HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Methods Mol Biol. 2022 ; 2405: 1–37. doi:10.1007/978-1-0716-1855-4_1.

Machine Learning Prediction of Antimicrobial Peptides

Guangshun Wang[1,*], **Iosif I. Vaisman**[2,*], **Monique L. van Hoek**[2,*]

[1]Department of Pathology and Microbiology, College of Medicine, University of Nebraska Medical Center, 985900 Nebraska Medical Center, Omaha, NE 68198-5900, USA;

[2]School of Systems Biology, George Mason University, 10920 George Mason Circle, Manassas, VA, 20110, USA.

Abstract

Antibiotic resistance constitutes a global threat and could lead to a different pandemic. One strategy is to develop a new generation of antimicrobials. Naturally occurring antimicrobial peptides (AMPs) are recognized templates and some are already in clinical use. To accelerate the discovery of new antibiotics, it is useful to predict novel AMPs from the sequenced genomes of various organisms. The antimicrobial peptide database (APD) provided the first empirical peptide prediction program. It also facilitated the testing of the first machine learning algorithms. This chapter provides an overview of machine-learning predictions of AMPs. Most of the predictors, such as AntiBP, CAMP, and iAMPpred, involve a single-label prediction of antimicrobial activity. This type of prediction has been expanded to antifungal, antiviral, antibiofilm, antiTB, hemolytic, and anti-inflammatory peptides. The multiple functional roles of AMPs annotated in the APD also enabled multi-label predictions (iAMP-2L, MLAMP, and AMAP), which include antibacterial, antiviral, antifungal, antiparasitic, antibiofilm, anticancer, anti-HIV, antimalarial, insecticidal, antioxidant, chemotactic, spermicidal activities and protease inhibiting activities. Also considered in prediction are peptide post-translational modification, 3D structure, and microbial species-specific information. We compare important amino acids of AMPs implied from machine learning with those frequent occurring residues of the major classes of natural peptides. Finally, we discuss advances, limitations and future directions of machine learning predictions of antimicrobial peptides. Ultimately, we may assemble a pipeline of such predictions beyond antimicrobial activity to accelerate the discovery of novel AMP-based antimicrobials.

Keywords

Multi-drug resistance; antimicrobial peptides; database; machine learning; peptide prediction

1. Introduction

The discovery and production of antibiotics has saved millions of lives. It is regarded as one of the greatest achievements of humankind in the twentieth century. However, pathogens fight back, leading to reduced potency of conventional antibiotics. To minimize

[*]**Corresponding to:** Dr. Monique van Hoek: mvanhoek@gmu.edu; Dr. Iosif Vaisman: ivaisman@gmu.edu; Dr. Guangshun Wang: gwang@unmc.edu.

toxic effects, bacteria can pump the drug out of the cells, reduce drug affinity to specific targets via mutations, and degrade antibiotics by proteases. Among various multi-drug resistant (MDR) microbes, the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) account for the 90% infections in hospitals [1]. There are also other emerging resistant pathogens, including human immunodeficiency virus type 1 (HIV-1), SARS-CoV2, Ebola, Zika viruses, resistant bacteria *Mycobacterium tuberculosis*, *Salmonella*, *Candida, Neisseria gonorrhoeae*, and *Clostridioides difficile*. If no action is taken, the projected annual deaths could reach 10 million by 2050 [2]. To meet this challenge, one fundamental strategy is to develop a new generation of antimicrobials that are capable of eliminating those MDR pathogens.

Antimicrobial peptides (AMPs) are considered as an alternative to conventional non-peptide antibiotics. This chapter focuses on prediction of antimicrobial peptides. First, we provide a brief introduction to AMPs. Second, we discuss the major prediction methods of AMPs. Third, both the data sets for predictions and the algorithms of machine learning are described. Fourth, we discuss the major machine learning prediction of AMPs. Fifth, we compare the prediction outcomes of machine learning in terms of accuracy on the same platform, results from test runs using new peptides not included in the training sets, and the important amino acids implied from machine learning with those derived from our database analysis of the major classes of natural AMPs. Then, we outline additional predictions that may speed up computer-aided novel antimicrobial discovery. Finally, we summarize the major achievements and limitations of AMP predictions and discuss future directions.

2. Innate immune antimicrobial peptides

Naturally occurring antimicrobial peptides are important components of innate immune systems. Such peptides are deployed in a variety of organisms such as plants and animals. They play a critical role in protecting organisms from infections. AMPs have remained potent for millions of years. As a consequence, they are recognized candidates for developing novel antimicrobials since they can kill drug-resistant pathogens, including bacteria, fungi, viruses, and parasites. AMPs are usually gene-encoded and can be expressed constitutively to guard certain niches or induced in response to invading pathogens [3–8]. According to the antimicrobial peptide database (APD, https://aps.unmc.edu), over 3000 natural AMPs have been discovered from six life kingdoms (bacteria, archaea, protists, fungi, plants, and animals) [9–11]. At present, 74% of the peptides originated from animals, while 11.2% and 11.1% were discovered in bacteria and plants, respectively. Most of natural AMPs (88%) are cationic and only a small portion (6%) are anionic. Anionic AMPs, such as daptomycin already in clinical use, may need metal to be active [12]. In the APD, the majority of AMPs possess hydrophobic contents (Pho) between 10 and 70% (defined in Table 1). Only about 1% such peptides have very high (>70%) or very low (<10%) Pho. In terms of length, 2879 peptides in the current APD3 (88%) are shorter than 50 amino acids. The average length of all AMPs (3257 as of January 2021) in the APD3 is 33.2 with an averaged net charge of +3.3. The frequently occurring amino acids (>8%) are leucine (L), glycine (G), and lysine (K) [10], while the least occurring amino acids (<2%) include methionine (M) and tryptophan (W) (Table 1). Such frequencies are

proportional to the percentage of natural AMPs containing one of the 20 amino acids also calculated in Table 1. The variation of the amino acid (composition) signatures of natural AMPs in different structure, activity, and source groups has been tabulated elsewhere [13]. Figure 1 displays amino acid signatures for known α-helical, β-sheet peptides (panel A), tryptophan-rich (Trp-rich), histidine-rich (His-rich), proline-rich (Pro-rich) AMPs, and leucine-rich (Leu-rich) temporins (panel B). It is evident that such signatures depend on the amino acid composition of a group of AMPs in the APD. The amino acid sequence of a peptide, however, clearly plays a role as well in determining peptide structure and activity [6,14]. Another important player is post-translational modification (e.g., amidation, glycosylation, halogenation, hydroxylation, and cyclization) of peptide sequences, with 24 types of modifications annotated in the current APD3 as of October 2020 [11,15]. Typically, cationic AMPs target anionic bacterial membranes due to the formation of the classic amphipathic helix structure [3–6]. However, such peptides can also attack other targets such as bacterial cell walls and ribosomes. It is believed that the simultaneous attack of more than one targets renders it difficult for bacteria to develop resistance to AMPs. Beyond bacterial killing and biofilm inhibition, AMPs are found to have other functional roles, ranging from pathogen toxin neutralization, wound healing to host immune regulation [4,5,16]. A total of 24 types of AMP functions are annotated in the APD3 [11,13].

