus allowing the use of a time step of 2 fs. The electrostatic interaction was treated by the particle grid Ewald (PME) method, with a truncation distance of 1.2 nm. Van der Waals interactions smoothly truncate between 1.0 and 1.2 nm using the force-switch function. The adjacency list is updated every 20 steps, and the non-key interaction adopts the Verlet truncation scheme. In all systems, unfavorable atomic contacts are eliminated by minimizing the energy of the steepest descent [49].

3 Results and Discussion

3.1 Model Construction and Evaluation Results

The FASTQ data of microbial DNA samples were obtained by metagenomic sequencing. After quality control and connector trimming, the sequencing data were processed to obtain clean data. The amino acid sequences of the open reading frames (sORF)

[2] https://zhanggroup.org/I-TASSER/

Table 1. Physicochemical properties of candidate peptides.

Seq	Sequence	Length	Hemolytic	Toxicity	Charge
3	SLGHAAGALAAHAR	14	Non-Hemolytic	Non-Toxin	2
5	GLGMGPVHLVGW	12	Non-Hemolytic	Non-Toxin	0.5
23	ALPSPLGCLQCAF	13	Non-Hemolytic	Non-Toxin	1.5
24	TCPAAGTGTCPTAATAR	17	Non-Hemolytic	Non-Toxin	1

were predicted using the getORF function, and the non-protein family sequences were eliminated using the BLASTP tool on the NCBI website. Short peptide sequences were obtained through in vitro simulated digestion with pepsin and trypsin II. Repetitive amino acid sequences were eliminated using the CD-HIT tool and compared with the known antimicrobial peptide database to remove the identified AMP. Finally, a predicted candidate peptide dataset containing 3,533 unique sequences was obtained.

The AMP-CLIP classification model was used to predict antimicrobial peptides, achieving an accuracy of 99% and a precision of 87.4%. It indicates that it is stable and suitable for the next step of antimicrobial peptide prediction. Figure 2 shows the amino acid composition distribution for the AMP and non-AMP sequences of the training set, as well as for the predicted AMP sequences, illustrating the characteristic differences and the rationality of the peptides screened by the model.

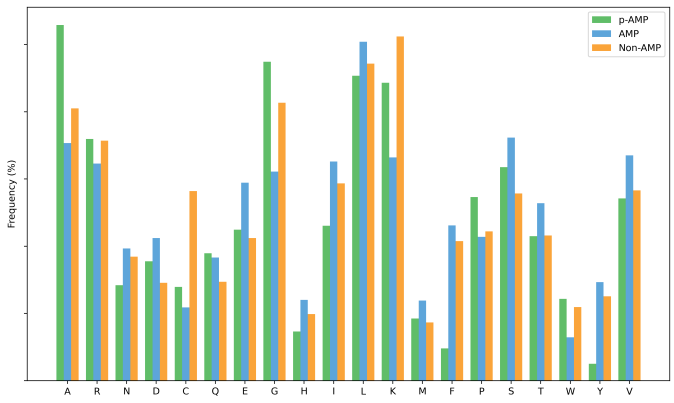

Figure 2. The figure presents the amino acid composition of the predicted p_AMPs from our study, alongside that of known AMPs and non-AMPs in the training datasets.

3.2 The Discovery of AMPs

The dataset was input into the prediction model for evaluation, and 644 peptides with potential activity were screened out. According to the model scoring results, the standard deviation thresholds of the scores of the four models, namely Transformer, LSTM, Attention, and CNN, were set at 0.2. Further selection finally led to 52 sequences. Toxicity and hemolytic activity were evaluated respectively using the online tools ToxinPred141 and HemoPI. Among them, it was found that four peptides were neither toxic nor hemolytic. Therefore, these four high-scoring, non-toxic and non-hemolytic peptides were identified as promising and safe antimicrobial candidates (see Table 1).

3.3 Analysis of SHAP Interpretability

After analysis, it was found that the 21st amino acid contributed the most to the model scores across all three models, as shown in Figures 3, 4 and 5. By writing a Python script, we identified that the most frequently occurring amino acids at position 21 are G (Glycine), A (Alanine), R (Arginine), and L (Leucine). These four amino acids make the greatest contribution to the antimicrobial peptide screening models and have the most significant impact on the prediction of antimicrobial activity. Glycine is the smallest amino acid, and because it lacks a side chain (having only a hydrogen atom), it provides greater conformational flexibility at that position in the peptide chain. Alanine and leucine are hydrophobic amino acids, and their hydrophobic nature promotes efficient insertion into bacterial membranes, thereby contributing to membrane disruption [19]; Arginine is a positively charged functional amino acid whose side chain enables strong electrostatic interactions with negatively charged bacterial membrane phospholipids. Extensive studies have demonstrated that positively charged residues are critical for antimicrobial activity, with arginine in particular exhibiting superior membrane-binding properties owing to its high polarity and capacity for hydrogen bonding [50].

In conclusion, the antimicrobial activity of peptides relies on a combination of interactions with bacterial membranes—namely electrostatic attraction, hydrophobic insertion, and conformational flexibility. The amino acids glycine (G), alanine (A), arginine (R), and leucine (L) exhibit these key properties,

Table 2. Hemolytic activity (%) of antimicrobial peptides at different concentrations.

Peptide (μg/mL)	6.25	12.5	25	50	100	200
ALPSPLGCLQCAF	0.38 ± 0.013	0.42 ± 0.013	0.51 ± 0.019	2.79 ± 0.11	5.85 ± 0.21	8.03 ± 0.34
GLGMGPVHLVGW	0.25 ± 0.011	0.37 ± 0.013	0.44 ± 0.017	3.82 ± 0.17	4.73 ± 0.22	8.29 ± 0.36
SLGHAAGALAAHAR	0.27 ± 0.012	0.31 ± 0.011	0.39 ± 0.014	2.88 ± 0.13	4.28 ± 0.24	7.31 ± 0.32
TCPAAGTGTCPTAATAR	0.32 ± 0.012	0.39 ± 0.014	0.47 ± 0.021	3.16 ± 0.14	4.92 ± 0.23	7.09 ± 0.31

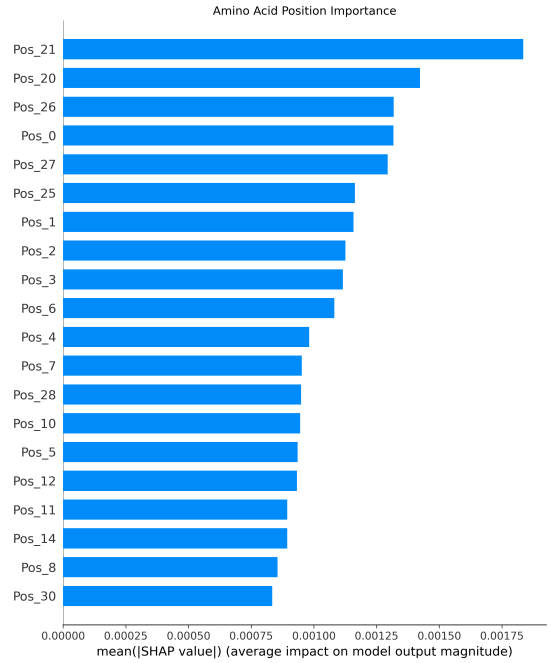

Figure 3. SHAP results of the CNN model.

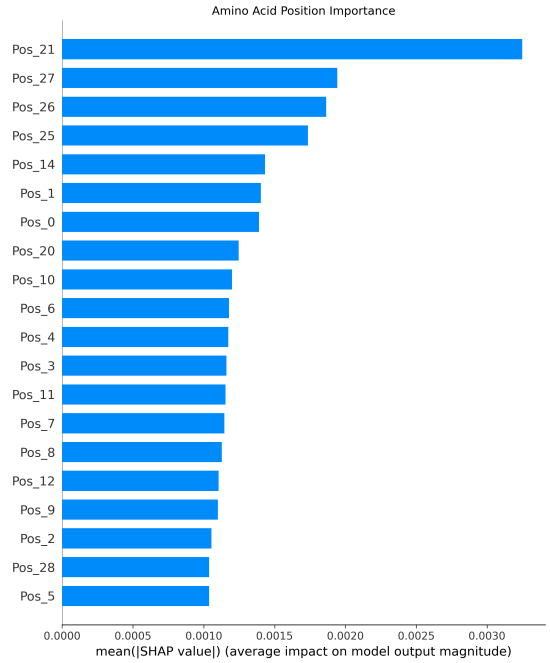

Figure 4. SHAP results of the LSTM model.

thereby exerting a notable impact on the prediction model's output. This highlights the model's capacity for biological interpretability.

3.4 Analysis of Wet Experiment Results

The four high-scoring, non-toxic and **non-hemolytic** peptides selected were chemically synthesized. The purity and molecular weight of the synthesized peptides were confirmed by mass spectrometry (MS), shown in Figure 6, ensuring the accuracy of the peptide sequences used for subsequent analyses. Circular dichroism spectroscopy analysis, hemolysis test, cytotoxicity test, and minimum inhibitory concentration test were continued.

