



OPEN ACCESS

EDITED BY

Guangshun Wang,
University of Nebraska Medical Center,
United States

REVIEWED BY

Monique L. Van Hoek,
George Mason University, United States
Cesar de la Fuente-Nunez,
University of Pennsylvania, United States

*CORRESPONDENCE

Jung-Suk Sung
sungjs@dongguk.edu
Wonhee Jang
wany@dongguk.edu

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to
Antimicrobials, Resistance and
Chemotherapy,
a section of the journal
Frontiers in Microbiology

RECEIVED 21 June 2022

ACCEPTED 27 July 2022

PUBLISHED 24 August 2022

CITATION

Lee B, Shin MK, Yoo JS, Jang W and Sung J-S (2022) Identifying novel antimicrobial peptides from venom gland of spider *Pardosa astrigera* by deep multi-task learning.
Front. Microbiol. 13:971503.
doi: 10.3389/fmicb.2022.971503

COPYRIGHT

© 2022 Lee, Shin, Yoo, Jang and Sung. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identifying novel antimicrobial peptides from venom gland of spider *Pardosa astrigera* by deep multi-task learning

Byungjo Lee^{1†}, Min Kyoung Shin^{1†}, Jung Sun Yoo²,
Wonhee Jang^{1*} and Jung-Suk Sung^{1*}

¹Department of Life Science, Dongguk University-Seoul, Goyang-si, South Korea, ²Animal Resources Division, National Institute of Biological Resources, Incheon, South Korea

Antimicrobial peptides (AMPs) show promises as valuable compounds for developing therapeutic agents to control the worldwide health threat posed by the increasing prevalence of antibiotic-resistant bacteria. Animal venom can be a useful source for screening AMPs due to its various bioactive components. Here, the deep learning model was developed to predict species-specific antimicrobial activity. To overcome the data deficiency, a multi-task learning method was implemented, achieving *F1* scores of 0.818, 0.696, 0.814, 0.787, and 0.719 for *Bacillus subtilis*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Staphylococcus epidermidis*, respectively. Peptides PA-Full and PA-Win were identified from the model using different inputs of full and partial sequences, broadening the application of transcriptome data of the spider *Pardosa astrigera*. Two peptides exhibited strong antimicrobial activity against all five strains along with cytocompatibility. Our approach enables excavating AMPs with high potency, which can be expanded into the fields of biology to address data insufficiency.

KEYWORDS

antimicrobial peptide, deep learning, multi-task learning, species-specific prediction, spider venom gland transcriptome

Introduction

The overuse and the misuse of antibiotics have contributed to the increased selective pressure for antibiotic resistance (Cantón and Morosini, 2011). As a result, the emergence and spread of antibiotic-resistant bacterial strains have risen dramatically during recent years and caused at least 35,000 deaths in the United States, according to a 2019 report (Centres for Disease Control and Prevention, 2019). As the widespread of antibiotic-resistant bacteria has become a serious medical threat, there is an urgent need for discovering novel antimicrobial reagents that can directly contribute to the development of next-generation antibiotics. Among various bioactive molecules, peptides are advantageous for therapeutic applications because of their distinctive characteristics and

physiological roles, such as immunomodulators, hormones, ligands, and signals (Cunha et al., 2008; Posner and Laporte, 2010; Krumm and Grisshammer, 2015). The intrinsic diversity in the sequence of peptides leads to a broad spectrum of molecular targets with high specificity and selectivity (Tsomaia, 2015). In addition, peptides have low cytotoxicity, low tissue accumulation, and cost-effectiveness compared with small molecules and recombinant proteins (Recio et al., 2017).

Antimicrobial peptides (AMPs) play a key role in the innate immune system, contributing to the defense against pathogens like bacteria, fungi, parasites, and viruses and exhibiting immunomodulatory activities (Yeung et al., 2011; Van Dijk et al., 2016; Mookherjee et al., 2020). Generally, AMPs are less than 50 amino acids (AAs) in length with hydrophobic residues and high positive net charges, leading to interactions with pathogen membranes and intracellular molecules (Huan et al., 2020). By virtue of their non-specific and rapid killing mechanisms against pathogens, AMPs are a promising novel class of antibiotics for preventing and combating antibiotic-resistant bacteria (Nuti et al., 2017; Lei et al., 2019). AMPs are found in almost every organism and are highly abundant in animal venoms. For example, in the antimicrobial peptide database 3 (APD3), 153 AMPs originated from the venom of various species, including spider (Wang et al., 2016). Spiders are among the most successfully evolved taxa of venomous animals, which led to the possession of venom with remarkably diverse mixtures comprising thousands of different components (Langenegger et al., 2019). Therefore, spider venom could be a useful source of novel AMPs (Budnik et al., 2004; Wang and Wang, 2016; Wadhvani et al., 2021).

The increased utilization of deep learning in biological and medical research enabled high-throughput analyses, aided by exponentially accumulating biological datasets (Schmidt and Hildebrandt, 2021). Various deep learning models have been developed for the prediction of miRNA targets, the identification of biological signals, the inference of gene expressions, and the prediction of DNA functions based on large amounts of genomic data (Chen et al., 2016; Lee et al., 2016; Quang and Xie, 2016; Tan et al., 2017). Although deep learning techniques have also been applied in neurotoxic, antibacterial, antifungal, and anticancer peptides discovery, the expansion of similar approaches to some fields is challenging due to the data limitations (Fjell et al., 2009; Lata et al., 2010; Waghu et al., 2016; Meher et al., 2017; Ashfaq et al., 2021; Charoenkwan et al., 2021; Lee et al., 2021). The diversity of bacterial genomes led to distinctive bioactivities of individual strains, as well as numerous defense mechanisms toward antibacterial agents (Zhou et al., 2015). Even though it is not our intention to diminish the importance of developing narrow-spectrum antibiotics, each type of pre-existing method may predict AMPs working on only one or a few types of bacterial species. Thus, we tried to build a generalized model that can predict AMPs working on multiple species to suggest reliable candidates of novel antibiotics.

In this study, an accurate deep learning model for predicting antimicrobial activities against five bacterial strains was developed

via multi-task learning (MTL), a method that can overcome the lack of data by improving generalization (Wang et al., 2017; Figure 1A). We curated the dataset with AMP and non-AMP data for model training and testing. Due to the extreme data limitation when training the model, prediction targets were restricted to five bacterial species, *Bacillus subtilis* (*B. subtilis*), *Escherichia coli* (*E. coli*), *Pseudomonas aeruginosa* (*P. aeruginosa*), *Staphylococcus aureus* (*S. aureus*), and *Staphylococcus epidermidis* (*S. epidermidis*). The MTL was applied for model training and was compared with the single-task learning (STL) application. The implementation of the MTL resulted in the best-performing model, which was used for screening potential AMPs from the transcriptome of the spider *Pardosa astrigera* (*P. astrigera*). Two spider-derived peptides, PA-Full and PA-Win, were selected and were verified of their antibacterial activities by functional evaluation. Notably, PA-Win was shown to be a strong AMP with low cytotoxicity in mammalian cells. Our results demonstrated that the deep learning model trained by the MTL achieved an improved prediction performance, overcame the data deficiency, and predicted novel AMPs with high cytocompatibility compared with the application of the STL.

Materials and methods

Dataset preparation

The AMP data for model training were collected from the database of antimicrobial activity and structure of peptides (DBAASP; Pirtskhalava et al., 2016). The dataset contains 14,708 monomer sequences of peptides with known minimum inhibitory concentration (MIC) values against diverse microbial species. The MIC values in the dataset expressed as micromolar (μM) were converted to microgram per milliliter ($\mu\text{g/ml}$) for consistency. For peptides that had multiple MIC test results in an identical species, the average value of MIC was calculated and used. Among the AMP data, several peptides had high MIC values that indicated extremely low antibacterial activity. Thus, we filtered the AMPs by a cutoff value of 267.7 $\mu\text{g/ml}$, which is the lower 95% range of MIC values from the entire AMP data. We obtained the non-AMP data from UniProt using the keyword "NOT antimicrobial" (Bateman et al., 2019).