3. An overview of prediction methods of antimicrobial peptides

The majority of natural AMPs were identified using the classic isolation and characterization methods [3–5]. Such peptide identification procedures are laborious and time-consuming. One alternative method is to predict AMPs by computers based on the current peptide knowledge and sequenced genomes of numerous organisms [9,17–19]. These prediction methods are grouped into five classes based on the information considered in programming [20]: (1) mature peptide (i.e., AMPs), (2) propeptide, (3) mature peptide and propeptide, (4) processing enzyme, and (5) genomic context (Figure 2). Some AMPs such as cathelicidins possess a conserved pro-sequence domain prior to the mature peptide. Such a conserved sequence pattern became one method for identifying uncharacterized cathelicidins from sequenced genomes for mammals, fish, reptiles, birds, and amphibians (method 2). The human cathelicidin was initially predicted as FALL-39 [21], which is merely 1–2 resides longer than the mature forms isolated in human neutrophils and reproductive system (LL-37 and ALL-38), respectively [22,23]. In the same vein, the discovery of bacteriocins from bacteria has been expanded from highly conserved processing enzymes (method 4a) to transporters (method 4b) and the entire gene clusters (i.e., genomic context; method 5). Computer programs such as BAGEL, antiSMASH, and BACIIα have been established for bacteriocin identifications [24–26]. Occasionally, both precursor and mature sequences (method 3) were considered in clustering AMPs probably due to the nature of a particular data set then available [27]. The most widely explored information for prediction are mature peptides (method 1). Sequence patterns such as multiple disulfide bonds were utilized for identifying defensin-like AMPs in plants, cattle, mice, and humans [28–30]. A GXC γ-core motif has also been identified in these peptides and utilized for AMP prediction [31].

The construction of databases for AMPs greatly facilitated the development of computer-based design [32] and prediction methods. Table 2 provides a list of databases for AMPs

[11,18,33–49]. In 2004, the APD and ANTIMIC were simultaneously published in the database issue of *Nucleic Acid Research* in 2004 [9,50]. The APD, with a focus on structure and activity of mature AMPs, was widely accepted and utilized by the AMP field [9]. Since then, more databases have been established with varying scopes or by entering additional details (Table 2). A systematic review on such databases has been described elsewhere [51]. Because of the model role of the APD, it is useful to describe its data scope and evolution. In the first two versions [9,10], the APD attempted to cover all AMP sequences: experimentally determined, predicted, and synthetic. This history can be seen from a small number of synthetic and predicted entries remaining in the current APD (72 synthetic peptides and 211 predicted peptides without activity data). There are three types of activity data annotated in the APD: (1) minimal inhibitory concentration (MIC); (2) diffusion distance; and (3) optical density decrease as an evidence of inhibition. Due to convenience, MIC values based on microdilution assays are frequently measured and reported. Since predicted peptides without experimental data might not be true AMPs [11], it was decided to postpone the collection of such peptides in the APD. Also, a large number of the synthetic peptides derived from the same template tended to dominate data filtering in the database, thereby deviating the database filtering from natural wisdom to artificial peptides. As a consequence, the APD also postponed the collection of synthetic peptides. Thus, the third version of the APD (APD3) [11] uses the following criteria to register AMPs: (1) natural peptides, (2) peptides with known amino acid sequences, (3) peptides with known activity (MIC < 100 μM), and (4) peptides less than 100 amino acids [11]. The last condition was relaxed to 200 amino acids to incorporate important human antimicrobial proteins. This practice generates a welcomed data set for AMP search, prediction and design.

Based on mature peptides, the first computer-based prediction was programmed in the APD in 2003 [9]. The program informs users whether the input sequence is likely to be an AMP based on some known AMP knowledge, such as positive charge and amphipathic nature. Later, it was improved based on the peptide parameter space (net charge, hydrophobic content, and peptide length) defined by the entire database [19]. If such parameters of a new sequence are out of the scope, the program will inform the users that the input sequence is less likely to be an AMP. The APD also outputs five peptide sequences most similar to the user's input.

Subsequently, Lata et al. first programmed artificial neural network (ANN), quantitative matrices (QM), and support vector machine (SVM) in 2007 based on the APD data set [17]. Since then, there has been a growing interest in AMP prediction at both the single-label and multi-label levels. The single-label prediction will predict the likelihood of being antimicrobial, while multi-label predictions were developed based on different functions of AMPs annotated in the APD3 [11], such as chemotaxis, toxin neutralization, protease inhibition, and wound healing. The first multi-label prediction [52] predicts antibacterial activity in the initial stage followed by predictions of other types of activities, including antifungal, antiviral, anti-HIV, and anticancer activities. CAMP collected both synthetic and predicted peptides. Its prediction tool [18,53] enables three tasks. First, users can predict the antimicrobial activity of a peptide sequence by four different models. Second, users can predict the antimicrobial region within a peptide sequence. Third, users can generate a large combinatorial list of sequences for a user-defined sequence and then can predict effect of

single residue substitutions on antimicrobial activity using the AMP predictor. Table 3 lists some major machine learning prediction programs [53–78].

4. Training data sets, machine learning models and algorithms for classification and prediction of antimicrobial peptides

Machine learning models are commonly used for classification and prediction of AMP. Nearly all machine learning predictions of AMPs are supervised. The quality of these models is determined by a number of different factors. Among the most important contributors to the model performance are training sets consisting of antimicrobial and non-antimicrobial peptides, features used to represent the peptides, classification schemes, and machine learning algorithms.