The secondary structure characteristics of the peptide segments were analyzed by circular dichroism spectroscopy. The measured peptide segments had typical α-helical structure characteristics in the 190–230 nm region, that is, a positive peak appeared near approximately 192 nm.

To evaluate the biocompatibility of the candidate antimicrobial peptides, the hemolytic activity of four peptides against human red blood cells (hRBCs)

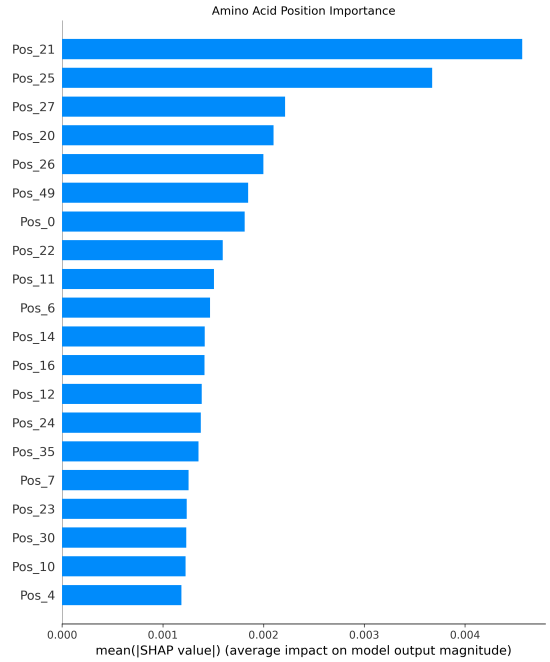

Figure 5. SHAP results of the Transformer model.

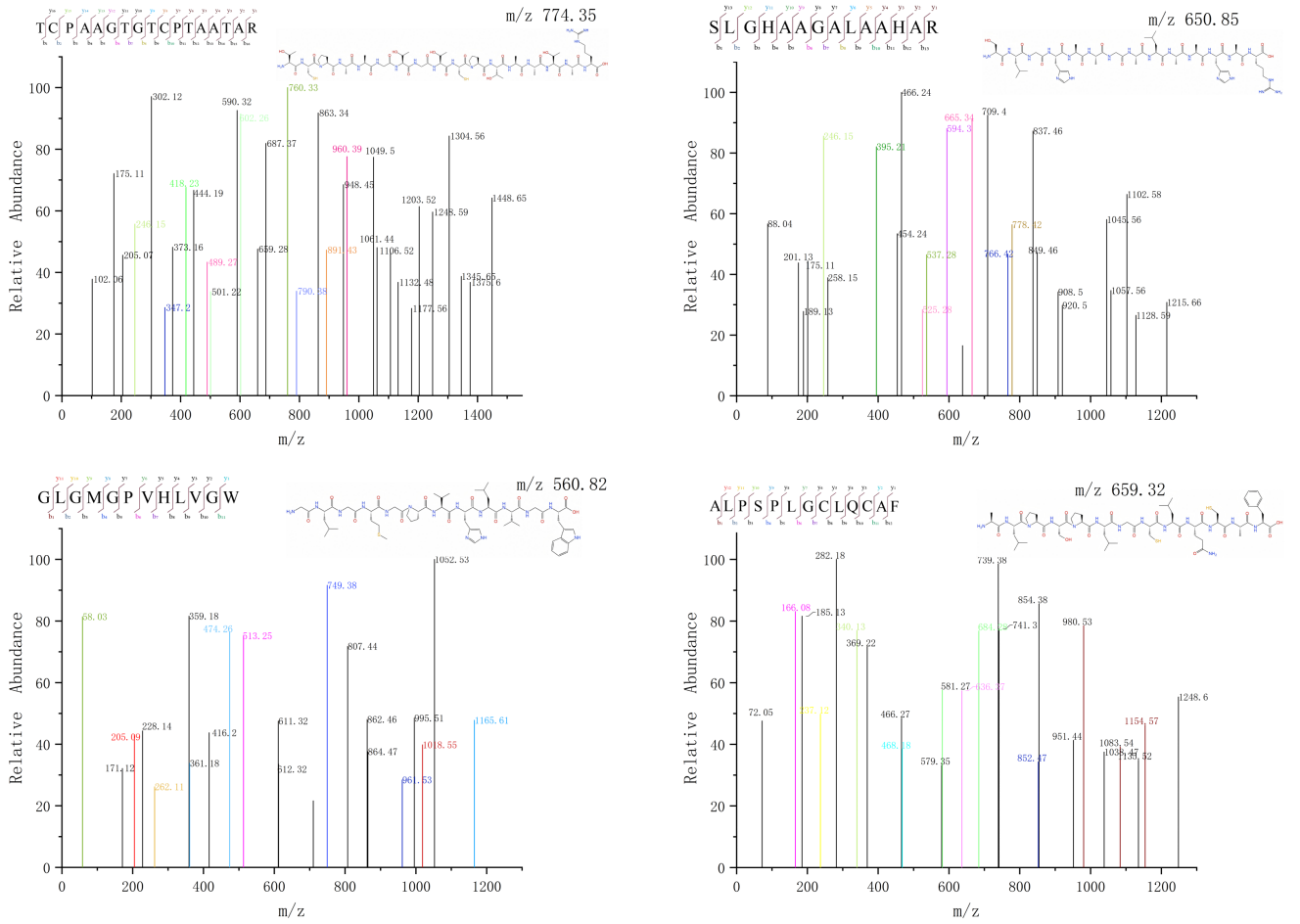

Figure 6. Liquid Chromatography–Mass Spectrometry (LC-MS) Analysis of the Selected Antimicrobial Peptides.

was examined at various concentrations ranging from 6.25 to 200 μg/mL. As summarized in Table 2, all peptides exhibited hemolysis rates below 1% at low concentrations (≤25 μg/mL), indicating minimal disruption to the erythrocyte membrane and good hemocompatibility. With increasing peptide concentration, the hemolysis rate gradually increased but remained at relatively low levels overall. At the highest tested concentration (200 μg/mL), the hemolysis rates of ALPSPLGCLQCAF, GLGMGPVHLVGW, SLGHAAGALAAHAR, and TCPAAGTGTCPTAATAR were 8.03%, 8.29%, 7.31%, and 7.09%, respectively (Table 2) —all well below the commonly accepted 10% threshold for significant hemolytic activity. These results indicate that none of the tested peptides caused notable erythrocyte lysis even at elevated concentrations, suggesting favorable blood compatibility and low cytotoxicity.

Cytotoxicity was evaluated in the NIH-3T3 cell line by the MTT method. The results showed that within the concentration range of 12.5–200 μg/mL, the cell survival rate was generally maintained at more than 70%, and the survival rate was close to 100% at lower concentrations. This further proves that polypeptides have relatively low toxicity to mammals and have potential biomedical application prospects.

Our preliminary results suggest that these four polypeptides may interact with the bacterial cell membrane, disrupt its integrity, and eventually lead to cell death. The minimum inhibitory concentrations (MICs) of the four candidate peptides against *K. pneumoniae*, *E. coli*, and *S. aureus* are summarized in Table 3, with vancomycin serving as the positive control. As shown in Figure 7, vancomycin was used as a positive control drug to evaluate the antimicrobial activity of the newly synthesized peptides. Vancomycin is an antibiotic commonly used in clinical practice. It belongs to the glycopeptide class of antibiotics and is widely applied in the treatment of infections caused by Gram-positive bacteria. The MIC test results showed that the minimum inhibitory concentration of Vancomycin against *S. aureus* was less than 1 μg/mL, demonstrating extremely high antimicrobial activity; The MIC for *K. pneumoniae*