There were many peptide sequences showing high similarity with other peptides and the peptide sequence redundancy can interfere with model training and performance evaluation. Therefore, the redundancy between training and test datasets as well as within the sequences in the test dataset were controlled by the CD-HIT program with a cutoff value of sequence identity threshold <0.5 to evaluate the model performance accurately (Fu et al., 2012). For the training data, the cutoff value of 0.9 was used to increase the data for model training. We selected five target bacterial species for prediction based on the amount of the AMP sequences that was greater than 190, of which the positive data for respective species were divided into train and test datasets. The

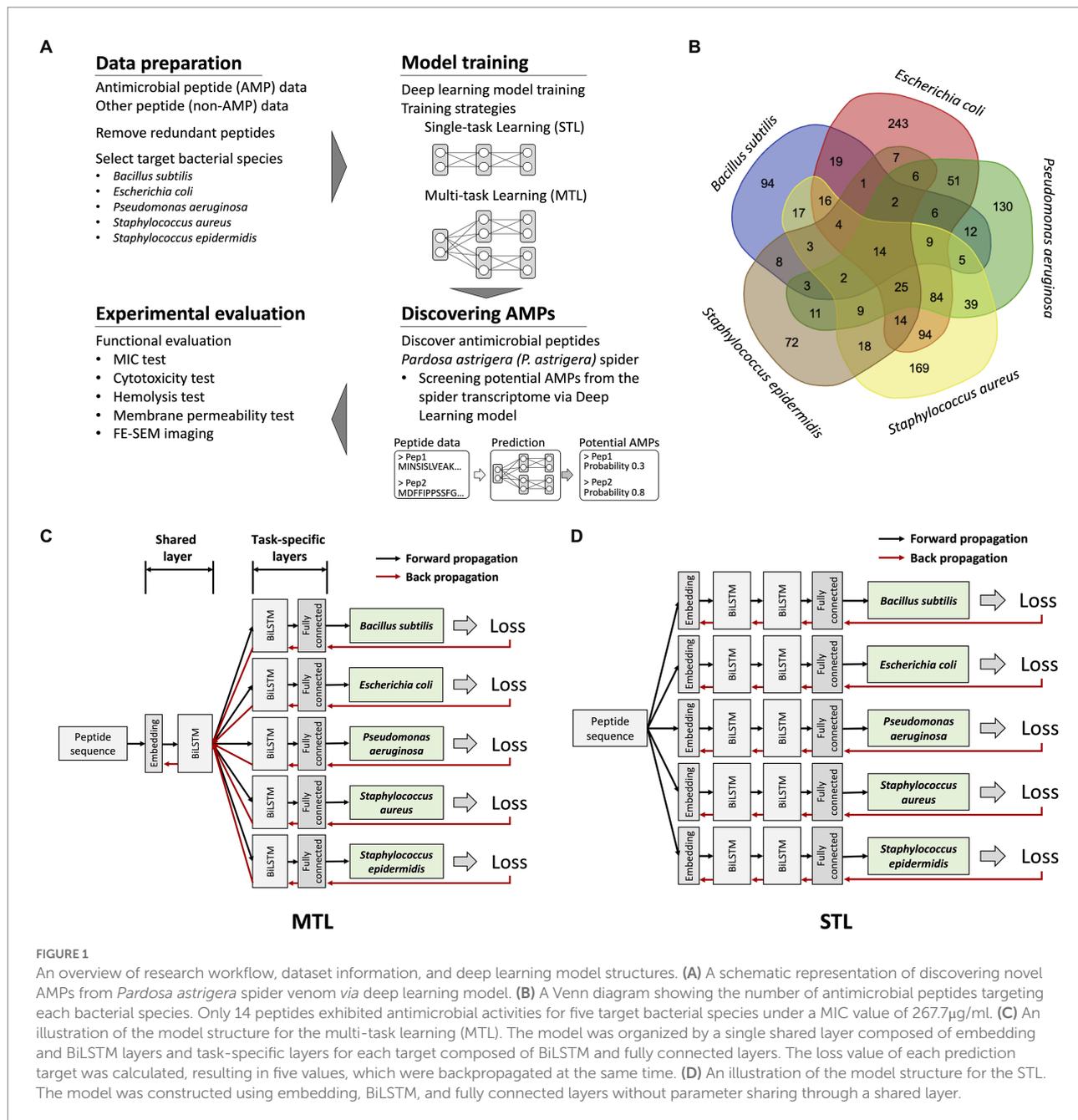


FIGURE 1

An overview of research workflow, dataset information, and deep learning model structures. (A) A schematic representation of discovering novel AMPs from *Pardosa astrigera* spider venom via deep learning model. (B) A Venn diagram showing the number of antimicrobial peptides targeting each bacterial species. Only 14 peptides exhibited antimicrobial activities for five target bacterial species under a MIC value of 267.7 μg/ml. (C) An illustration of the model structure for the multi-task learning (MTL). The model was organized by a single shared layer composed of embedding and BiLSTM layers and task-specific layers for each target composed of BiLSTM and fully connected layers. The loss value of each prediction target was calculated, resulting in five values, which were backpropagated at the same time. (D) An illustration of the model structure for the STL. The model was constructed using embedding, BiLSTM, and fully connected layers without parameter sharing through a shared layer.

non-AMP data possessed more diverse functionality than the AMP data, so we prepared two types of datasets with ratios of 1:1 and 1:3 between the non-AMP and AMP data, and test datasets were organized with the ratio 1:10 (Table 1).

Prediction model structure and training

There are various types of deep learning networks, among which a recurrent neural network (RNN)-based model is known as an effective way to process sequential data. Peptide data can be considered sequential data because the function of a peptide is

determined by a sequential chain of AA residues. Thus, we organized the model structure based on long-term short memory (LSTM), a type of RNN that can tackle the vanishing or explosion gradient problem of RNN (Hochreiter and Schmidhuber, 1997). In addition, bidirectional long short-term memory (BiLSTM) was implemented to reflect both the forward and backward direction of the peptide in the model prediction.

For model training, the MTL and STL were applied, respectively. In the MTL, a shared layer was connected to five different task-specific layers to predict the antimicrobial activity for each target bacterial species. The model architecture of the STL was identical to that of MTL but without a sharing layer with other

TABLE 1 The number of peptide sequences used for model training and testing.

Target species	Training dataset (1:1)		Training dataset (1:3)		Test dataset	
	AMP	Non-AMP	AMP	Non-AMP	AMP	Non-AMP
<i>Bacillus subtilis</i>	155	155	155	465	60	600
<i>Escherichia coli</i>	435	435	435	1,305	160	1,600
<i>Pseudomonas aeruginosa</i>	301	301	301	903	107	1,070
<i>Staphylococcus aureus</i>	380	380	380	1,140	142	1,420
<i>Staphylococcus epidermidis</i>	145	145	145	435	54	540

prediction targets. The configuration of the model architectures was organized with varying BiLSTM node sizes (Supplementary Table 1). We applied hyperparameters for model training with the Adam optimizer with a learning rate of 1E-4 and a maximum training epoch of 500 (Kingma and Ba, 2014).

The performances of the trained models were measured by precision, recall, *F1* score, and Matthews correlation coefficient (MCC) as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$F1 \text{ score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}$$

where TP stands for a true positive number, TN for a true negative number, FP for a false positive number, and FN for a false negative number. We selected the best-performing model in the MTL and STL, respectively, according to the highest *F1* score.

Discovering novel AMPs from *Pardosa astrigera* spider transcriptome data

The transcriptome of the venom gland of *P. astrigera* spider obtained from the previous study was utilized (Shin et al., 2020). The unigenes were processed for abundance estimation and open reading frame (ORF) region prediction by RSEM and

TransDecoder program, respectively (Li and Dewey, 2011). The predicted coding regions was translated into AA residues, and the translated coding regions were used for discovering novel AMPs by deep learning model.

The translated sequences were prepared for two types of input data, using full sequence under length of 50 AAs and partial sequence of the full sequence by sliding window. The sliding window truncated the sequence for desired length and step size from the original sequence, where the truncates were generated with 20-mer length with a step size of five AA in this study. Two types of prepared input data were fed into the selected best-performing model via MTL, and the probabilities of antimicrobial activity against five strains were provided. The two peptides, PA-Full and PA-Win, were selected from the model, and net charge and water solubility were calculated using PepCalc.¹

Bacterial strains and cell lines

Bacterial strains of *E. coli* (KCCM 11234), *P. aeruginosa* (ATCC 9027), *B. subtilis* (ATCC 6051), *S. epidermidis* (ATCC 12228), *S. aureus* (KCCM 11335), and methicillin-resistant *S. aureus* (MRSA, ATCC 33591) were used in this study. Every strain was maintained in Mueller-Hinton broth (MHB, Difco Laboratories, Detroit, MI, United States) under shaking conditions at 37°C. In addition, human cell lines of A549 (ATCC CCL-185), MCF7 (ATCC HTB-22), HaCaT (CLS 300493), NHA (Sciencell 1800), and NHDF (ATCC PCS-201-012) were used. The cells were cultured under a humidified atmosphere with 5% CO₂ at 37°C. A549 cells were cultured in RPMI-1640 (Gibco, Grand Island, NY, United States) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin and streptomycin (PS, Gibco). HaCaT and NHDF cells were cultured in Dulbecco's modified Eagle medium supplemented with 10% FBS and 1% PS. MCF7 cells were cultured in minimum essential media (Gibco) supplemented with 10% FBS, 1% PS, and 1 × non-essential amino acids solution (Gibco). Finally, NHA cells were cultured in astrocyte medium (Gibco).