4.1. Training sets for predictions

4.1.1. Positive training set—Quality of the training set is critically important for the model performance, since it is the only source of information the model uses to learn. AMP sequences for the training set are usually extracted from one or more of AMP databases. The growing number of AMP databases (some examples are listed in Table 2) represents a wide range of approaches to data collection, data curation, and data management. For the purpose of training set design, it is important to take into account that AMP databases vary in size, sources of information, amount and quality of annotations, and other parameters. Sizewise, the current versions stretch from over 3,000 peptides in the APD [9–11] to 10,000 in CAMP [18,53], 12,000 in dbAMP [48], 16,000 in DBAASP [33], and 23,000 in LAMP2 [40]. Some of the larger databases (e.g., LAMP2 [40]) may contain the entire content of the smaller ones by copying the peptide entries from existing databases. At the same time, the non-overlapping components are frequently present, primarily in the scope of synthetic peptides and due to different definitions of AMPs. Some specialized databases have expanded the data set by including other types of peptides, which do not necessarily fall into the definition of classic AMPs [44,49]. For instance, antiviral peptides can also be designed by investigators in the laboratories based on the viral machinery such as proteases. As a result, the distribution of peptides by sequence length in databases can be different as well. The APD contains mostly natural AMPs, which are templates for making synthetic peptides. For example, there are hundreds of LL-37 derived peptides. 88% of the entries in the APD are less than 50 amino acids and only 80 peptides out of 3257 have a length greater than 100 residues. Similarly, most peptides in DBAASP database are shorter than 50 residues. Only 20 entries in DBAASP are longer than 100 residues, while CAMP contains 1,850 such sequences. The longest sequence in APD and DBAASP is less than 190 residues compared to 1,256 residues in CAMP.

The first training set for machine learning model test was extracted from the APD [17]. Another data set used in AMP prediction was derived from the CAMP [18]. Because the majority of natural AMPs in the CAMP were taken from the APD, there is a significant overlap between these two data sets. Some recent studies generated a hybrid data set by merging the peptide sequences from different databases [61,62,70,71,78]. The size of the positive data set appears to influence prediction outcome [61]. Species-specific predictions

of AMPs [69] were made based on the DBAASP, which annotate antimicrobial activity in more details [33]. For 3D structural data, the APD has direct links to the Protein Data Bank (PDB) [79]. Hence, a list of training peptides with 3D structures can also be generated without redundancy (i.e., multiple coordinates for the same peptide).

4.1.2. Negative data set—Ideally, the negative set should consist of peptides which were tested experimentally and displayed no antimicrobial activity against one or more relevant pathogens. Non-AMP sequences are a natural byproduct of any wet lab screening for antimicrobial peptides. However, negative results are rarely published and as a result the large sets of validated non-AMP sequences are likely sitting in the drawers of investigators and not available to the public. Creating a database of non-AMP sequences and convincing researchers to contribute data into this database would be a helpful step in improving the quality of the training sets.

Bioinformaticians/computing scientists have taken an alternative approach to obtaining negative data sets. The AntiBP [17] generated the first negative data set based on the Uniprot [80]. The negative part of the training set is usually selected from the random sequences in the protein sequence database, which are not annotated as antimicrobial, secretory, toxins, etc. Sequences in the negative set can be controlled by the level of sequence identity, sequence composition, similarity to the sequences in the positive set, structural and other properties. Since the protein sequence databases are very large (the October 2020 release of UniProt database contains more than 200 million sequences) [80], the supply of sequences for the negative sets is practically unlimited. There are caveats with these data. The sequences in the negative set may possess antimicrobial properties, although the probability of this is relatively low. Also, antimicrobial activities of AMPs are very sensitive to sequence variation [81]. Such features may not be represented in the current negative data set. Training the models on different combinations of a positive set with several independent negative sets may provide insights into the scale of negative set contamination by hitherto unknown antimicrobial peptides.

In many cases it is advisable to use a balanced training set, where the AMP and non-AMP sequences are equally represented. AMP sequences can be selected from AMP databases (Table 2). Normally, only a subset of the entire database (or several databases) can be used to compile a positive part of the training set. Sequences from the database are filtered by length, activity, sequence identity, and other parameters. In most studies the positive sets range from several hundred to several thousand sequences, while the size of the negative set from the Uniprot can be much larger. However, the data sets for numerous species-specific predictions were much smaller due to limited MIC data [69].

4.2. Descriptors and features

Many different features of peptides can be used to characterize their antimicrobial activity and discriminate between antimicrobial and non-antimicrobial peptides. Frequently these features are based on identities, physico-chemical properties, structural properties, and compositions of individual amino acid residues and their combinations [61,82–84]. Physical and chemical properties of amino acids which are most likely to improve

machine learning (ML) model performance include hydrophobicity, electrostatic charge, and polarity. Similarly important are structural properties such as helical propensity and solvent accessibility. In many models feature vectors include residue locations in the sequence, compositional characteristics and sequence patterns. The overall number of features can be very large, in those cases feature selection can help to reduce the size of the feature vector by removing features with relatively low contributions to the model performance.

4.3. Machine learning algorithms

A large number of different machine learning algorithms (Table 3) have been implemented in AMP classification and prediction models since the first papers reporting this approach were published in 2007 [17, 27, 85]. ML methods successfully used in AMP modeling include K-nearest neighbor [52, 86], hidden Markov models (HMMER) [27], naïve Bayes [86], neural networks (NN) (including their deep learning varieties) [63,71,72,87–90], support vector machines [17,18,58,59,61,64,66,73,76], random forests (RF) [18,60,62,65,70,74,77], zero-shot learning (ZSL) [69] and many others (Table 3).

Support vector machines classification maps feature vectors representing the peptides in the training set into a higher dimensional space. Then the algorithm constructs an optimal hyperplane which separates two classes of peptides, AMPs and non-AMPs, with the maximal margin of separation between the classes. This hyperplane serves as a decision boundary in the original space. The hyperplane divides the entire higher dimensional space into two half-spaces, and each new peptide from the prediction set is going to be located in one of these two half-spaces. This location will determine the predicted class for new peptides.

Decision tree (DF) classifier has the form of a rooted binary tree. A divide-and-conquer approach is used during model training. It traverses the tree starting from the root, and at each node an input feature is selected that best separates the output classes. Learned trees are frequently pruned to decrease overfitting. After the tree is created using a training set, a new peptide can be sorted down the tree based on the values of the input features on the corresponding node, and the appropriate branch is followed to the next node. The recursive process terminates once the peptide reaches a leaf node, where the peptide class, AMP or non-AMP is identified. Random forests algorithm is an ensemble method based on decision trees. It generates multiple bootstrapped datasets, each dataset trains a classification tree by randomly selecting a fixed-size subset of the available predictors for splitting at each node, and predictions are made by majority vote over all trees. Random forests help to avoid many pitfalls of the decision tree algorithm, particularly overfitting.