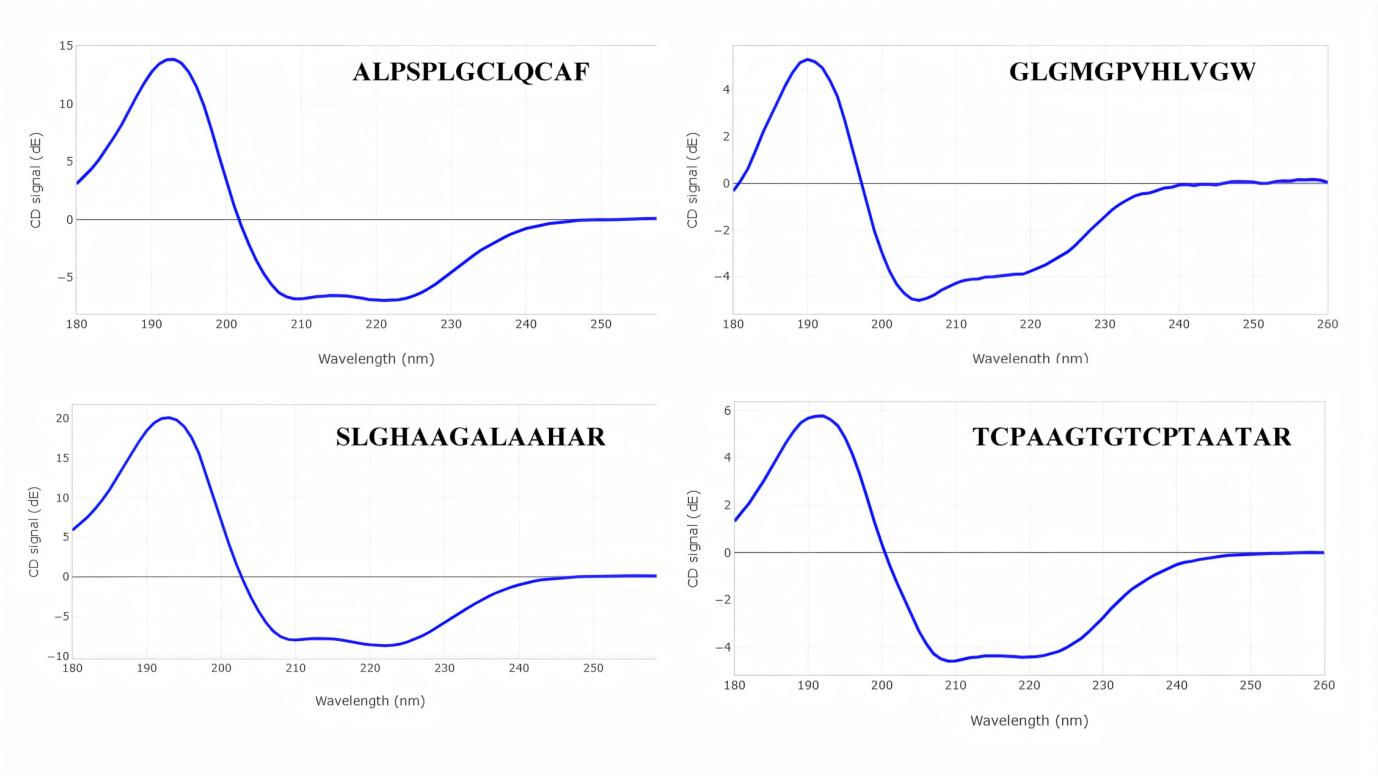

Figure 7. Circular Dichroism (CD) Analysis of Synthetic Peptides.

and *E. coli* was 16 μg/mL respectively, indicating that its antimicrobial ability against some Gram-negative bacteria was relatively weak. This result is in line with the characteristic that Vancomycin is only effective against Gram-positive bacteria, as it cannot penetrate the outer membrane of Gram-negative bacteria.

In contrast, the MIC values of ALPSPLGCLQCAF and TCPAAGTGTCPTAATAR in *S. aureus* were 16 μg/mL and 8 μg/mL, respectively (Table 3), slightly higher than those of Vancomycin, but still showed good antimicrobial activity. Furthermore, although the MIC values (32–64 μg/mL) of these peptides in *K. pneumoniae* and *E. coli* were higher than those in Vancomycin (Table 3), some existing strains have developed resistance to vancomycin at present. Clinical studies have shown that many enterococcal strains and some *Staphylococcus aureus* strains are resistant to vancomycin. When infected with *S. aureus* strains, the therapeutic effect of the standard dose of vancomycin may not be optimal. Even within the safe dose range, it is difficult to achieve the ideal antimicrobial activity [51].

Table 3. The MIC values (μg/mL) of peptides and Vancomycin against *K. pneumoniae*, *E. coli* and *S. aureus*.

Sample	*S. aureus*	*K. pneumoniae*	*E. coli*
ALPSPLGCLQCAF	16	32	32
GLGMGPVHLVGW	128	128	64
SLGHAAGALAAHAR	16	64	32
TCPAAGTGTCPTAATAR	8	64	64
Vancomycin	<1	16	16

3.5 Analysis of molecular dynamics simulation results

To investigate the molecular-level interactions between the selected antimicrobial peptides and bacterial membrane mimics, 500 ns molecular dynamics simulations were performed for peptide 3 and peptide 5 in three distinct membrane systems: POPE:POPG (1:3) representing Gram-positive bacterial membranes, POPE:POPG (2:1) representing Gram-negative bacterial membranes, and DPPC representing mammalian cell membranes. Figures 8 and 9 present representative snapshots captured at different time points during these simulations, illustrating the conformational changes and membrane insertion behavior of peptide 3 and peptide 5, respectively. The following subsections provide quantitative analyses of structural deviations, mass density distributions, compactness parameters, and hydrogen bonding interactions.

3.5.1 Structural Deviations

In the course of molecular dynamics simulations, the structural drift of peptides was used to assess their

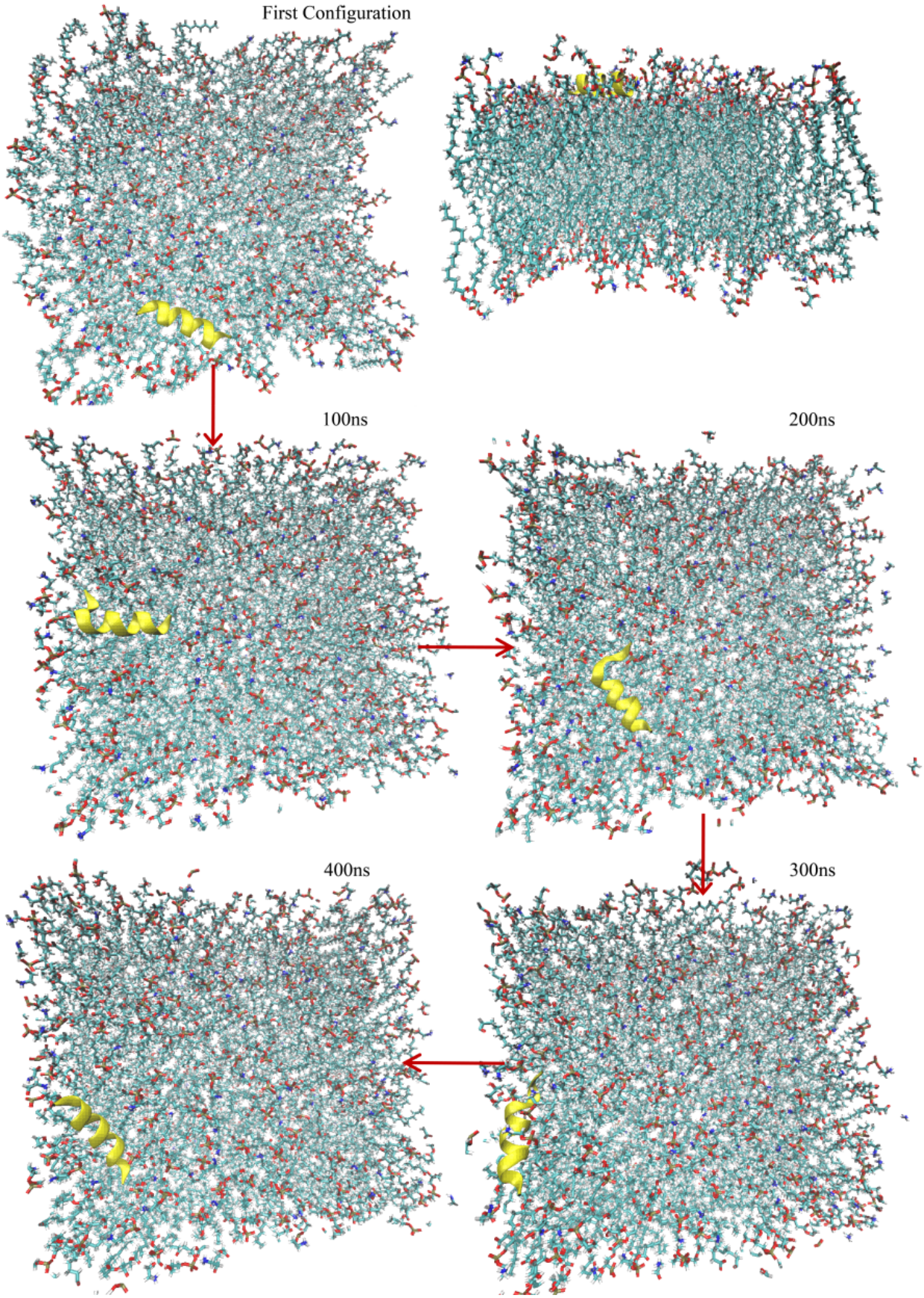

Figure 8. Snapshots of molecular dynamics simulations captured at different time frames, illustrating the conformational changes of the peptide3.