Antibacterial activity assessment

The MIC values of peptides were measured by performing the broth microdilution method. Bacterial strains were cultured in MHB at 37°C to reach an exponential phase. The bacterial suspension was prepared into 2 × 10⁵ CFU/ml and transferred 50 μl to 96-well microplates. An equal volume of peptides was added to each well, reaching a final concentration from 0.125 to 256 μg/ml of the peptides. The wells containing either only media or the mixture of inoculant and media served as blank and control, respectively. The bacterial growth was measured using a microplate

¹ <http://pepcalc.com>

reader (Molecular Devices, Sunnyvale, CA, United States) at 600 nm after incubation at 37°C for 18 h. MIC values of the peptides were determined of the lowest concentration without observable growth of bacteria. To assess the bactericidal effect of the peptides, 50 µl of the samples from each well below the MIC values were plated onto tryptic soy agar (Difco Laboratories) plates. The agar plates were incubated at 37°C overnight, and the minimum bactericidal concentration (MBC) values were defined as the lowest peptide concentrations without any colony formation.

Measuring bacterial membrane permeabilization

The measurement of outer membrane permeabilization and cytoplasmic membrane depolarization was performed using 1-*N*-phenyl-naphthylamine (NPN) and 3,3'-dipropylthiadicarbocyanine iodide (DiSC₃(5)), respectively, as previously described (Shin et al., 2022).

Field emission scanning electron microscope (FE-SEM) imaging

Morphological changes upon treating peptides were observed using FE-SEM imaging. Target bacteria strains in the exponential phase were diluted in medium to reach 1×10^7 CFU/ml in MHB and seeded onto cover glass coated with poly-L-lysine. After 2 h at room temperature (RT), the bacteria were washed with PBS and treated with $1 \times$ MIC of each peptide for 4 h at RT. The samples were washed with PBS and fixed with 2.5% glutaraldehyde solution at 4°C overnight. After washing with distilled water, bacterial samples were dehydrated with a series of gradient ethanol concentrations (30, 50, 60, 70, 90%, twice for 100%) and then air-dried. Following the platinum coating, bacterial morphology was observed under the FE-SEM sigma instrument (Carl Zeiss Microscopy, Jena, Germany).

Cytotoxicity assay

For cytotoxicity assay of the peptides, the cells were seeded at a density of 1×10^4 cells/well in 96-well microplates. Various concentrations of peptides ranging from 1 to 256 µg/ml were treated and then incubated for 24 h. The microplate was incubated for 1 h after adding Quanti-Max WST-8 Cell Viability assay solution (Biomax, Seoul, South Korea) to each well. For calculating relative cell viability, the absorbance (Ab) was read using a microplate reader (Molecular Devices) at 450 nm.

Hemolysis assay

To determine the hemolytic effects, bovine red blood cells (RBCs, Innovative Research, Novi, MI, United States) were used.

RBC suspensions of 100 µl were transferred to each microtube and were combined with an equal volume of peptide solution. PBS and 0.1% Triton X-100 were used as the negative and positive controls, respectively. The samples were incubated for 1 h at 37°C and then centrifuged at $3,000 \times g$ for 10 min at 4°C. After transferring the supernatants to a 96-well microplate, the Ab was measured using a microplate reader (Molecular Devices) at 450 nm. The hemolytic activity was calculated using the following equation:

$$\text{Hemolytic activity (\%)} = \frac{\text{Ab}_{\text{peptide}} - \text{Ab}_{\text{PBS}}}{\text{Ab}_{\text{Triton X-100}} - \text{Ab}_{\text{PBS}}} \times 100.$$

Statistical analysis

All experimental work was independently conducted, and the results were expressed as mean \pm standard error of the mean (SEM). The statistical significance of the data was evaluated by performing a one-way analysis of variance (ANOVA) test, followed by Bonferroni correction using GraphPad Prism 9.0 (GraphPad Software, La Jolla, CA, United States). *p*-Values of < 0.05 were considered statistically significant.

Results

Training deep learning model for antimicrobial activity prediction *via* MTL

To predict antimicrobial activities targeting a selected set of bacteria, a model should provide several independent prediction results for individual targets. The prediction targets were five different bacterial species, which could be considered multiple related tasks. Multi-label classification could be one approach to model training (Zhang and Zhou, 2013). However, because only 14 peptide sequences were fully labeled for all prediction targets, multi-label classification was not applicable (Figure 1B). Thus, we applied the MTL on non-overlapping datasets by optimizing the model with multiple loss functions (Zhang and Yang, 2017). The model structure was organized as one shared layer connected to several task-specific output layers, otherwise known as hard parameter sharing, and the method effectively prevented an overfitting problem (Figure 1C; Strezoski et al., 2019). During model training, the gradients from the individual prediction target flowed through each task-specific layer and were merged into a shared layer. In comparison, the model for the STL was organized as serially stacked BiLSTM layers without a shared layer, where one model predicts for each target species only (Figure 1D). Considering that the antimicrobial activity of the peptide is exhibited regardless of sequence direction, the BiLSTM-based prediction model was used. Essentially, the STL and the MTL adopted an identical model structure for predicting individual

target species, but only the MTL with the shared layer is trained by broad features of AMPs from the integrated dataset. To determine the model structure with better performance than others, we applied various BiLSTM unit sizes and hyperparameters. The input sequence length fed into the model was restricted to be under 50 AAs, as most AMPs in the database rarely exceed 50 AAs. As multiple trained weights were generated during model training, the optimized model in each training strategy and hyperparameter that resulted in the highest *F1* score on the validation dataset was selected for further evaluation of prediction performance.

To evaluate the trained model performance, the prediction performances of the MTL and the STL were compared by performance metrics. As the number of positive class dataset was extremely small for model performance test, the non-AMP dataset was sampled 10 times larger than the positive data for the test dataset for the better evaluation of the model. Hence, three performance metrics, precision, recall, and *F1* score, were determined. When comparing the box plots of model performance grouped by the STL and the MTL, the MTL showed the best prediction performance in every performance metric and for every target species (Figure 2A). From the scatter plot comparing the prediction performance values trained by the STL versus the MTL under identical model structures, 92.2% of the total results showed improved performances by the MTL than the STL based on the *F1* score (Figure 2B). Finally, we achieved the best performing model in the STL and the MTL strategies based on the *F1* score of the test dataset, and the precision-recall curves of the models were exhibited (Figure 2C). A model with the best performance in the MTL outperformed that in the STL for every prediction target species, except for *S. epidermidis* (Table 2). The MTL achieved a precision-recall area under the curve (PRAUC) score of 0.758 for *B. subtilis*, 0.738 for *E. coli*, 0.886 for *P. aeruginosa*, 0.826 for *S. aureus*, and 0.918 for *S. epidermidis*. In addition, the size differences between the non-AMP and the AMP datasets affected the model performance, where the ratio of 1:3 showed better results than that of 1:1 in precision, *F1* score, and PRAUC, but not recall (Supplementary Figure 1). Overall, the MTL improved the model prediction of the antimicrobial activities of peptides against five species, suggesting that the shared layer boosted the extraction of the latent representation of peptides.

Discovering novel AMPs from the transcriptome of *Pardosa astrigera* spider venom via the MTL model

In order to discover novel AMPs from the spider venom, venom gland transcriptome analysis of *P. astrigera*, a wandering spider, was performed (Shin et al., 2020). The venom gland was separated from the spider, and RNA-sequencing and *de novo* assembly were conducted after total RNA isolation (Haas et al., 2013). A total of 149,710 genes with expression levels > 1.0 fragments per kilobase of transcript per million (FPKM) among 169,160 unigenes were identified. The ORF region was predicted by the TransDecoder program, resulting in total 56,373 coding sequence regions (Hwang

et al., 2021). The coding regions were translated into AA residues, and the sequences were used for further discovery of AMPs.