While most of the predictions aimed to discriminate AMP and non-AMP (i.e., single-label), several labs have attempted a multi-label prediction based on the multi-functional data annotated in the APD3 [11,13]. The four multi-label predictions (iAMP-2L, MLAMP, AMAP, and AMPfun) all conduct predictions in two levels [52,60,66,68]. Similar to the single-label prediction described above, the first level of the multi-label prediction predicts whether the peptide is an AMP or non-AMP. If it is, then the program moves onto the second level prediction to predict the likelihood of other functions the peptide may have. These can include antibacterial, antibiofilm, antiviral, anti-HIV, antifungal, antiparasitic,

antimalarial, anticancer, insecticidal, antioxidant, chemotactic, enzyme inhibitors, and spermicidal activity. It appears that AMAP is best in terms of accuracy. It also predicted more biological functions of AMPs at the second level.

To evaluate the performance of an algorithm on a training set, cross-validation (CV) and random split into two subsets are commonly used. Implementation of tenfold CV begins with a random grouping of the training set peptides into ten equally sized subsets. Stratification is applied to maintain class proportions of the full training set in each of the subsets. At the next step, one of the subset is held out while the remaining nine subsets (90% of the original training set) are combined into one set that is used to train a model. The held - out subset (10% of the original training set) is then treated as a test set, and the trained model predicts the class for each peptide in the subset. Then the procedure is repeated for the remaining nine combinations. The iterative procedure yields a single prediction for each of the peptides in the original training set, which is then compared to the actual class. These comparisons allow to calculate the numbers of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions. Commonly used performance measures, such as sensitivity, specificity, precision, balanced error rate and Matthew's correlation coefficient, are all functions of these four numbers. Many published ML models report CV accuracy values which are close to 100%. The actual real world performance of these models on predicting novel antimicrobial peptides may be lower due in part to the extremely complex AMP activity landscape.

5. Machine learning predictions of special antimicrobial peptides

5.1. Utility and Main drawbacks of AMP prediction algorithms

Overall, our ability to accurately predict the antimicrobial activity, hemolytic activity or cytotoxic activity of any peptide sequence is a developing field. While advances in machine learning, positive and negative data-sets and analytic approaches have been made, the accuracy of predicting the properties of a new peptide sequence is still low, too low to be of reliable use in a screening step for example. Improvements in the peptide sorting and analysis, especially thinking about the different surface properties of gram-negative and gram-positive bacteria, could yield significant advancements in accuracy, which would significantly advance the field. This lack of reliability is the main drawback of AMP prediction algorithms and the main hindrance in their use in high-throughput design programs to generate new AMPs.

5.2. Antiviral peptide predictors and data

The antiviral activity of antimicrobial peptides is of considerable interest. In particular, antiviral peptides (AVPs) appear to have activity against membrane-enveloped viruses, such as LL-37 against influenza virus [91,92]. Some peptides (e.g., LL-37 and θ-defensins) have been found to have HIV inhibitory activities [93]. Antiviral peptides (AVPs) have been shown to exert their activities at various steps in the viral lifecycle, including impeding attachment to host cells, altering viral replication within cells or indirectly by recruiting other parts of the immune system to promote host defense [93]. The antimicrobial peptide LL-37 has been shown to be effective to inhibit attachment and entry of the influenza virus

[91,92]. As an example of the indirect mode of antiviral activity, the Rhesus theta-defensin has been shown to be indirectly antiviral against SARS-CoV-1 [94], with the major effect being an increase in the host defense that allows survival of the mice against this infection. LL-37 is also active against Zika virus [95]. Recently, several highly effective AMPs were designed that show significant activity against Ebola virus (EBOV) infection of cells [96]. These peptides were designed or "engineered" fragments of LL-37 peptide [7], and were found to strongly inhibit EBOV entry into in cell lines and human primary macrophages, but not viral replication [96]. This study represents an exciting advance in both the design of active antiviral peptides and their application to important diseases such as Ebola.

Several websites [97–99] have been established to assist the prediction of AVPs (Table 4). Using database analysis and a feature reduction technique (recursive feature elimination (RFE) algorithm), one group generated a software tool to predict antiviral peptides with this advance, Feature-Informed Reduced Machine Learning for Antiviral Peptide Prediction (FIRM-AVP) [99]. The analysis assembled 649 features that correlated with antiviral activity and then applied a reduction of the number of features to 169 based on the Pearson's correlation coefficient and computed MDGI (mean decrease of Gini index) values. They then applied the RFE technique to order the features by importance and to identify the most important features. Three features that were identified in common between two different parts of the analysis include "PseAAC (pseudo amino acid composition) feature for leucine (L) amino acid", "PseAAC feature for lysine (K) amino acid", and "Location oriented feature for α-helix" [99]. This suggests that these features may have strong contribution to the physicochemical features of an effective antiviral peptide. Overall, this is in agreement with the general observation that anti-viral peptides are often alpha-helical and positively charged peptides [93].

5.3. Antifungal peptide predictors and data

Specific databases and prediction models [100,101] have been developed for antifungal peptides (AFPs) (Table 5). Antifungal peptides appear to have a prominence of the amino acids cysteine (C), glycine (G), histidine (H), lysine (K), arginine (R), and tyrosine (Y) in their amino acid sequences [101]. A similar set of frequently occurring amino acids L, C, alanine (A), G, K, and R is obtained when 1210 antifungal AMPs in the APD was statistically analyzed [11]. Positional analysis suggests that the amino-terminus of antifungal peptides may predominately be R, valine (V) or K, while C and H are predominant at the carboxyl terminus of the peptide. This is different from the most common amino acids (G, L, A, and K) found in antibacterial helical peptides [10,11].

5.4. Specific and unique peptide prediction tools

Many other specialized prediction algorithms for peptides have been developed in recent years [102–104]. While anti-inflammatory and pro-inflammatory activities are closely linked to infection outcomes, these peptides may not be directly antimicrobial. However, it may be of interest to antimicrobial peptide researchers, especially since many antimicrobial peptides, such as LL-37, are known to have host-directed effects in addition to antibacterial effects [108]. Some websites have been developed for predicting very specific kinds

of activities that may be of interest to antimicrobial peptide researchers, including anti-inflammatory peptides, pro-inflammatory peptides and anti-tubercular peptides (Table 6).

5.5. Tuberculosis

Tuberculosis (TB) continues to be a plague on humanity, infecting more than 10 million people each year worldwide, and is responsible for approximately 2 million annual deaths globally. The emergence of multi-drug resistant and extremely multidrug resistant (XDR) strains of TB, especially in prisons and other enclosed conditions, is an extreme challenge to society and to the medical community to develop new approaches to treat these infections. Antimicrobial peptides may represent one new approach to treating *Mycobacterium* infection [105–107], likely in combination with other treatments. The AntiTBpred website has been developed to help researchers parse through antimicrobial peptide sequences and to try to identify candidates that might be useful against this recalcitrant and challenging organism.