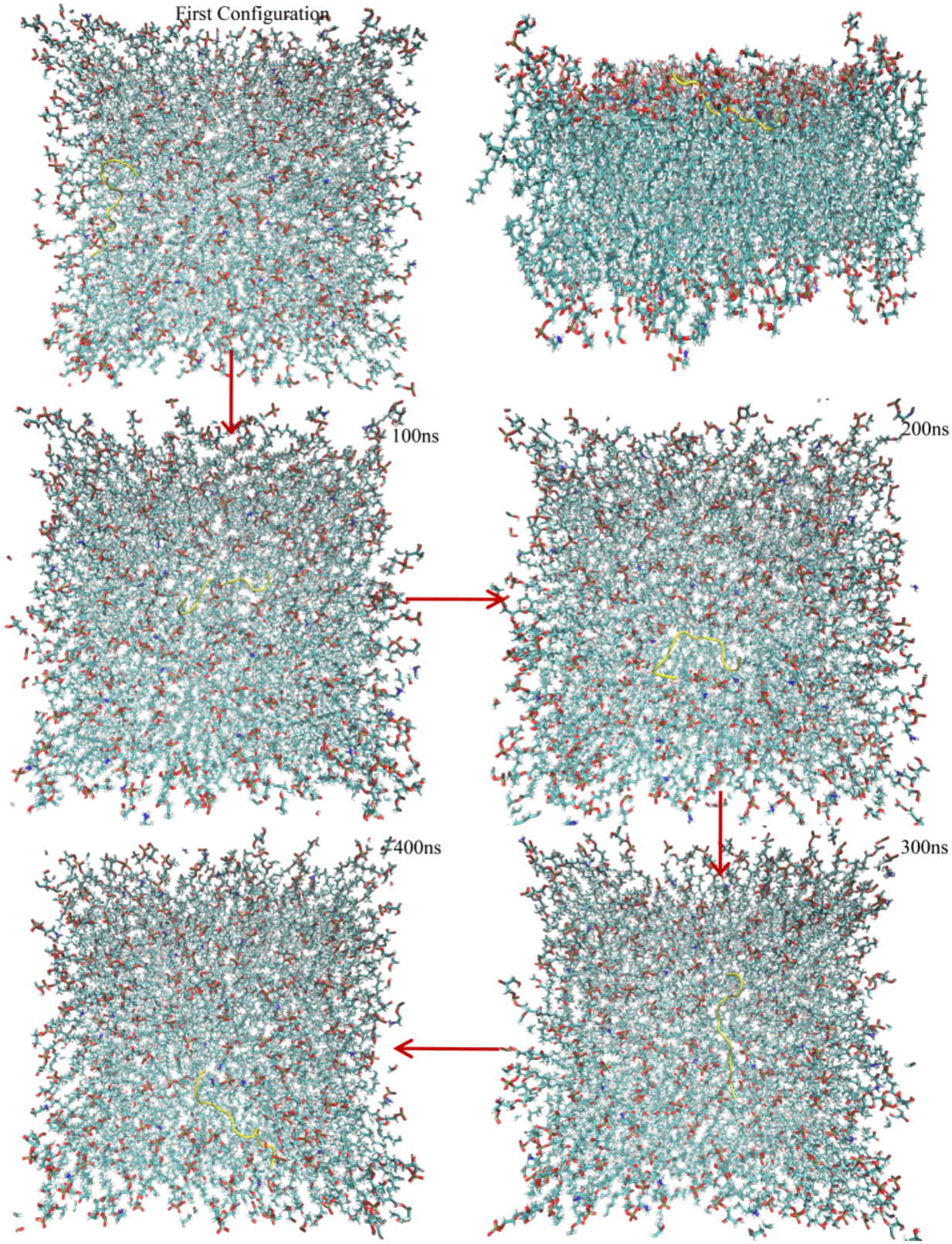

Figure 9. Snapshots of molecular dynamics simulations captured at different time frames, illustrating the conformational changes of the peptide5.

conformational stability and dynamic behavior within different membrane environments. To quantify this drift, we computed the root-mean-square deviation (RMSD) of the Cα atoms of each candidate peptide, taking the initial structure at time $t = 0$ as the reference [46], the result is shown in Figure 10 (A). The analysis shows that the peptide 3 has an average RMSD of 0.5315 nm with a fluctuation range of $\pm$0.060 nm in

the POPE/POPG (1:3) membrane system, indicating relatively stable conformation. In contrast, the peptide exhibits an average RMSD of 0.607 nm with a fluctuation range of ±0.102 nm in the DPPC membrane system, suggesting a less stable conformation with greater fluctuations. Further comparison reveals that the average RMSD in the POPE/POPG (2:1) membrane system is only 0.287 nm, significantly lower than that in the DPPC system, with a fluctuation range of merely ±0.067 nm, demonstrating higher conformational stability. Therefore, it can be concluded that the POPE/POPG membrane environment is more favorable for maintaining the peptide's conformational stability, implying potentially better biological activity in such bacterial membrane environments.

The RMSD variations of peptide 5 in different membrane systems reveal differences in its conformational stability. As shown in the Figure 10 (B), in the POPE/POPG (2:1) membrane system, peptide 5 has an average RMSD of 0.463 nm with a fluctuation range of ±0.075 nm, indicating relatively stable conformation. In the POPE/POPG (1:3) membrane system, the average RMSD is 0.397 nm with a fluctuation range of ±0.107 nm; although the fluctuations are slightly larger, the overall structural changes are gradual and without abrupt jumps, demonstrating better stability. In contrast, in the DPPC membrane system, the average RMSD is 0.476 nm, slightly higher than the other two, with a fluctuation range of ±0.098 nm, indicating a more pronounced deviation of the peptide chain from the reference conformation. These RMSD differences suggest that the peptide exhibits distinct dynamic behaviors throughout the MD simulations in different membrane environments.

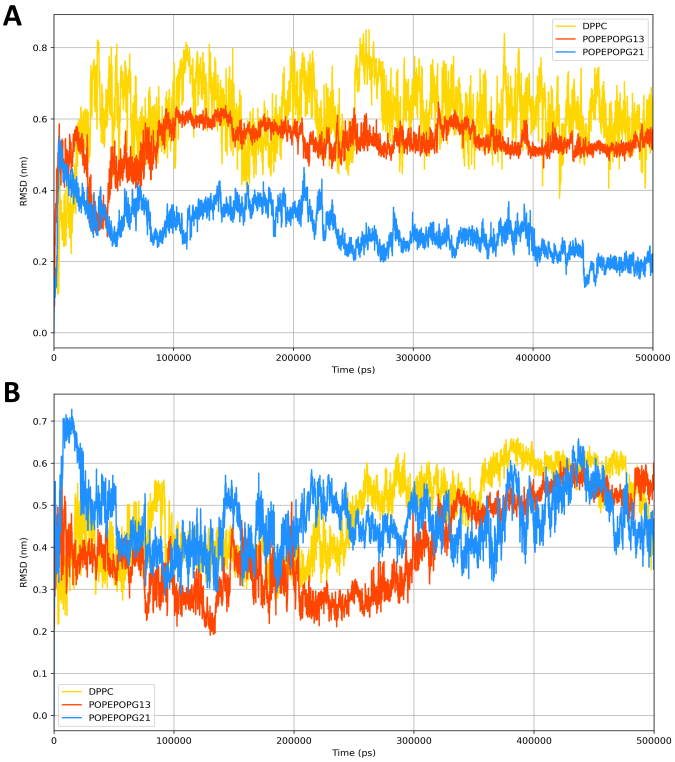

Figure 10. RMSD analysis of antimicrobial peptides in different membrane systems. (A) RMSD of peptide 3 during molecular dynamics simulation in three membrane systems. (B) RMSD of peptide 5 during molecular dynamics simulation in three membrane systems.

3.5.2 Mass Density

The mass density distributions of peptides 3 and 5 across the three membrane systems are presented in Figure 11. The analysis revealed that the localization behavior of peptides in different membrane systems was significantly different. In the POPE/POPG (1:3) membrane system, Peptide 5 primarily localizes near the lipid polar headgroup region, with a moderate degree of insertion but does not penetrate deeply into the hydrophobic tail region Figure 11(A). In contrast, in the POPE/POPG (2:1) membrane system, the mass density profile exhibits a higher and broader peak at around 6 nm, indicating deeper peptide insertion, with some regions possibly extending into the lipid tail region Figure 11(B). This suggests that the membrane interior is more compact, with tighter lipid chain packing, facilitating deeper peptide embedding. In comparison, in the DPPC membrane system, Peptide 5 remains mostly at the membrane surface or polar headgroup region, showing shallow insertion and no significant tendency to enter the hydrophobic core Figure 11(C).