Transcripts often consist of signal and propeptide sequences that are cleaved to produce the mature protein with a shortened peptide length but full biological activity (Kuhn-Nentwig, 2021). However, the prediction model was designed to have a maximum input sequence length limitation of 50 AAs, leading to a problem that a substantial portion of the peptides with signal and/or propeptide sequences cannot be used for the antimicrobial activity prediction. Thus, we applied the “sliding window” technique for discovering potential AMPs from not only short sequences but long sequences, as 64% of the transcript was over 50 AAs in length (Figure 3A). The “sliding window” technique truncated sequences with the desired length by horizontally moving the “window frame” on an input peptide (Figure 3B). We selected a window size of 20 AAs because the length of most known AMPs is around 20 AAs.

We predicted AMPs from *P. astrigera* data using the input of full sequences under length of 50 AAs and partial sequences by sliding window technique (Figure 3C). The histogram presented the count of peptides by the total number of bacterial targets predicted as AMPs. For example, the case of “2” represents the total sum of AMPs predicted for two species in every combination among five targets. Initially, 20,239 peptide sequences from *P. astrigera* were predicted to have antimicrobial activities without the sliding window technique, and 588 sequences were predicted AMPs against the five bacterial species. When the sliding window technique was implemented using the window size of 20 AAs and step size of 5 AAs, we obtained 1,581,209 partial peptide sequences, of which 167,267 were predicted to be AMPs against five bacterial species. In the case of other window size ranging from 14 to 24, the results are shown in Supplementary Table 2.

Among numerous potential sequences with antibacterial effects, we selected two peptide sequences from each screening method that met the following criteria: FPKM > 10, net charge above + 3, and good water solubility. Each peptide selected from the results with or without the sliding window technique was named PA-Win or PA-Full, respectively (Table 3). The model predicted the two putative AMPs with almost 100% probability, and the peptides were not homologous to any known AMPs found by APD3 and NCBI BLAST (Camacho et al., 2009; Wang et al., 2016). Only uncharacterized proteins with unknown function from diverse spiders were annotated to PA-Win with an *E*-value of $3E-13$. Notably, the discovery of PA-Win would not have been possible from the prediction model were it not for local activity prediction via the sliding window technique. The two selected peptides, PA-Full and PA-Win, were synthesized and used for further experimental evaluation.

Functional evaluation of the PA-Full and PA-Win predicted from the model

The functional properties of the selected peptides were evaluated by *in vitro* experiments. First, PA-Full and PA-Win were

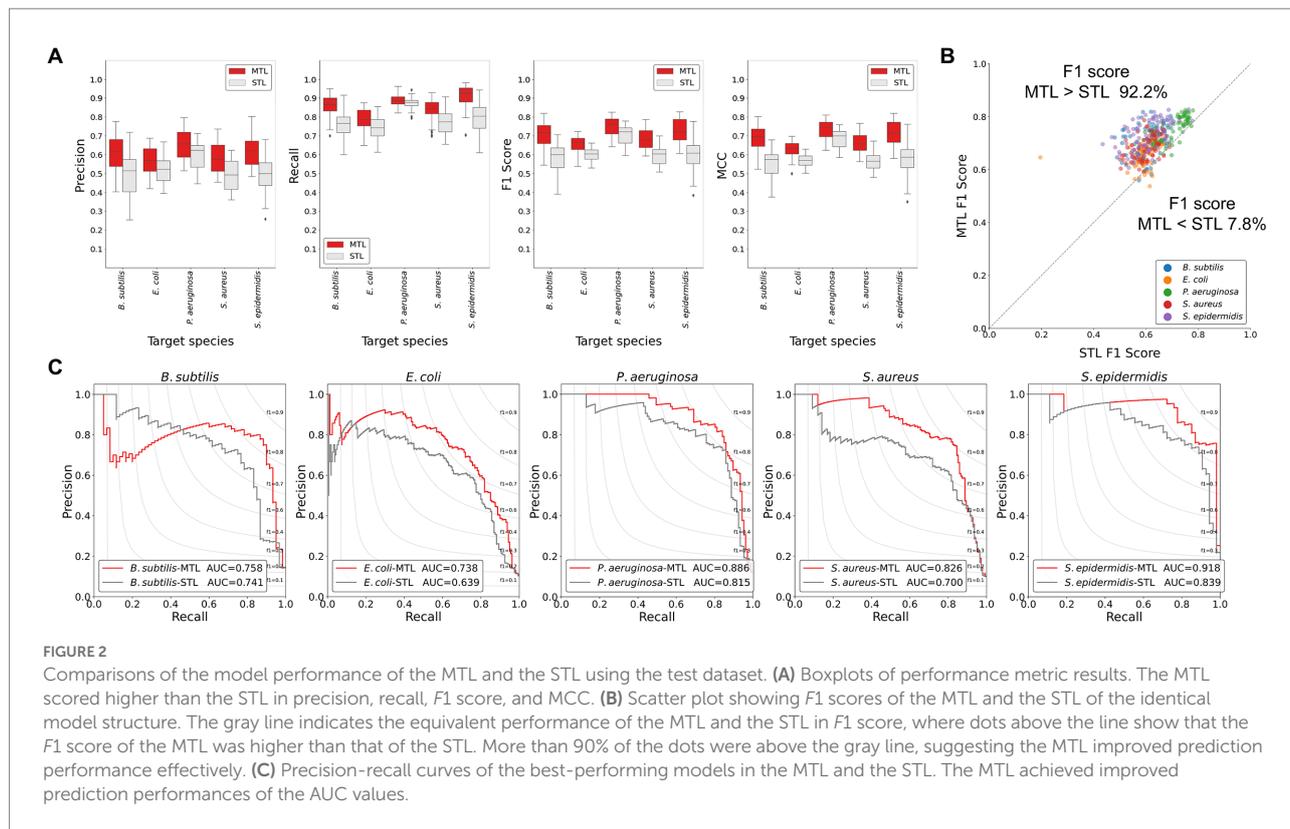


TABLE 2 Prediction results of the best-performing model by multi-task learning (MTL) and single-task learning (STL) on the test dataset.