Using LL-37, the human cathelicidin, as an example, AntiTBpred analysis suggests that this peptide either may or may not be an anti-tubercular peptide. Studies have shown that *in vitro* and *in vivo*, LL-37 is antibacterial for *Mycobacterium tuberculosis* (MTb) and can reduce bacilli counts in a mouse model [108]. Further studies have shown that LL-37 is required to control intracellular MTb replication [106–108]. The antimicrobial peptide HBD2 has also been shown to have antibacterial activity against MTb *in vitro* [109]. In the output example below, these two peptide sequences were analyzed using all 4 models within AntiTBPred. Only 1 of the 4 models correctly predicted (grey highlights) that HBD2 was antiTB, and it also predicted that LL-37 would be antiTB.

5.6. Antibiofilm peptide predictors and data

Biofilm formation by bacteria is a major contributor to colonization, persistence and difficulty in treatment of bacterial infections. Chronic, non-healing diabetic wounds on the lower extremities, lung infections in cystic fibrosis patients, hip-replacement and other orthopedic implants and chronic bladder infections all have bacterial biofilm as a major component of their etiology. In recent years, as our understanding of bacterial biofilms has increased [110–112], it has become clear that some antimicrobial peptides have the ability to either prevent the attachment and formation of biofilm or can induce the dispersal of bacterial biofilms [113–120]. Several databases and websites [11,35,121–123] have been developed to gather the information on antibiofilm peptides and to try to predict their activity (Table 8).

Although not strictly a peptide-focused resource for peptide researchers, a related tool aBiofilm (https://bioinfo.imtech.res.in/manojk/abiofilm/) [124] may be of interest to antibiofilm peptide researchers. This tool provides a database, an antibiofilm predictor and data-visualization tools.

6. Antimicrobial prediction outcome comparison

6.1. Prediction comparison on the same platform

The prediction accuracy of AMPs can be determined by numerous factors, ranging from data sets, peptide sequence information encoding, to algorithms. Which data set to use depends on the aim of the prediction and personal knowledge. How to represent the peptide faithfully in a manner which is understandable by computers is a challenging task by itself. This is further complicated by numerous types of chemical modifications annotated in the APD3 [11]. An optimized prediction requires a sufficient definition of both the types and numbers of peptide features. Such peptide features range from a dozen to hundreds. The algorithms or models may be used alone or in combination.

Data sets: A reliable data set is critical to obtain useful predictions. Machine-learning predictions normally use a balanced positive and negative data ratio of 1:1 to avoid a biased prediction toward the large data set. CAMP used a positive:negative ratio of 1:1.5 [18]. AmPEP tested numerous ratios and achieved a higher accuracy when a 1:3 ratio was utilized [62]. A too high ratio is undesired as the prediction will tilt toward negative sequences, thereby reducing the overall performance of machine learning in predicting AMPs. Meher and colleagues tested the effect of the size of positive peptides. They found that the more positive peptides, the better the prediction [61]. This makes sense because the prediction program is better trained with more positive examples (synthetic + natural AMPs). When more and more synthetic peptides are included, however, the prediction accuracy toward natural AMPs may drop. This is undesired when the goal is to scan the genomes to discover novel antibiotics.

Peptide features: A thorough description of the peptide sequence would require numerous features. The first prediction noticed the need of a more complete representation of peptide information. A higher accuracy was achieved when the peptide features from both the N and C-termini were considered [17]. Wang et al. [54] utilized 270 sequence features to represent each AMP. These include 20 standard amino acids (AAC) and 50 pseudo-amino acid compositions (PseAAC) that describe the peptide sequence based on positional correlations between amino acids. Each PseAAC is also linked with five features: polarity, secondary structure, molecular volume, codon diversity, and electrostatic charge (50×5). However, each peptide feature may not play the same role in prediction. In pattern recognition, it is most important to identify the major features significant for peptide classification. CAMP started with 257 features and found 64 features were best for RF [18]. It is possible to further reduce the peptide features required for prediction. Bhadra et al. were able to reduce the features from 105 to 23 without a loss of prediction accuracy [62]. Tripathi and Tripathi utilized merely 15 peptide features to reach a comparable prediction accuracy, including the consideration of the sequence shuffling effect [70]. It appears that only a dozen of key peptide features are needed to achieve a comparable prediction accuracy.

Algorithms/models: Tripathi and Tripathi applied different algorithms (RF, J48, SVM, and Naïve Bayes) to peptide prediction based on the same data set. They found Random Forest is best [70]. Also, Yan et al. found that deep learning (CNN) performed similarly to

RF but better than SVM [71]. However, both SVM (8 studies) and RF (7 cases) are popular in Table 3. To reduce overfitting, there is also an attempt to utilize an ensemble approach by involving multiple models [78]. Lin and Xu [60] revealed a higher accuracy of the more recent multi-label prediction methods such as iAMP-2L and MLAMP (92.2% and 94.7%) than those programmed in the CAMP (SVM, RF, and DA at 57.8%−77.5% accuracy) [18]. It appears that the high accuracy reported for machine learning does not match the outcomes of real tests (below). There is a room to improve for all the existing programs.

6.2. Testing the prediction outcomes by using peptides not included in the training set

How each program performs in AMP prediction can be put into practice. We tested the AntiBP program by using newly discovered natural AMPs, which were not included in the training set. Among the 17 peptides with known activity, 71% were predicted correctly [20]. Another test was conducted in 2015 using 10 new peptides (APD ID: 2399–2408) [51]. AntiBP SVM predicted 70% correctly, whereas the RF, SVM, ANN, and DA programs in CAMP [18] obtained 60–80% correctness. iAMP-2L [52] achieved a similar prediction of 80%. Bishop et al. [125] identified 568 novel peptides from alligator plasma. From 45 predicted to be AMPs by CAMP [18], eight peptides were chemically synthesized and subjected to antibacterial assays. Five were experimentally proved to be antimicrobial (a prediction accuracy of 5/8 = 62.5%). Yan et al. [71] developed Deep-AmPEP30 and predicted three antimicrobial sequences from the genome of *Candida glabrata*, and one peptide was proved active against Gram-positive bacterium *Bacillus subtilis* and Gram-negative *Vibrio parahaemolyticus*. These tests underscore the limitations of existing programs. Porto et al. [81] found that the machine-learning programs worked well only for peptides resembling the trained data set. However, they failed to predict sequence shuffled peptides [14], indicating an insufficient consideration of peptide sequence information.