In the POPE/POPG (1:3) membrane system, Peptide 3 primarily distributes near the outer aqueous phase adjacent to the membrane surface, interacting with the membrane interface but without deep penetration Figure 11(D). However, in the POPE/POPG (2:1) membrane system, its insertion is more pronounced, with the distribution curve shifting toward the membrane center, suggesting stronger affinity for this lipid composition Figure 11(E). In the DPPC membrane system, Peptide 3 exhibits a weaker mass density distribution, mainly confined to the aqueous phase or membrane surface Figure 11(F). DPPC (1,2-dipalmitoyl-sn-glycero-3-phosphocholine) is a saturated phospholipid with no double bonds in its fatty acid chains, causing it to form a highly ordered, densely packed gel-phase structure at physiological temperatures. This tight lipid arrangement makes the hydrophobic core region of the membrane particularly

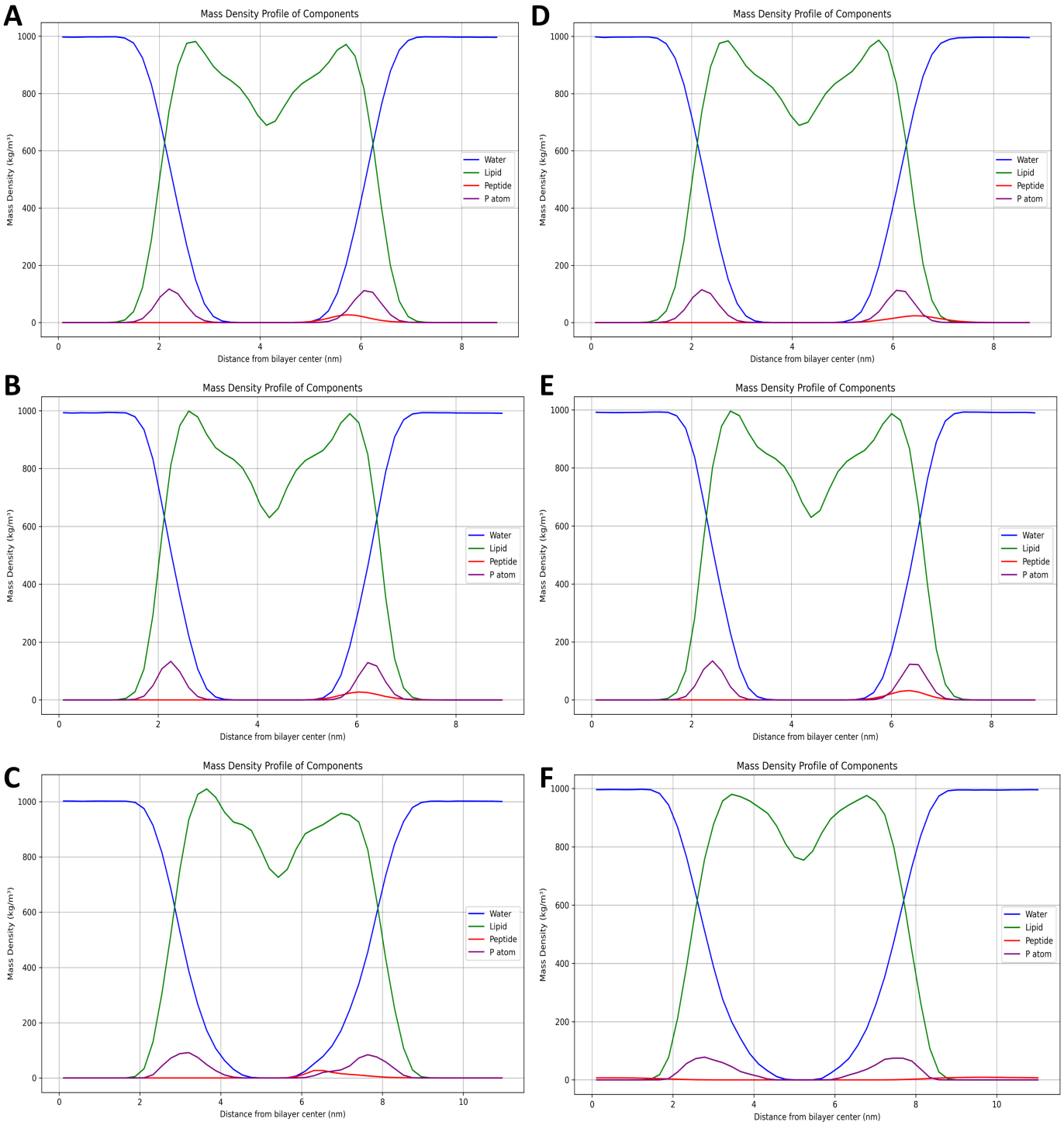

Figure 11. Mass Density Profiles of Peptides 3 and 5 in Different Membrane Systems. (A) Mass Density of Peptide 5 in the POPE:POPG (3:1) Membrane System. (B) Mass Density of Peptide 5 in the POPE:POPG (2:1) Membrane System. (C)Mass Density of Peptide 5 in the DPPC Membrane System. (D) Mass Density of Peptide 3 in the POPE:POPG (3:1) Membrane System. (E) Mass Density of Peptide 3 in the POPE:POPG (2:1) Membrane System. (F) Mass Density of Peptide 3 in the DPPC Membrane System.

resistant to penetration by peptide molecules [52]. This may be due to the highly ordered and dense structure of DPPC bilayers or charge repulsion effects that restrict deeper peptide insertion.

3.5.3 Compactness

SASA (Solvent Accessible Surface Area) analysis further elucidates the exposure levels of hydrophilic/hydrophobic residues and the conformational behavior of peptides in different

membrane environments [53]. SASA is a metric used to evaluate the solvent accessibility of a molecular surface, reflecting the depth of peptide insertion and residue exposure within membrane systems.

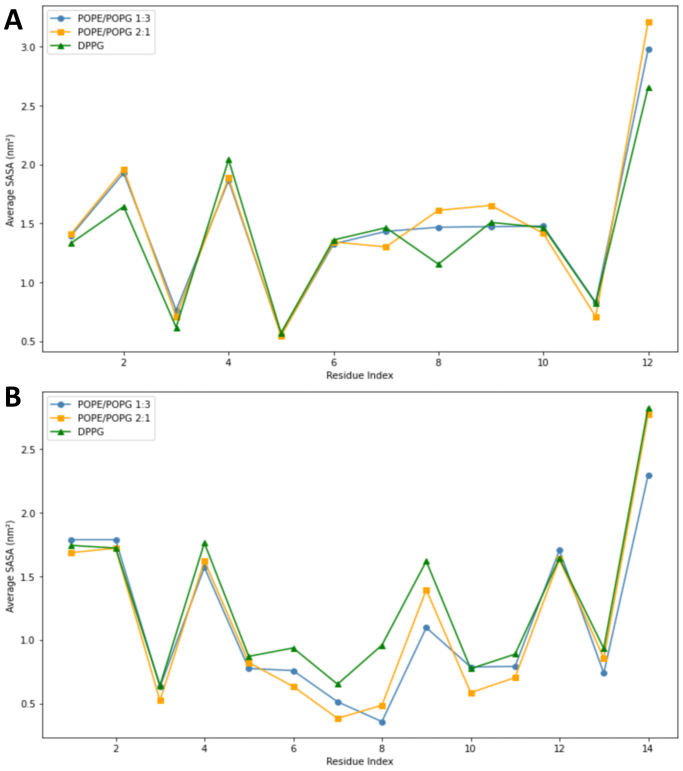

Figure 12. (A) Per-residue SASA of Peptide 5 in different membrane systems. (B) Per-residue SASA of Peptide 3 in different membrane systems.

As shown in Figure 12 (A), for Peptide 5, the residues with the lowest SASA values in all three membrane systems were Gly3, Gly5, and Gly11—all glycine residues. For Peptide 3, the lowest SASA values were observed at Gly3, Ala5, Ala7, Ala10, and Ala11. These residues possess small side-chain volumes and are either weakly hydrophilic or hydrophobic, neutral amino acids. They tend to embed into the hydrophobic core of the membrane, minimizing water contact, which explains their low SASA values [54].

In contrast, Peptide 5 exhibited the highest SASA values at Leu2, Met4, Pro6, Val7, His8, Val10, and Trp12—residues that are predominantly hydrophobic or positively charged polar residues. Their elevated SASA values suggest greater exposure at the membrane surface or membrane-water interface, likely due to a more flexible peptide conformation.For Peptide 3, the highest SASA values were observed at Ser1, Leu2, His4, Leu9, His12, and Arg14, indicating that hydrophilic and positively charged residues (Ser, His, Arg) were highly exposed at the membrane surface. Additionally, hydrophobic residues like Leu also showed relatively high SASA values, suggesting that the peptide likely adopts a loose conformation near the membrane surface rather than deeply inserting into the bilayer.

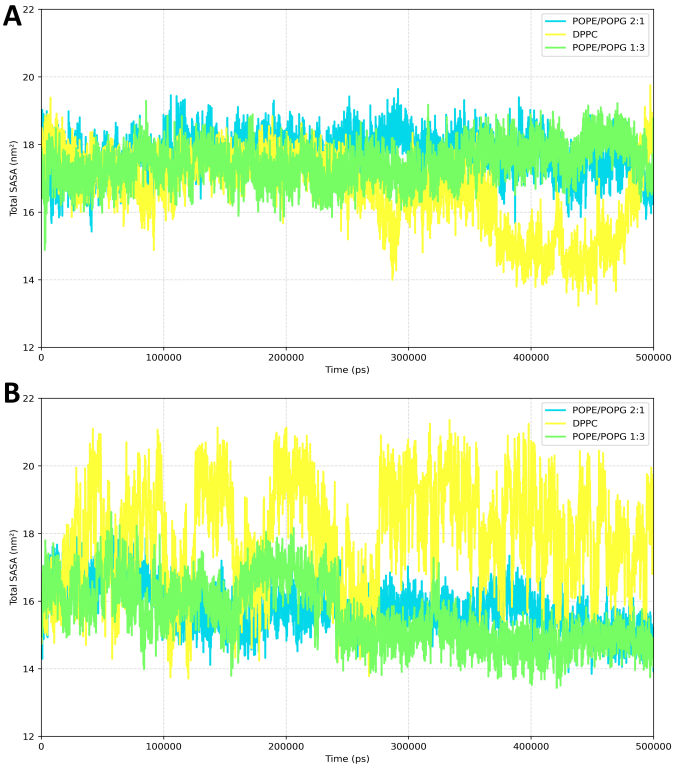

Figure 13. (A) SASA of peptide5 in Different Membrane Systems. (B) SASA of peptide3 in Different Membrane Systems.