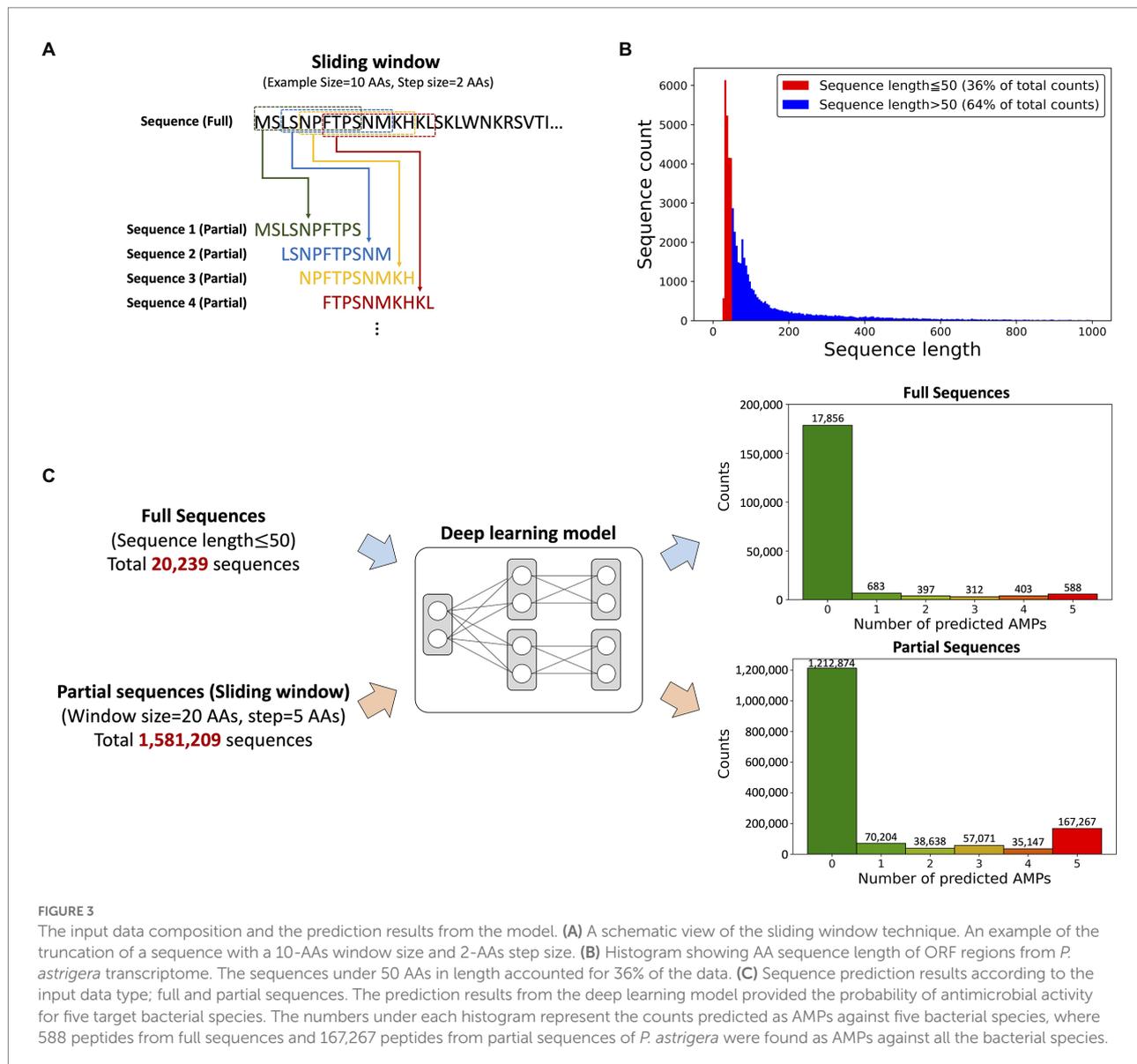
Bacterial strain	Training strategy	Precision	Recall	PRAUC	<i>F1</i> score	MCC
<i>Bacillus subtilis</i>	STL	0.632	0.800	0.741	0.706	0.678
	MTL	0.750	0.900	0.758	0.818	0.802
<i>Escherichia coli</i>	STL	0.587	0.756	0.639	0.661	0.629
	MTL	0.628	0.781	0.738	0.696	0.667
<i>Pseudomonas aeruginosa</i>	STL	0.713	0.860	0.815	0.780	0.759
	MTL	0.789	0.841	0.886	0.814	0.796
<i>Staphylococcus aureus</i>	STL	0.624	0.796	0.700	0.700	0.672
	MTL	0.736	0.845	0.826	0.787	0.766
<i>Staphylococcus epidermidis</i>	STL	0.681	0.907	0.839	0.778	0.762
	MTL	0.663	0.981	0.918	0.791	0.784

The numbers in boldface indicate the best performance.

tested for their antimicrobial activities against Gram-negative bacteria *E. coli* and *P. aeruginosa*, and Gram-positive bacteria *B. subtilis*, *S. epidermidis*, and *S. aureus*, which are the five strains used for securing the AMP sequence data. As a representative AMP, melittin, a major component of bee venom, was used as a positive control because it rapidly kills microbes (Lee et al., 2013). The bacteria were treated with peptide concentrations between 0.125 and 256 µg/ml; that is, under the cutoff value of 267.7 µg/ml applied for pre-processing the model dataset. As shown in Table 4, the two peptides exhibited significant growth inhibition of the target strains. PA-Win showed stronger antibacterial effects than PA-Full, having MIC values ranging from 1 to 8 µg/ml on every strain. The potency of bacterial eradication was then measured by

the MBC of the peptides. The MBC of PA-Win was obtained against every species, whereas PA-Full was measured only on *B. subtilis*. It was confirmed that the two peptides have antimicrobial function against five different strains, and PA-Win was demonstrated to be a strong AMP, showing a strength comparable to melittin.

The toxicity of the peptides on mammalian cells was also tested to investigate the cytocompatibility for further biological applications. The cytotoxic effect was measured on various cell lines, such as human lung carcinoma (A549), immortalized human keratinocytes (HaCaT), human breast adenocarcinoma (MCF7), normal human astrocytes (NHA), and normal human dermal fibroblasts (NHDF), using the WST assay



(Supplementary Figure 2). The half-maximum inhibitory concentration (IC_{50}) values and their 95% confidence interval were calculated from three individual tests for each cell line. When treated with PA-Full, cell lines were not affected and remained viable even at peptide concentrations above 256 $\mu\text{g}/\text{ml}$ (Figure 4A). The IC_{50} values of PA-Win on cell lines were at least eight times higher than the highest MIC value, 8 $\mu\text{g}/\text{ml}$ (Figure 4B). The hemolytic activities of the peptides were evaluated using bovine erythrocytes. Hemolytic activity was presented as the relative value to 0.1% Triton-X treatment, exhibiting 100% hemolysis, and the concentrations of 0.5 \times , 1 \times , and 2 \times MIC of PA-Full, PA-Win, and melittin were tested. When treated with 2 \times MIC, PA-Full and PA-Win caused almost no hemolysis which was below 3%, whereas the melittin caused hemolysis over 90% (Figure 4C). Both peptides showed inhibitory concentrations that can exert sufficient antibacterial activity without affecting mammalian cells, and

PA-Win was observed to be the stronger AMP among the two peptides.

Effects of PA-Full and PA-Win on bacterial membrane integrity

It is well known that AMPs exert their antibacterial activity by disrupting the membrane, causing pore formation and leakage of cellular components. Fluorescent dyes DiSC₅(5) and NPN were used to evaluate the effects of the selected AMPs on bacterial membrane integrity, individually targeting the cytoplasmic and outer membrane. Along with the two selected peptides, 1 \times MIC of melittin (shown in Table 3) was again used as a positive control because it kills microbes by membrane disruption. As depicted in Figure 5A, bacterial cells treated with the two respective peptides

TABLE 3 The prediction results and characteristics of the PA-Full and PA-Win.

Name	Sequence	Predicted results of antimicrobial activity (%)					Sequence similarity			Net charge	Water solubility	FPKM
		<i>Bacillus subtilis</i>	<i>Escherichia coli</i>	<i>Pseudomonas aeruginosa</i>	<i>Staphylococcus aureus</i>	<i>Staphylococcus epidermidis</i>	APD3	BLAST				
PA-Full	WVLSKIKRKKKEN	100.0	100.0	100.0	99.0	99.8	-	-	+10.8	Good	10.8	
	SDHQTKFSKVKTKR											
PA-Win	WKRFFPRVV	100.0	100.0	100.0	100.0	100.0	-	Hypothetical protein	+9.8	Good	13.5	
	RKIFRRRIKR											

showed either an instant or time course increase in relative fluorescence intensity by DiSC₅(5) due to a depolarized cytoplasmic membrane. PA-Full and PA-Win caused sharp increases in the relative fluorescence of Gram-positive strains than Gram-negative bacteria. In Gram-negative strains, the NPN uptake assay was additionally conducted to measure the outer membrane disruption caused by treating with the peptides (Figure 5B). The maximum relative fluorescence intensity was reached within a minute for both peptides, where the relative fluorescence intensity obtained by treating with PA-Full exceeded that by melittin. The results showed that PA-Full and PA-Win effectively led to biomembrane defects.

In agreement with the DiSC₅(5) and NPN fluorescent measurements, FE-SEM revealed the morphological changes upon peptide treatment (Figure 5C). The control groups were treated with an equal volume of PBS. While the control group retained their normal shape with an intact surface, the bacteria treated with the two respective peptides for 4 h showed apparent damages on cells. The peptide-treated groups showed blebs and deformations on the cell surface and even complete rupture. Altogether, these results indicated that PA-Full and PA-Win permeabilized the cells and disrupted bacterial membranes, like other known AMPs.