6.3. Comparison with existing AMP knowledge

Every machine learning algorithms is essentially a black box. It is not surprising that there is no direct link between the computing outcome and AMP biology. AmPEP compared various descriptors that distinguish the AMPs from non-AMPs and identified charge as the most important descriptor [62]. The iAMPpred program [61] also found the importance of net charge followed by isoelectric point of the peptides in the training set. The iAMP-2L program reveals that amino acid composition accounts for 60% of the weightings [52]. Taken together, the AMP charge and composition are two major features for AMP differentiation. Overall, these machine learning findings agree with the research results of AMPs that cationicity and hydrophobicity are the two most important factors that determine peptide antimicrobial activity. Amino acid composition is important in determining peptide activity spectrum as well [9,126,127].

Some programs documented selected amino acids to be important predictors of AMPs. Based on the APD3 data set, the AMAP study [66] identified amino acids C, K, V, and phenylalanine (F) for AMP prediction, whereas aspartic acid (D), glutamic acid (E), L, Y, proline (P), R, and asparagine (N) are indicators for non-AMPs. Using a merged data set, iAMPpred identified amino acids K, P, C, and isoleucine (I) [61]. Wang [54] found C, P, R, W, and H based on both natural and patented AMPs in the CAMP database. In another

study, amino acids G, F, P, and W were identified [44] based on the DBAASP data set [33]. It is evident that there is a low level of consensus from different prediction studies. This may result from differences in the training data sets, algorithms, and the assessment of important features during prediction.

It may be useful to compare the above amino acids with the frequently occurring amino acids (~10%) discovered from analyses of the major classes of natural peptides in the APD3 [10]. G, L, A, and K are frequently occurring (abundant) amino acids (~10% or more) in 463 known helical AMPs. In contrast, amino acids C, G, and R are abundant in natural AMPs with a known β-sheet structure (87 in the APD3) (Figure 1A). For the "rich" families, His-rich AMPs are clearly rich in H and A, while Pro-rich AMPs are rich in P and R. Also, Trp-rich peptides are abundant in W and R (Figure 1B). When combined, we have G, L, A, K, C, R, H, P, R, and W. Most of the machine learning discovered amino acids correspond to the frequently occurring amino acids of AMPs discovered in the APD3 [13]. Machine learning also identified hydrophobic V, F, and I. While F and I are abundant in helical AMPs from fish and mammals, V is abundant in lactone and lactam types of bacteriocins [13]. It is puzzling why both L and A were not identified by any machine learning. Leucine is clearly rich in 121 amphibian temporins (Figure 1B) and important for peptide design [32]. Alanine is particularly high in amphibian AMPs from South America [13]. Increased conversations between AMP and bioinformtic people may improve the prediction outcomes in the future.

7. Beyond antimicrobial properties and proposed prediction integration toward future medicine

7.1. Antimicrobial peptide properties that contribute to AMP activity

As discussed above, the general properties of peptides that appear to be positively correlated with AMP activity have been identified from experience and usually include the following physico-chemical parameters: (1) peptide length, (2) amphipathicity, (3) hydrophobicity and (4) cationicity. However, the translation of these general principles into very specific physico-chemical rules by which certain sequences can be included or excluded or predicted to have antimicrobial activity or not has been the challenge of the last decades since their discovery. As discussed above, there are many detailed bioinformatic and computational approaches that seek to solve this problem of AMP prediction (Table 3).

7.2. Important antimicrobial peptide properties in addition to AMP activity

Additional properties of peptides will contribute to them being "successful" antimicrobial peptides besides AMP activity. These properties, beyond antimicrobial peptide activity, include: toxicity towards host cells, ability to penetrate microbial or eukaryotic membranes, susceptibility to host proteases and "stickiness", the propensity to be bound to albumin or other high-abundance proteins in the host, among others. Host-cell toxicity can include hemolytic activity and cytotoxicity, or it can be observed *in vivo* through toxicity trials. Cell permeability of the peptide can be a critical factor if the target of the AMP is an intracellular bacteria for example. "Stickiness" to high-abundance host proteins or high susceptibility to host proteases can affect the *in vivo* availability of the peptide and its half-life, aspects of pharmacodynamics (PD) and pharmacokinetics (PK) that have significant implication for

future clinical success. Unfortunately, the PK/PD data for AMPs are sparse, since most of the peptides have not been advanced to that level [6]. Some of the major parameters for consideration and possible inclusion in a computational approach are listed in Table 9. Many tools for computing these properties are available online, for example in R (Peptides, https://rdrr.io/cran/Peptides/man/), ExPASy (expasy.org), and the calculation tool of the APD3 [11].

LL-37 is a widely studied human cathelicidin peptide encoded by the single CAMP gene. It is stored in and released from neutrophils and expressed in other types of human cells as well. Depending on the cells and physiological conditions, the precursor of human cathelicidin may be cleaved into different mature peptides. This peptide has been found to be antibacterial against many pathogens, including resistant strains, persisters and biofilms. It belongs to the classic amphipathic helical family with a short tail at the C-terminus (PDB: 2K6O) [7]. In Table 9A, the major physicochemical properties of LL-37 are shown as computed by one of the many websites described below. This peptide is short (37 aa), amphipathic (>1), cationic (net charge +6), has a high pI (>10) and has a low molecular weight (under 5 kDa). ExPASy ProtParam tool provides instability index (23.34) and aliphatic index 89.46. The APD website calculates GRAVY (−0.724), Boman index (2.99 kcal/mol), and Wimley-White whole residue hydrophobicity (12.83) for LL-37. As a well-studied peptide, we will use LL-37 as an example in our discussion of the online tools described below.

7.3. Host-cell toxicity and hemolysis

Host-cell cytotoxicity and hemolysis are critical to the clinical potential of any antimicrobial peptide. Thus, we propose that this issue needs to be considered early, right after identification of desired antimicrobial activity of any peptide as a potential strong counter-selection criterion. Although sequence features such as multiple lysines and high hydrophobicity are known to contribute to host-cell cytotoxicity, it appears to remain challenging to “design-out” host-directed toxicity of active peptides while retaining the desired antimicrobial activity of the sequence. The combined AMP selection and counter-selection procedure leads to a short list of AMPs with high therapeutic indexes for experimental validation.

There are multiple online programs available for the computational prediction of toxicity and hemolysis of antimicrobial peptides. For example, Gupta et al have published a method of *in silico* toxicity prediction for peptides [128,129]. This site is called ToxinPred and has two algorithms available, ToxinPred SVM-SwissProt, ToxinPred QM-di-SwissProt. To illustrate the use of this website, we submitted the sequence of LL-37, the human cathelicidin, to compare the prediction versus in-laboratory data (Table 9A, B). It can be seen that experimentally, the cytotoxicity of LL-37 is dose-dependent, and increases with increasing concentration of peptide (Table 9B). However, this subtlety of concentration of peptide is not captured by the predictors, which just predict one result for some unknown concentration of peptide. Thus, just like a stopped clock is correct twice a day, the predictor is correct at some concentrations of LL-37 and is incorrect at higher concentrations. This concentration-dependence of the real-life data needs to be integrated with computational

predictors in the future, perhaps by including the concentrations at which the results are included in the dataset as an “antibacterial” or “non-cytotoxic” peptide.