Figure 13 illustrates the time-dependent SASA (solvent-accessible surface area) profiles of the peptides in three different membrane systems. In the POPE/POPG (1:3) and POPE/POPG (2:1) bilayers, Peptide 5 exhibits high conformational stability and tends to bind near the membrane surface. The overall SASA values in the POPE/POPG (1:3) system are slightly higher than those in the POPE/POPG (2:1) system, indicating that the former allows Peptide 5 to be more exposed to the solvent.In contrast, in the DPPC membrane, the behavior of Peptide 5 is more complex—its SASA decreases around 350 ns, then rises again—suggesting a dynamic process of insertion, deeper embedding or compaction, followed by re-exposure to the solvent. This greater fluctuation in SASA implies that the interaction between Peptide 5 and the DPPC membrane may be weaker or less stable.For Peptide 3, the SASA values in both POPE/POPG (1:3) and (2:1) membranes remain relatively low, possibly due to its more compact

structure. A slight decreasing trend after 200 ns suggests that the peptide may be inserting deeper into the membrane over time [55].

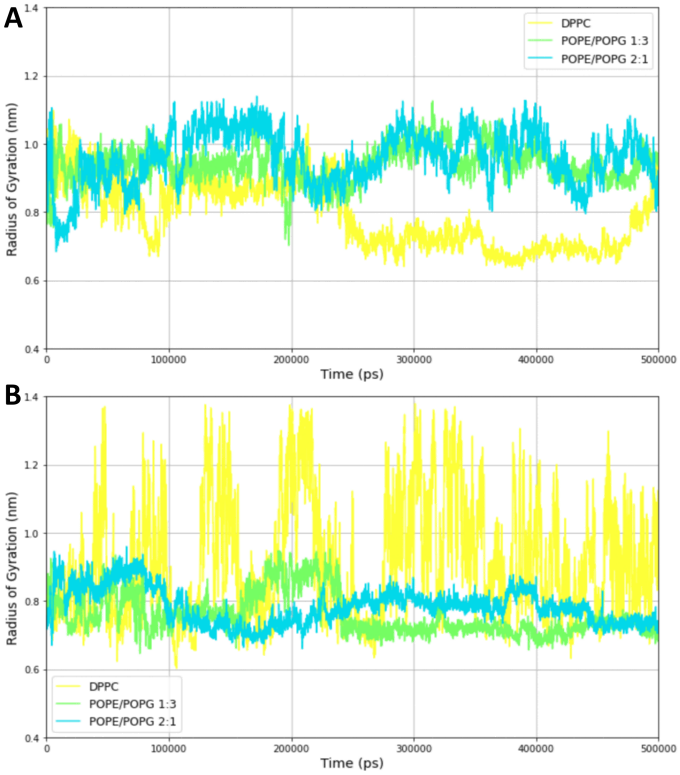

Figure 14. (A) Time evolution of the radius of gyration (Rg) value for peptide5 in both kinds of bilayer. (B) Time evolution of the radius of gyration (Rg) value for peptide3 in both kinds of bilayer.

By calculating the radius of gyration (Rg), the overall properties and behavior of the peptides in the bilayer model can be further understood, such as their compactness and the time required to reach structural and conformational equilibrium. As shown in Figure 14, the Rg values of peptide 5 in the POPE/POPG (1:3) and POPE/POPG (2:1) membrane systems remain relatively stable overall, further indicating that its conformation is stable in these two membrane environments and that the overall structural looseness remains consistent, which aligns with the trends observed in SASA. In contrast, in the DPPC membrane system, the Rg of peptide 5 decreases after 300 ns and begins to increase again around 450 ns, corresponding to the changes in SASA and reflecting a dynamic process of contraction and re-expansion of the peptide in the DPPC membrane.

Compared to peptide 5, peptide 3 shows slightly lower overall Rg values in the POPE/POPG (1:3) and POPE/POPG (2:1) membrane systems, and a slight decreasing trend after 200 ns, indicating a more compact structure and deeper membrane insertion, which is consistent with the gradual decrease in its SASA values. In the DPPC membrane system, however, the Rg values of peptide 3 fluctuate more dramatically, indicating that its overall conformation is highly unstable in this environment. This observation further corresponds to the large variations in SASA, supporting the conclusion that peptide 3 is unable to maintain a stable binding state in the DPPC membrane.

3.5.4 Hydrogen Bond

To gain deeper insights into the binding behavior and dynamic characteristics of peptides in different membrane systems, we further analyzed the temporal evolution of the number of hydrogen bonds formed between peptides and membrane or water molecules during the simulation [56]. As shown in the Figure 15 (A), peptide 5 formed the highest number of hydrogen bonds with membrane molecules in the POPE/POPG (1:3) membrane system. The number of hydrogen bonds showed a gradual increase and eventually stabilized at around 10, indicating progressively enhanced and stable peptide–membrane interactions. In the POPE/POPG (2:1) membrane system, the number of hydrogen bonds was slightly lower, stabilizing around 5. In contrast, in the DPPC membrane system, peptide 5 formed the fewest hydrogen bonds—typically fewer than 5—suggesting weak and dynamically unstable binding in this membrane environment. These results are consistent with trends observed in the SASA and radius of gyration (Rg) analyses, further confirming that peptide 5 preferentially forms stable interactions with the negatively charged POPE/POPG membranes, while exhibiting weaker binding affinity in the neutral DPPC membrane environment, where it is more likely to interact with surrounding water molecules rather than membrane components.

For peptide 3, As shown in the Figure 15 (B), the highest number of hydrogen bonds was observed in the POPE/POPG (2:1) membrane system, ranging between 5 and 15, indicating strong membrane binding capacity. In the POPE/POPG (1:3) membrane, the number of hydrogen bonds was slightly lower, fluctuating between 5 and 10. In the DPPC system, peptide 3 formed the fewest hydrogen bonds, consistently fewer than 5, suggesting an unstable binding behavior and relatively weak membrane affinity in this environment.

Therefore, the hydrogen bond analysis further supports the conclusion that peptides are more