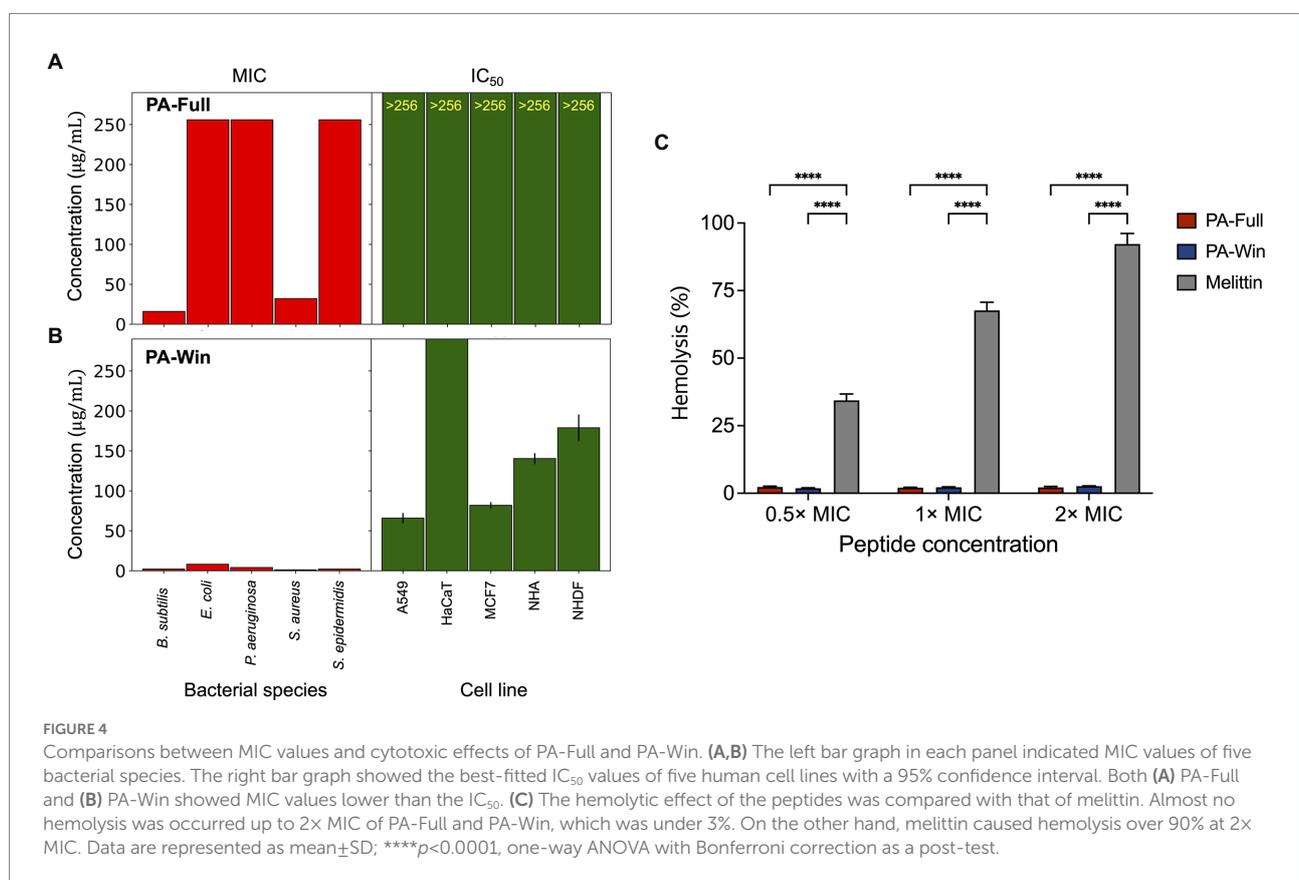
Discussion

Antimicrobial peptides (AMPs) have recently received major attention for the potential new therapeutic agents because of their ability to efficiently kill a broad range of microorganisms by permeating biomembranes, damaging cellular components, and inhibiting metabolic processes (Mwangi et al., 2019). Animal venoms comprise various biological components that are used in predation and defense, such as AMPs. However, identifying potential AMPs from animal venoms via conventional experimental methods is time-consuming and labor-intensive. Therefore, we developed a deep learning model that predicts potential candidate AMPs based only on AA sequences and then used the transcriptome data obtained from the venom gland of spider *P. astrigera* to discover novel AMPs.

Spider venom is a rich source of functional biomolecules as it is well known of its diversity in venom components with various effects, such as antimicrobial, analgesic, antimalarial, and anti-arrhythmic activities. Thus, we obtained venom gland transcriptome data from the spider *P. astrigera* to identify novel AMPs. Many studies are based on homology search when handling the transcriptome data for investigating and characterizing animal venoms. However, because homology search is based on the sequence similarity with already discovered peptides, it has limitations in identifying completely novel functional peptides with no resemblance to known peptide sequences. By contrast, the deep learning approach predicts the antimicrobial functionality based on the peptide sequence per se; that is, the antimicrobial activity is predicted based on the input sequences of AAs without considering any physicochemical properties, only using the latent representation

TABLE 4 The MIC, MBC, and IC50 values of PA-Full and PA-Win.

Bacterial strain	MIC ($\mu\text{g/ml}$)			MBC ($\mu\text{g/ml}$)			IC ₅₀ ($\mu\text{g/ml}$) (95% CI)	
	PA-Full	PA-Win	Melittin	PA-Full	PA-Win	Cell line	PA-Full	PA-Win
<i>Bacillus subtilis</i>	16	2	8	16	2	A549	>256	65.9 (59.6–72.5)
<i>Escherichia coli</i>	256	8	4	>256	8	HaCaT	>256	>256
<i>Pseudomonas aeruginosa</i>	256	4	16	>256	8	MCF7	>256	81.7 (77.6–86.0)
<i>Staphylococcus aureus</i>	32	1	2	>256	4	NHA	>256	140.3 (133.7–147.4)
<i>Staphylococcus epidermidis</i>	256	2	2	>256	2	NHDF	>256	178.8 (162.9–196.3)



of the AA sequences. Thus, the deep learning model can broaden the search for AMPs that the homology search cannot discover.

In order to discover peptides with antibacterial activity against a broad range of bacterial species, we aimed to develop a deep learning model for predicting antimicrobial functionality against five bacterial species. Among the various criteria employed for categorizing AMPs by several databases, a cutoff value of MIC of 267.7 $\mu\text{g/ml}$ was applied to obtain more data in the means of increasing the prediction target species. Lowering the MIC cutoff may possibly identify stronger AMPs than the current model, however, it should limit the prediction target by diminishing the

number of peptides that can be utilized to train the model. Finally, the AMP data used for this study were selected based on available peptides that met the minimum number applicable for the deep learning model training. However, about 60% of the peptides were demonstrated to kill only single bacterial species among five target species (Figure 1B). The majority of the positive data was AMPs with narrow-spectrum antimicrobial activity; although such sequences may be active against species that were not tested yet, it still provides information only on narrow-spectrum until experimentally validated. In this context, a prediction model for AMPs without considering species-specific targets will provide