Hemolytic activity is the ability of a peptide to lyse red blood cells. This assay is normally performed with a washed 2% solution of red blood cells, following a standard protocol [135,136]. Many different red blood cell types can be used, depending on the intent of the experiment, such as sheep [135–137], horse [138], chicken [139] or mouse [140,141], which may be more sensitive to peptide hemolysis than human red blood cells [141]. Often it is desirable to use de-identified human blood to test hemolytic activity, which can be obtained from companies like BioIVT and used in these assays [141]. Computational predictors of hemolytic activity can be used to compute an estimate of hemolytic activity. For example, HemoPred [142], HemoPI/Hemolytik [143] and HAPPENN [144] are some of the websites currently available (Table 10). HemoPred utilizes a random forest classifier based on amino acid sequence, dipeptide composition and physicochemical parameters [142]. HemoPI is based on comparing a dataset of highly hemolytic peptides to a random dataset of peptides from SwissProt [143]. Finally, HAPPENN tool employs neural networks based on classification of known peptides as hemolytic and non-hemolytic to predict the hemolytic activity from a new peptide’s primary sequence [144].

As an exercise, we ran the sequence of the LL-37 peptide through the various hemolysis predictors (Table 11) and compared the results to published laboratory generated data regarding hemolytic activity (Table 12).

From the literature, the following hemolysis data was obtained for the LL-37 peptide (Table 12), as an example. This is not a comprehensive meta-analysis, but shows data from several papers that contained data over a wide range of peptide concentrations and hemolytic results [145–151]. Of course, there is no indication from these computational predictors of dose-dependence of the effect, although “the dose makes the poison” in most cases with antimicrobial peptides, including LL-37. The prediction results vary from absolutely one end of the hemolytic activity spectrum to the other – one analysis result says “Not Hemolytic”, one result is “Somewhat hemolytic” and one result is “Hemolytic”. This small analysis suggests that there is significant room for improvement in the accuracy of these predictors compared to actual experimental data generated in the laboratory (Table 12 and Figure 3).

7.4. Bacterial Cell-penetrating Peptides:

Another factor that may need to be considered in computational prediction of AMP activity is the characteristic of cell-penetration of the pathogen itself: bacteria, membrane-virus, fungal cell, etc. While the main mechanism of action of AMPs is clearly membrane targeting and disruption, there are multiple, well-defined examples of intra-bacterial targets of AMPs that may contribute to their physiological effect, especially at Sub-MIC levels *in vivo*. These can include targeting bacterial enzymes critical for bacterial survival, or direct interference of the AMP with the bacterial DNA. One example of the association of AMPs with critical bacterial enzymes is the identification of Acyl Carrier Protein as a target of LL-37, the human cathelicidin protein. This association was first determined biochemically by binding the bacterial proteins to immobilized peptide and identifying high-affinity binding proteins [152]. Another example of intra-bacterial targets of AMPs is

the association of LL-37 directly with bacterial DNA within the cell, leading to mutations of critical genes [153,154]. This work includes a compelling visualization of the AMP inside the live *Pseudomonas* bacteria, associated with the DNA. This property of AMPs to enter the bacteria to exert some direct, non-membrane acting effect could be computationally assessed using cell-penetrating peptide (CCP) analysis, such as is done for other well-known CPPs [155]. Unlike AMPs, CPPs for bacterial pathogens should have the property of being non-killing but membrane-penetrating, and comparison of these sets of peptide sequences may reveal some interesting differences. It might be possible to use the CPP algorithm to counter-select for peptides that do not have this property if a membrane-targeting peptide was desired to possibly achieve bactericidal activity.

7.5. Inclusion of additional parameters in drug development

It would be useful if these computational predictors could be used in a combinatorial fashion to achieve the goals of the researcher in designing new AMPs, such as was designed in the database filtering technology approach [156, 157]. For example, perhaps one seeks a short, helical antimicrobial peptide that has activity against gram-negative bacteria and especially has anti-biofilm activity and low hemolytic activity. It would be useful to have separate analytical tools linked together to generate the desired output. With the ever increasing number of modules available in R, and web-based prediction and analysis tools, this analysis could be done from small scale to high-throughput sequence analysis to design novel peptides. If the computational predictors could be made more accurate, this could be useful in drug-development projects upstream of *in vitro* screening programs for example, to increase hit efficacy. The inclusion of pre-screening for hemolysis and cytotoxicity would be very useful to reduce the number of hits that have poor *in vivo* performance characteristics. In addition, high throughput peptide sequencing could enable the generation of high quality training sets and negative data sets.

8. Current achievements and future directions

8.1. Achievements

In summary, antimicrobial peptide prediction is in essence a peptide classification problem. Different supervised learning algorithms have been trained to predict AMPs (Table 3). The major achievements include the following:

- Construction of AMP databases that facilitated machine learning prediction. The APD database, initially online in 2003 and updated regularly, provides a platform for understanding the structure and activity relationship of natural AMPs.
- Generation of hypothetically negative data sets based on UniProt.
- Successful encoding peptide features for machine learning prediction.
- Programming of various machine-learning algorithms with more or less similar prediction outcomes.
- Execution of both single and multi-label predictions as well as ensemble predictions of AMPs.

- Consideration of the impact of the peptide sequence in addition to amino acid composition.
- Consideration of post-translational modifications and 3D structure of AMPs.
- Species-specific prediction of AMPs.

8.2. Future directions

Machine learning prediction of AMPs remains a challenging task. The success rate is modest and not yet perfect because numerous factors are in play. We anticipate that the quality of AMP prediction will improve with the development of the following aspects:

- More complete positive data set for AMPs from continued peptide search and database update. There are two types of positive data. First, a continued expansion of *natural AMPs* in the APD will increase the accuracy of identifying natural AMP sequences. Second, data merging from different databases are anticipated to continue and a large data set with more and more *synthetic peptides* may improve the prediction of artificial sequences.
- Experimentally validated negative data sets for AMPs. Our ongoing collection of such peptides will reduce false positives in ML predictions.
- Ranking peptide activity data based on the same scale (e.g., MIC, diffusion distance, and E-test). This is a challenging task due to limited activity analysis under various lab conditions. A recommended guide for antimicrobial assays of AMPs may be helpful.
- Increased use of information about the target organism in classification and analysis of AMPs (e.g., Target is a Gram-positive vs Gram-negative bacteria, or a specific pathogen).
- Continued improvement of peptide encoding for rapid and accurate computing identification.
- Increased use of peptide information on chemical modifications and their relationship with activity.
- Increased high-quality 3D structures and their applications in AMP prediction. This is yet another challenging task as currently only ~13% AMPs are known to have 3D structures in the APD3 and high quality structures are not easy to obtain [11].
- Development of more powerful machine learning/artificial intelligence algorithms to better handle sequence and structural diversity and data imbalance of AMPs. Combined use of various ML models (i.e., ensemble) may improve predictions.
- Increased communication between AMP investigators and machine learning/AI scientists.