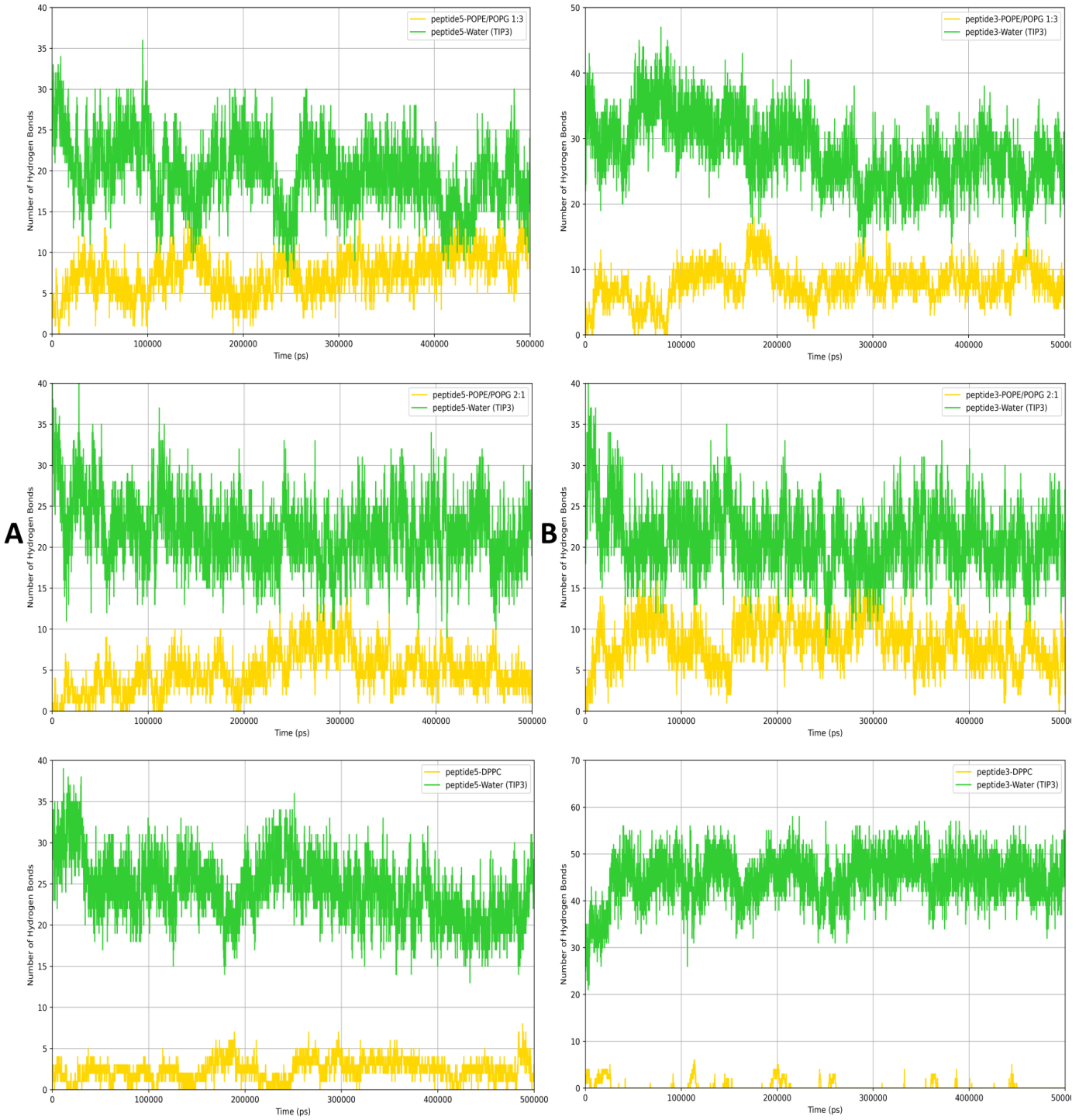

Figure 15. (A) The number of hydrogen bond between the peptide5 and membrane as well as water moleculesin the 500 ns. (B) The number of hydrogen bond between the peptide3 and membrane as well as water moleculesin the 500 ns.

likely to form stable and persistent associations in POPE/POPG membrane environments compared to DPPC membranes.

Based on a comprehensive analysis of multiple molecular dynamics parameters—including solvent-accessible surface area (SASA), radius of gyration (Rg), and the number of hydrogen bonds—of peptide 5 and peptide 3 in three different membrane systems (POPE/POPG 1:3, POPE/POPG 2:1, and DPPC), the following conclusions can be drawn: the composition and surface charge of the membrane significantly influence the binding behavior and conformational stability of the peptides. Overall, the negatively charged POPE/POPG membranes, especially the POPE/POPG (1:3) system,

are more favorable for stable peptide binding. These environments promote deeper insertion of peptides into the hydrophobic core of the membrane while maintaining relatively stable conformations, as reflected by higher hydrogen bond numbers, smaller Rg fluctuations, and corresponding SASA trends.

In contrast, in the neutral DPPC membrane, both peptides exhibit weaker binding abilities and larger conformational fluctuations, as indicated by pronounced changes in SASA and Rg values and fewer hydrogen bonds. This suggests that the peptides tend to remain near the membrane surface or at the membrane–water interface and have difficulty achieving stable insertion into the membrane interior.

Peptide 5 demonstrates stronger membrane affinity and greater conformational flexibility than peptide 3 in all membrane systems, particularly in POPE/POPG membranes, where it binds more tightly, inserts deeper, and maintains higher structural stability. Due to its more compact structure, peptide 3 generally shows lower SASA and Rg values. While it can still form a certain number of hydrogen bonds in POPE/POPG membranes, its stability is markedly reduced in the DPPC environment.

In summary, POPE/POPG membranes—especially at the 1:3 ratio—are more effective in enhancing peptide–membrane interactions. Peptide 5 shows greater potential as a membrane-active peptide for further functional or mechanistic studies.

4 Conclusion

In this study, novel antimicrobial peptides were screened from the metagenomic sequences of soil manure in breeding farms through bioinformatics methods. A deep learning AMP-CLIP classification model was adopted, which integrates four deep learning frameworks: GCN, LSTM, Attention, and Transformer, Designed specifically to improve the recognition accuracy of AMPs, it fully mines the characteristic information in peptide sequences. It is significantly superior to existing methods in terms of accuracy and performance, and is particularly suitable for high-throughput mining of potential antimicrobial peptides from complex metagenomic data. However, the multi-layer structure of the model leads to high consumption of computing resources, high training and prediction costs, and is not conducive to large-scale deployment. Finally, four novel antimicrobial peptides with high scores that were non-toxic and non-hemolytic were screened out. Preliminary experiments indicated that they had inhibitory effects on *K.peneumoniae*, *E. coli*, and *S. aureus*.

Our preliminary research indicates that peptide 5 and peptide 3 target bacterial cell membranes to exert their antimicrobial effects. A membrane-peptide complex system was constructed and subjected to 500ns kinetic simulation using GROMACS. The simulation results show that the selected peptides exhibit deeper insertions, more stable conformations, and stronger hydrogen bond interactions in negatively charged POPE/POPG membranes. It indicates that it may exert antimicrobial activity by targeting bacterial membranes; In the DPPC membrane system of animal model membranes, the binding of peptides to the membrane is relatively weak, mainly remaining on the membrane surface or in the aqueous phase region, and the stability is poor. These peptides can be effectively used in food processing and they can significantly reduce the risk of pathogen growth. Then, its antimicrobial activity may be affected by the complex environmental conditions in food, so further research is needed to optimize its application potential in the food industry.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The author declares no conflicts of interest.

AI Use Statement

The author declares that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

This study involving human blood samples was reviewed and approved by the Ethics Committee of Beijing Technology and Business University. Written informed consent was obtained from all blood donors. The study was conducted in accordance with the Declaration of Helsinki and relevant Chinese ethical regulations for biomedical research. The NIH-3T3 cell line used in the cytotoxicity test is a standard commercial immortalized cell line and did not require separate ethical approval.

References

[1] Cheng, Q., Zeng, P., Chi Chan, E. W., & Chen, S. (2022). Development of Peptide-based Metallo-β-lactamase Inhibitors as a New Strategy to Combat Antimicrobial Resistance: A Mini-review. *Current Pharmaceutical Design*, *28*(44), 3538–3545. [CrossRef]

[2] Morris, C. F. (2019). The antimicrobial properties of the puroindolines, a review. *World Journal of Microbiology and Biotechnology*, *35*(6), 86. [CrossRef]

[3] Bai, H., He, L. Y., Wu, D. L., Gao, F. Z., Zhang, M., Zou, H. Y., ... & Ying, G. G. (2022). Spread of airborne antibiotic resistance from animal farms to the environment: dispersal pattern and exposure risk. *Environment international*, *158*, 106927. [CrossRef]

[4] Furlan, J. P. R., Lopes, R., Ramos, M. S., Rosa, R. D., Savazzi, E. A., & Stehling, E. G. (2024). The detection of KPC-2, NDM-1, and VIM-2 carbapenemases in international clones isolated from fresh vegetables highlights an emerging food safety issue. *International Journal of Food Microbiology*, *420*, 110765. [CrossRef]

[5] Xu, C., Han, A., Tian, Y., Fu, M., Gao, X., & Liu, H. (2025). Based on computer simulation and experimental verification: mining and characterizing novel antimicrobial peptides from soil microbiome. *Food Chemistry*, *467*, 142275. [CrossRef]

[6] Pashaei, F., Bevalian, P., Akbari, R., & Pooshang Bagheri, K. (2019). Single dose eradication of extensively drug resistant Acinetobacter spp. In a mouse model of burn infection by melittin antimicrobial peptide. *Microbial Pathogenesis*, *127*, 60–69. [CrossRef]

[7] Memariani, H., Shahbazzadeh, D., Ranjbar, R., Behdani, M., Memariani, M., & Pooshang Bagheri, K. (2017). Design and characterization of short hybrid antimicrobial peptides from pEM-2, mastoparan-VT 1, and mastoparan-B. *Chemical Biology & Drug Design*, *89*(3), 327–338. [CrossRef]

[8] Lohner, K. (2009). New strategies for novel antibiotics: peptides targeting bacterial cell membranes. *General Physiology and Biophysics*, *28*(2), 105–116. [CrossRef]

[9] Santos-Júnior, C. D., Pan, S., Zhao, X. M., & Coelho, L. P. (2020). Macrel: antimicrobial peptide screening in genomes and metagenomes. *PeerJ*, *8*, e10555. [CrossRef]

[10] Akbar, S., Raza, A., & Zou, Q. (2024). Deepstacked-AVPs: predicting antiviral peptides using tri-segment evolutionary profile and word embedding based multi-perspective features with deep stacking model. *BMC Bioinformatics*, *25*(1), 102. [CrossRef]

[11] Li, Z. R., Lin, H. H., Han, L. Y., Jiang, L., Chen, X., & Chen, Y. Z. (2006). PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic acids research*, *34*(suppl_2), W32-W37. [CrossRef]

[12] Waghu, F. H., Barai, R. S., Gurung, P., & Idicula-Thomas, S. (2016). CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Research*, *44*(D1), D1094–D1097. [CrossRef]