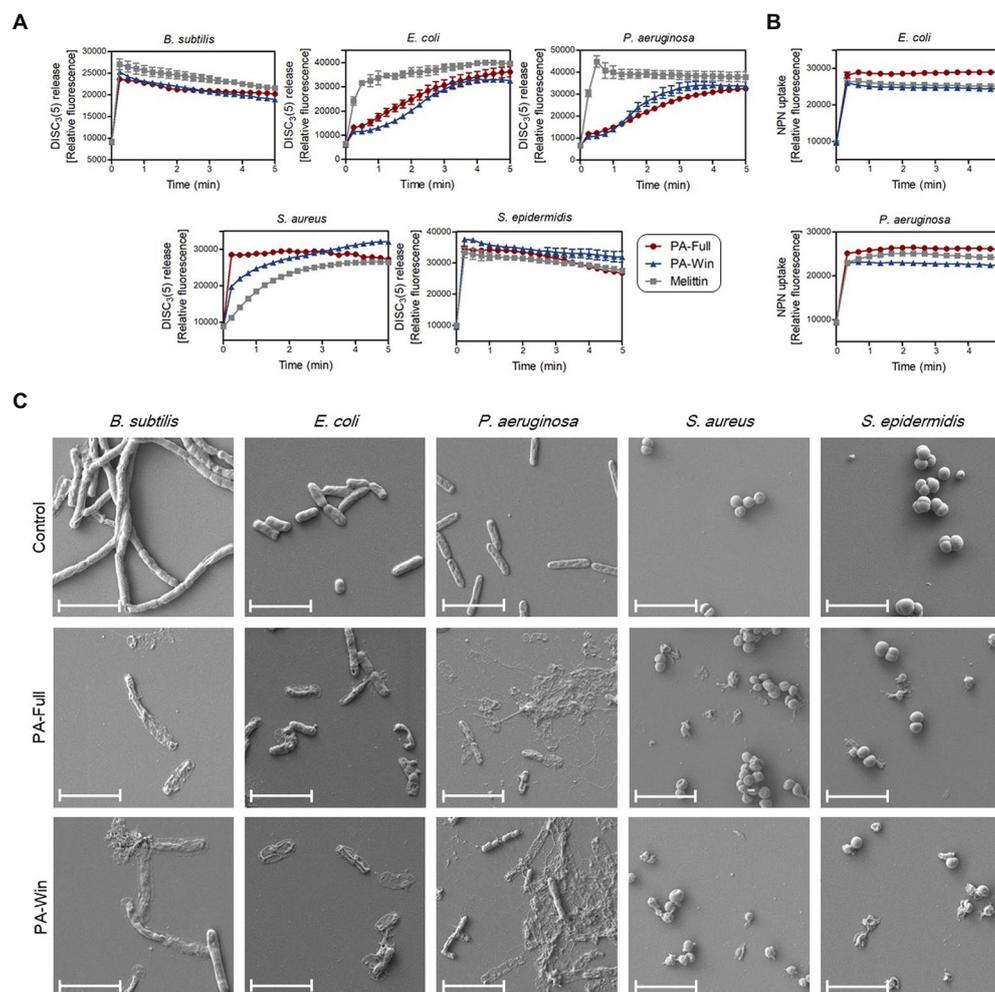


FIGURE 5
Disruption of bacteria membrane upon PA-Full and PA-Win treatment. **(A)** The cell membrane permeability of DiSC₃(5) was measured to evaluate the integrity of bacterial cytoplasmic membranes. The increase in fluorescence intensity was observed when Gram-positive and Gram-negative bacteria were treated with 1 × MIC of the peptides, PA-Full, PA-Win, or melittin. **(B)** The permeabilization of the outer membranes of Gram-negative bacteria was observed by the NPN uptake assay. Treatment with PA-Full or PA-Win showed sharp increases in fluorescence intensity comparable to that of melittin. **(C)** Morphologies of bacteria imaged by FE-SEM. Every strain treated with 1 × MIC of peptides for 4 h showed damage to bacterial membranes with blebs and ruptures compared with the negative control treated with an equal volume of PBS. The scale bars represent 4 μm.

AMP sequences with a narrow-spectrum, targeting only one or two strains. Therefore, we designed a species-specific prediction model that provides information about antimicrobial activity against multiple bacterial species, two Gram-negative and three Gram-positive species, to avoid the discovery of narrow-spectrum AMPs.

The data for AMPs against *B. subtilis*, *E. coli*, *P. aeruginosa*, *S. aureus*, and *S. epidermidis* barely met the minimum amount of data for the deep learning approach, while the other strains were not even close. Thus, we developed a deep learning model *via* the MTL approach to overcome and compensate for these data limitations. The MTL is known to improve predictive performance by training multiple related tasks simultaneously, increasing data efficiency, and reducing overfitting (Crawshaw, 2020). Several efforts were made to increase the classification ability of the model using the data size between AMPs and non-AMPs. When it comes

to the feature distribution on peptides, the AMP class can be said to be “local,” whereas the non-AMP is scattered as “global,” comprising diverse peptides without distinctive functions. For the model to improve the discrimination between both classes, the test dataset was constructed with AMPs and non-AMPs with a 1:10 ratio. In addition, two types of model training datasets were configured for each target bacterial species with the ratio between AMP and non-AMP as 1:1 and 1:3 for comparison. The MTL showed better predictive performance in the test dataset compared with STL. The weights of the shared layer were tuned by the gradient flows from five different tasks, which may facilitate the shared layer to learn more general latent representations of the peptides. When comparing the results based on the different amount of non-AMP in the training dataset, using a ratio of 1:3 between AMP and non-AMP data in both MTL and STL showed

better performance than a ratio of 1:1 in precision. Meanwhile a ratio 1:1 partially performed better in recall metric, which may be due to the imbalance of the training data. A ratio of 1:3 resulted in improved prediction performance based on both the *F1* score and PRAUC, which denote the balanced model performance, suggesting the distinguishability between AMP and non-AMP has further improved. Finally, based on the highest *F1* score, we selected the best-performing model for discovering novel AMPs.

The best-performing model was provided with two input data types from the transcriptome of the spider *P. astrigera*, the full sequences and the partial sequences created by the sliding window, for antimicrobial activity prediction. Input data was about 78 times larger in the case of the partial sequences by sliding window and predicted about 284 times larger potential AMPs compared with the full sequence data. The sliding window technique facilitated the expansion of the input data, as well as the prediction results from the same transcriptome data. We selected PA-Full and PA-Win each from the predicted results using the full and partial sequences as the input data. The transcripts showed no homology with known peptides found by NCBI BLAST, resulting in PA-Full with no match and PA-Win with a hypothetical protein of unknown function (*E*-value of $3E-13$). Both peptides were identified as AMPs through experimental evaluation, and, particularly, PA-Win showed outstanding antimicrobial activity with low toxicity to mammalian cells. Our deep learning model discovered novel AMPs that otherwise cannot be suggested by the traditional methods, including but not limited to BLAST. The model also demonstrated that PA-Win would not have been identified without implementing the sliding window technique. Finally, PA-Win has shown significant antibacterial activity comparable to that of a representative AMP, melittin, against all the tested strains and exhibited cytocompatibility without hemolytic activity, further raising its value. By using the model that we developed, it enables high-throughput prediction of novel functional peptides with low cost and effort, excavating peptides with high potential and high applicability as AMPs, such as PA-Win.

Our model can predict AMP functionality from any sequence data based on AAs, including the transcriptomes of other venomous species. The sliding window technique even accelerates the production of vast prediction results, which can contribute to the acquisition of AMP candidates for next-generation antibiotics. Conclusively, as our MTL-based deep model has successfully discovered an ideal AMP (i.e., PA-Win), we believe the methodology can be applied to the fields of biology with insufficient data for utilizing specific biological data and exploring multiple functional resources, including but not limited to AMPs.

References

- Ashfaq, A., Shahid, A., Salman, K., Maqsood, H., Farman, A., Aftab, A., et al. (2021). Deep-AntiFP: prediction of antifungal peptides using distant multi-informative features incorporating with deep neural networks. *Chemom. Intell. Lab. Syst.* 208:104214. doi: 10.1016/j.chemolab.2020.104214
- Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Alpi, E., Bely, B., et al. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. doi: 10.1093/nar/gky1049

Data availability statement

The datasets generated for this study can be found at <https://github.com/bzlee-bio/AMPSpeciesSpecific>.

Author contributions

MKS and BL conceptualized the study, developed the deep learning model, performed experimental validation, and wrote the manuscript. JSY acquired the funding, coordinated the project, and provided revision for the study. J-SS and WJ supervised the research and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by a grant from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIBR202231204).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.971503/full#supplementary-material>

- Budnik, B., Olsen, J., Egorov, T., Anisimova, V., Galkina, T., Musolyamov, A., et al. (2004). De novo sequencing of antimicrobial peptides isolated from the venom glands of the wolf spider *Lycosa singoriensis*. *J. Mass Spectrom.* 39, 193–201. doi: 10.1002/jms.577

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10:421. doi: 10.1186/1471-2105-10-421

- Cantón, R., and Morosini, M.-I. (2011). Emergence and spread of antibiotic resistance following exposure to antibiotics. *FEMS Microbiol. Rev.* 35, 977–991. doi: 10.1111/j.1574-6976.2011.00295.x
- Centres for Disease Control and Prevention (2019). *Antibiotic Resistance Threats in the United States, 2019*. Atlanta, GA: US Department of Health and Human Services, Centres for Disease Control and Prevention.
- Charoenkwan, P., Chiangjong, W., Lee, V. S., Nantasenamat, C., Hasan, M. M., and Shoombuatong, W. (2021). Improved prediction and characterization of anticancer activities of peptides using a novel flexible scoring card method. *Sci. Rep.* 11:3017. doi: 10.1038/s41598-021-82513-9
- Chen, Y., Li, Y., Narayan, R., Subramanian, A., and Xie, X. (2016). Gene expression inference with deep learning. *Bioinformatics* 32, 1832–1839. doi: 10.1093/bioinformatics/btw074
- Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey. *arXiv preprint arXiv:2009.09796* [Epub ahead of preprint]. doi: 10.48550/arXiv.2009.09796
- Cunha, F. M., Berti, D. A., Ferreira, Z. S., Klitzke, C. F., Markus, R. P., and Ferro, E. S. (2008). Intracellular peptides as natural regulators of cell signaling. *J. Biol. Chem.* 283, 24448–24459. doi: 10.1074/jbc.M801252200
- Fjell, C. D., Jenssen, H., Hilpert, K., Cheung, W. A., Panté, N., Hancock, R. E. W., et al. (2009). Identification of novel antibacterial peptides by chemoinformatics and machine learning. *J. Med. Chem.* 52, 2006–2015. doi: 10.1021/jm8015365
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Huan, Y., Kong, Q., Mou, H., and Yi, H. (2020). Antimicrobial peptides: classification, design, application and research progress in multiple fields. *Front. Microbiol.* 11:11. doi: 10.3389/fmicb.2020.582779
- Hwang, I.-W., Shin, M. K., Lee, Y.-J., Kim, S. T., Lee, S. Y., Lee, B., et al. (2021). N-type Cav channel inhibition by spider venom peptide of *Argiope bruennichi*. *Mol. Cell. Toxicol.* 17, 59–67. doi: 10.1007/s13273-020-00109-2
- Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* [Epub ahead of preprint]. doi: 10.48550/arXiv.1412.6980
- Krumm, B. E., and Grishammer, R. (2015). Peptide ligand recognition by G protein-coupled receptors. *Front. Pharmacol.* 6:48. doi: 10.3389/fphar.2015.00048
- Kuhn-Nentwig, L. (2021). Complex precursor structures of cytolytic cupiennins identified in spider venom gland transcriptomes. *Sci. Rep.* 11:4009. doi: 10.1038/s41598-021-83624-z
- Langenegger, N., Nentwig, W., and Kuhn-Nentwig, L. (2019). Spider venom: components, modes of action, and novel strategies in transcriptomic and proteomic analyses. *Toxins* 11:611. doi: 10.3390/toxins11100611
- Lata, S., Mishra, N. K., and Raghava, G. P. S. (2010). AntiBP2: improved version of antibacterial peptide prediction. *BMC Bioinform.* 11:S19. doi: 10.1186/1471-2105-11-S1-S19
- Lee, B., Baek, J., Park, S., and Yoon, S. (2016). “DeepTarget: end-to-end learning framework for MicroRNA target prediction using deep recurrent neural networks.” in *BCB '16: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 434–442.
- Lee, B., Shin, M. K., Hwang, I.-W., Jung, J., Shim, Y. J., Kim, G. W., et al. (2021). A deep learning approach with data augmentation to predict novel spider neurotoxic peptides. *Int. J. Mol. Sci.* 22:12291. doi: 10.3390/ijms222112291
- Lee, M.-T., Sun, T.-L., Hung, W.-C., and Huang, H. W. (2013). Process of inducing pores in membranes by melittin. *Proc. Natl. Acad. Sci.* 110, 14243–14248. doi: 10.1073/pnas.1307010110
- Lei, J., Sun, L., Huang, S., Zhu, C., Li, P., He, J., et al. (2019). The antimicrobial peptides and their potential clinical applications. *Am. J. Transl. Res.* 11, 3919–3931.
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12:323. doi: 10.1186/1471-2105-12-323
- Meher, P. K., Sahu, T. K., Saini, V., and Rao, A. R. (2017). Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Sci. Rep.* 7:42362. doi: 10.1038/srep42362
- Mookherjee, N., Anderson, M. A., Haagsman, H. P., and Davidson, D. J. (2020). Antimicrobial host defence peptides: functions and clinical potential. *Nat. Rev. Drug Discov.* 19, 311–332. doi: 10.1038/s41573-019-0058-8
- Mwangi, J., Hao, X., Lai, R., and Zhang, Z.-Y. (2019). Antimicrobial peptides: new hope in the war against multidrug resistance. *Zool. Res.* 40, 488–505. doi: 10.24272/j.issn.2095-8137.2019.062
- Nuti, R., Goud, N. S., Saraswati, A. P., Alvala, R., and Alvala, M. (2017). Antimicrobial peptides: a promising therapeutic strategy in tackling antimicrobial resistance. *Curr. Med. Chem.* 24, 4303–4314. doi: 10.2174/0929867324666170815102441
- Pirtskhalava, M., Gabrielian, A., Cruz, P., Griggs, H. L., Squires, R. B., Hurt, D. E., et al. (2016). DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res.* 44:6503. doi: 10.1093/nar/gkw243
- Posner, B. I., and Laporte, S. A. (2010). Cellular signalling: peptide hormones and growth factors. *Prog. Brain Res.* 181, 1–16. doi: 10.1016/S0079-6123(08)81001-1
- Quang, D., and Xie, X. (2016). DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. *Nucleic Acids Res.* 44:e107. doi: 10.1093/nar/gkw226
- Recio, C., Maione, F., Iqbal, A. J., Mascolo, N., and De Feo, V. (2017). The potential therapeutic application of peptides and peptidomimetics in cardiovascular disease. *Front. Pharmacol.* 7:526. doi: 10.3389/fphar.2016.00526
- Schmidt, B., and Hildebrandt, A. (2021). Deep learning in next-generation sequencing. *Drug Discov. Today* 26, 173–180. doi: 10.1016/j.drudis.2020.10.002
- Shin, M. K., Hwang, I.-W., Kim, Y., Kim, S. T., Jang, W., Lee, S., et al. (2020). Antibacterial and anti-inflammatory effects of novel peptide toxin from the spider *Pardosa astrigera*. *Antibiotics* 9:422. doi: 10.3390/antibiotics9070422
- Shin, M. K., Lee, B., Kim, S. T., Yoo, J. S., and Sung, J.-S. (2022). Designing a novel functional peptide With dual antimicrobial and anti-inflammatory activities via in Silico methods. *Front. Immunol.* 13:821070. doi: 10.3389/fimmu.2022.821070
- Strezoski, G., Van Noord, N., and Worring, M. (2019). “Learning task relatedness in multi-task learning for images in context”, in *Proceedings of the 2019 on international conference on multimedia retrieval*, 78–86.
- Tan, J., Doing, G., Lewis, K. A., Price, C. E., Chen, K. M., Cady, K. C., et al. (2017). Unsupervised extraction of stable expression signatures from public compendia with an ensemble of neural networks. *Cell Syst* 5, 63–71.e6. doi: 10.1016/j.cels.2017.06.003
- Tsomaia, N. (2015). Peptide therapeutics: targeting the undruggable space. *Eur. J. Med. Chem.* 94, 459–470. doi: 10.1016/j.ejmech.2015.01.014
- Van Dijk, A., Van Eldik, M., Veldhuizen, E. J. A., Tjeerdma-Van Bokhoven, H. L. M., De Zoete, M. R., Bikker, F. J., et al. (2016). Immunomodulatory and anti-inflammatory activities of chicken Cathelicidin-2 derived peptides. *PLoS One* 11:e0147919. doi: 10.1371/journal.pone.0147919
- Wadhvani, P., Sekaran, S., Strandberg, E., Bürck, J., Chugh, A., and Ulrich, A. S. (2021). Membrane interactions of laticarins: antimicrobial peptides from spider venom. *Int. J. Mol. Sci.* 22:10156. doi: 10.3390/ijms221810156
- Waghu, F. H., Barai, R. S., Gurung, P., and Idicula-Thomas, S. (2016). CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.* 44, D1094–D1097. doi: 10.1093/nar/gkv1051
- Wang, G. S., Li, X., and Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* 44, D1087–D1093. doi: 10.1093/nar/gkv1278
- Wang, L., Li, Y., Zhou, J., Zhu, D., and Ye, J. (2017). “Multi-task Survival Analysis”, in *2017 IEEE international conference on data mining (ICDM)*, 485–494.
- Wang, X., and Wang, G. (2016). Insights into antimicrobial peptides from spiders and scorpions. *Protein Pept. Lett.* 23, 707–721. doi: 10.2174/0929866523666160511151320
- Yeung, A. T., Gellatly, S. L., and Hancock, R. E. (2011). Multifunctional cationic host defence peptides and their clinical applications. *Cell. Mol. Life Sci.* 68, 2161–2176. doi: 10.1007/s00018-011-0710-x
- Zhang, Y., and Yang, Q. (2017). A survey on multi-task learning. *arXiv preprint arXiv:1707.08114* [Epub ahead of preprint]. doi: 10.48550/arXiv.1707.08114
- Zhang, M.-L., and Zhou, Z.-H. (2013). A review on multi-label learning algorithms. *IEEE Trans. Knowl. Data Eng.* 26, 1819–1837. doi: 10.1109/TKDE.2013.39
- Zhou, G., Shi, Q.-S., Huang, X.-M., and Xie, X.-B. (2015). The three bacterial lines of defense against antimicrobial agents. *Int. J. Mol. Sci.* 16, 21711–21733. doi: 10.3390/ijms160921711