[13] Li, C., Zou, Q., Jia, C., & Zheng, J. (2024). AMPpred-MFA: An Interpretable Antimicrobial Peptide Predictor with a Stacking Architecture, Multiple Features, and Multihead Attention. *Journal of Chemical Information and Modeling*, *64*(7), 2393–2404. [CrossRef]

[14] Veltri, D., Kamath, U., & Shehu, A. (2018). Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, *34*(16), 2740–2747. [CrossRef]

[15] Lee, H., Lee, S., Lee, I., & Nam, H. (2023). AMP-BERT: Prediction of antimicrobial peptide function based on a BERT model. *Protein Science*, *32*(1), e4529. [CrossRef]

[16] Yang, S., Yang, Z., & Ni, X. (2023). AMPFinder: A computational model to identify antimicrobial peptides and their functions based on sequence-derived information. *Analytical Biochemistry*, *673*, 115196. [CrossRef]

[17] Teufel, F., Refsgaard, J. C., Madsen, C. T., & Nielsen, H. (2023). DeepPeptide predicts cleaved peptides in proteins using conditional random fields. *Bioinformatics*, *39*(10), btad616. [CrossRef]

[18] Wang, Y., Zhao, T., Wei, D., Strandberg, E., Ulrich, A. S., & Ulmschneider, J. P. (2014). How reliable are molecular dynamics simulations of membrane active antimicrobial peptides? *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1838*(9), 2280–2288. [CrossRef]

[19] Nguyen, L. T., Haney, E. F., & Vogel, H. J. (2011). The expanding scope of antimicrobial peptide structures and their modes of action. *Trends in Biotechnology*, *29*(9), 464–472. [CrossRef]

[20] Mahmoodzadeh, A., Zarrinnahad, H., Bagheri, K. P., Moradia, A., & Shahbazzadeh, D. (2015). First report on the isolation of melittin from Iranian honey bee venom and evaluation of its toxicity on gastric cancer AGS cells. *Journal of the Chinese Medical Association*, *78*(10), 574–583. [CrossRef]

[21] Harris, F., Dennison, S. R., Singh, J., & Phoenix, D. A. (2013). On the selectivity and efficacy of defense peptides with respect to cancer cells. *Medicinal Research Reviews*, *33*(1), 190–234. [CrossRef]

[22] Navgire, G. S., Goel, N., Sawhney, G., Sharma, M., Kaushik, P., Mohanta, Y. K., ... & Al-Harrasi, A. (2022). Analysis and Interpretation of metagenomics data: an approach. *Biological Procedures Online*, *24*(1), 18. [CrossRef]

[23] Han, A., Liu, H., Dai, Y., Chen, W., & Zhao, C. (2024). Screening of umami peptides from fermented grains by machine learning, molecular docking and molecular dynamics simulation. *Food Bioscience*, *62*, 105536. [CrossRef]

[24] Wang, P., Ge, R., Liu, L., Xiao, X., Li, Y., & Cai, Y. (2017).

Multi-label learning for predicting the activities of antimicrobial peptides. *Scientific reports, 7*(1), 2202. [CrossRef]

[25] Wang, G., Li, X., & Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Research*, *44*(D1), D1087–D1093. [CrossRef]

[26] Kang, X., Dong, F., Shi, C., Liu, S., Sun, J., Chen, J., ... & Zheng, H. (2019). DRAMP 2.0, an updated data repository of antimicrobial peptides. *Scientific data, 6*(1), 148. [CrossRef]

[27] Wang, B., Lin, P., Zhong, Y., Tan, X., Shen, Y., Huang, Y., ... & Wu, Y. (2025). Explainable deep learning and virtual evolution identifies antimicrobial peptides with activity against multidrug-resistant human pathogens. *Nature microbiology, 10*(2), 332-347. [CrossRef]

[28] Porto, W. F., Pires, A. S., & Franco, O. L. (2017). Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnology Advances*, *35*(3), 337–349. [CrossRef]

[29] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*.

[30] Azodi, C. B., Tang, J., & Shiu, S. H. (2020). Opening the black box: interpretable machine learning for geneticists. *Trends in genetics, 36*(6), 442-455. [CrossRef]

[31] Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society*, *85*(14), 2149–2154. [CrossRef]

[32] Micsonai, A., Wien, F., Kernya, L., Lee, Y. H., Goto, Y., Réfrégiers, M., & Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences, 112*(24), E3095-E3103. [CrossRef]

[33] Domon, B., & Aebersold, R. (2006). Mass Spectrometry and Protein Analysis. *Science*, *312*(5771), 212–217. [CrossRef]

[34] Mohamed, M. F., Abdelkhalek, A., & Seleem, M. N. (2016). Evaluation of short synthetic antimicrobial peptides for treatment of drug-resistant and intracellular Staphylococcus aureus. *Scientific Reports*, *6*(1), 29707. [CrossRef]

[35] Das, P., Sercu, T., Wadhawan, K., Padhi, I., Gehrmann, S., Cipcigan, F., ... & Mojsilovic, A. (2021). Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering, 5*(6), 613-623. [CrossRef]

[36] Murzyn, K., Róg, T., & Pasenkiewicz-Gierula, M. (2005). Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophysical journal, 88*(2), 1091-1103. [CrossRef]

[37] Yusuf, M., Destiarani, W., Firdaus, A. R. R., & Rohman, F. (2022). Residual Interactions of LL-37 with POPC and POPE:POPG Bilayer Model Studied by All-Atom Molecular Dynamics Simulation. *International Journal of Molecular Sciences*, *23*(21), 13413. [CrossRef]

[38] Travkova, O. G., Moehwald, H., & Brezesinski, G. (2017). The interaction of antimicrobial peptides with membranes. *Advances in colloid and interface science, 247*, 521-532. [CrossRef]

[39] Jo, S., Lim, J. B., Klauda, J. B., & Im, W. (2009). CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal*, *97*(1), 50–58. [CrossRef]

[40] Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., & Im, W. (2014). CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry*, *35*(27), 1997–2004. [CrossRef]

[41] Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., ... & Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation, 12*(1), 405-413. [CrossRef]

[42] Venable, R. M., Sodt, A. J., Rogaski, B., Rui, H., Hatcher, E., MacKerell, A. D., ... & Klauda, J. B. (2014). CHARMM all-atom additive force field for sphingomyelin: elucidation of hydrogen bonding and of positive curvature. *Biophysical journal, 107*(1), 134-145. [CrossRef]

[43] Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., ... & Pastor, R. W. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B, 114*(23), 7830-7843. [CrossRef]

[44] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics, 79*(2), 926-935. [CrossRef]

[45] Jorgensen, W. L. (1982). Revised TIPS for simulations of liquid water and aqueous solutions. *The Journal of chemical physics, 77*(8), 4156-4163. [CrossRef]

[46] Aghazadeh, H., Ganjali Koli, M., Ranjbar, R., & Pooshang Bagheri, K. (2020). Interactions of GF-17 derived from LL-37 antimicrobial peptide with bacterial membranes: a molecular dynamics simulation study. *Journal of Computer-Aided Molecular Design, 34*(12), 1261-1273. [CrossRef]

[47] Parrinello, M., & Rahman, A. (1980). Crystal structure and pair potentials: A molecular-dynamics study. *Physical review letters, 45*(14), 1196. [CrossRef]

[48] Hess, B., Bekker, H., Berendsen, H. J., & Fraaije, J. G. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry, 18*(12), 1463-1472. [CrossRef]

[49] Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation, 4*(3), 435-447. [CrossRef]

[50] Chen, C. H., & Lu, T. K. (2020). Development and Challenges of Antimicrobial Peptides for Therapeutic Applications. *Antibiotics*, *9*(1), 24. [CrossRef]

[51] Mei, H., Wang, J., Che, H., Wang, R., & Cai, Y. (2019). The clinical efficacy and safety of vancomycin loading dose: A systematic review and meta-analysis. *Medicine, 98*(43), e17639. [CrossRef]

[52] Wang, B., Zhang, J., Zhang, Y., & Wang, Y. (2018). The penetration of a charged peptide across a membrane under an external electric field: a coarse-grained molecular dynamics simulation. *RSC Advances*, *8*(72), 41517–41525. [CrossRef]

[53] Stock, P., Monroe, J. I., Utzig, T., Smith, D. J., Shell, M. S., & Valtiner, M. (2017). Unraveling Hydrophobic Interactions at the Molecular Scale Using Force Spectroscopy and Molecular Dynamics Simulations. *ACS Nano*, *11*(3), 2586–2597. [CrossRef]

[54] Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology, 157*(1), 105-132. [CrossRef]

[55] Lee, M. T., Hung, W. C., Chen, F. Y., & Huang, H. W. (2008). Mechanism and kinetics of pore formation in membranes by water-soluble amphipathic peptides. *Proceedings of the National Academy of Sciences, 105*(13), 5087-5092. [CrossRef]

[56] Guerrero, M., Ayala, N., Rafael, D., Andrade, F., Marican, A., Vijayakumar, S., & Durán-Lara, E. F. (2025). Hydrogel-antimicrobial peptide association: A novel and promising strategy to combat resistant infections. *Colloids and Surfaces B: Biointerfaces, 247*, 114451. [CrossRef]