- Establishment of a pipeline of predictions of peptide properties required as a medicine by considering antimicrobial activity, cell toxicity *in vitro* and *in vivo*, and peptide bioavailability for efficacy *in vivo*.

Besides AMP prediction, another goal of the APD database is to help design novel peptides to combat antibiotic-resistant pathogens [9]. Different methods have been demonstrated [32]. The frequently occurring amino acids, such as glycine, leucine, and lysine, are sufficient in designing peptides with antibacterial activity comparable to human cathelicidin LL-37 [10,13]. Interestingly, a substitution of leucine in the database designed peptide DFTamP1 with isoleucine or valine led to activity or solubility decrease [156], underscoring the significance of nature's choice of leucine as a frequently occurring amino acid in AMPs [10]. Also, there is an inverse correlation between peptide length and leucine content of over 1000 amphibian peptides in the APD [160]. Our screening of representative peptides from the APD led to the identification of different sets of AMPs against methicillin-resistant *Staphylococcus aureus* (MRSA) and HIV-1 [161,162]. The grammar approach emphasizes the unique sequences in the database and their combinations [14]. The database filtering technology (DFT) is an *ab initio* approach, thereby providing another avenue [156]. The database derived parameters are useful to make peptide mimics [163] or to design even short peptides to decrease the production cost [6]. Our expansion of the DFT from *in silico* filtering to *in vitro* and *in vivo* filtering establishes a pipeline for peptide discovery [157]. This idea can be harnessed to establish a pipeline of machine learning predictions to accelerate peptide discovery. When quantitative MIC values are used to train ML algorithm, it becomes possible to rank the peptide activity to identify most potent sequences [164]. Likewise, a subsequent counterselection can be conducted by ranking peptide toxicity to host cells (Table 10) so that less toxic peptides can be selected for experimental validation. Ultimately, one may be able to generate an expert system that automatically designs and produces personalized antimicrobials with designed activity spectrum and molecular target for patients to treat a particular pathogen-caused infection. The multiple functions of AMPs annotated in the APD3 imply other potential applications as well.

Acknowledgements

This study was supported by Joint Warfighter Medical Research Program (JWMRP) JW200188 (MVH) and the NIH grant R01GM138552 (GW). Thanks to Fahad Alsaab and Maxwell Tabarrok for assistance with the hemolytic data.

References

1. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, Scheld M, Spellberg B, Bartlett J (2009) Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin Infect Dis. 48:1–12. [PubMed: 19035777]
2. O'Neill J (2014) The Review on Antimicrobial Resistance: Tracking Drug resistant Infections Globally, UK.
3. Boman HG (2003) Antibacterial peptides: basic facts and emerging concepts. J Inter Med 254:197–215.
4. Mangoni ML, McDermott AM, Zasloff M (2016) Antimicrobial peptides and wound healing: biological and therapeutic considerations. Exp Dermatol 25:167–73. [PubMed: 26738772]
5. Hancock REW, Sahl HG (2006) Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat Biotechnol 24:1551–1557. [PubMed: 17160061]

6. Lakshmaiah Narayana J, Mishra B, Lushnikova T, Wu Q, Chhonker YS, Zhang Y, Zarena D, Salnikov ES, Dang X, Wang F, Murphy C, Foster KW, Gorantla S, Bechinger B, Murry DJ, Wang G (2020) Two distinct amphipathic peptide antibiotics with systemic efficacy. Proc Natl Acad Sci USA 117:19446–19454. [PubMed: 32723829]
7. Wang G, Narayana JL, Mishra B, Zhang Y, Wang F, Wang C, Zarena D, Lushnikova T, Wang X (2019) Design of Antimicrobial Peptides: Progress Made with Human Cathelicidin LL-37. Adv Exp Med Biol. 1117:215–240. [PubMed: 30980360]
8. Browne K, Chakraborty S, Chen R, Willcox MD, Black DS, Walsh WR, Kumar N (2020) A New Era of Antibiotics: The Clinical Potential of Antimicrobial Peptides. Int J Mol Sci. 21:7047.
9. Wang Z, Wang G (2004) APD: the antimicrobial peptide database. Nucleic Acids Res 32:D590–D592. [PubMed: 14681488]
10. Wang G, Li X, Wang Z (2009) The updated antimicrobial peptide database and its application in peptide design. Nucleic Acids Res 37:D933–D937. [PubMed: 18957441]
11. Wang G, Li X, Wang Z (2016) APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res 44:D1087–1093. [PubMed: 26602694]
12. Kreutzberger MA, Pokorny A, Almeida PF (2017) Daptomycin-Phosphatidylglycerol Domains in Lipid Membranes. Langmuir. 33:13669–13679. [PubMed: 29130685]
13. Wang G (2020) The antimicrobial peptide database provides a platform for decoding the design principles of naturally occurring antimicrobial peptides. Protein Sci. 29(1):8–18. [PubMed: 31361941]
14. Loose C, Jensen K, Rigoutsos I, Stephanopoulos G (2006) A linguistic model for the rational design of antimicrobial peptides. Nature. 443(7113):867–9. [PubMed: 17051220]
15. Wang G (2012) Post-translational Modifications of Natural Antimicrobial Peptides and Strategies for Peptide Engineering. Curr Biotechnol. 1:72–79. [PubMed: 24511461]
16. Wang G, Mishra B, Lau K, Lushnikova T, Golla R, Wang X (2015) Antimicrobial peptides in 2014. Pharmaceuticals (Basel). 8:123–50. [PubMed: 25806720]
17. Lata S, Sharma BK, Raghava GP (2007) Analysis and prediction of antibacterial peptides. BMC bioinformatics 8:263. [PubMed: 17645800]
18. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S (2010) CAMP: a useful resource for research on antimicrobial peptides. Nucleic Acids Res 38:D774–D780. [PubMed: 19923233]
19. Wang G (2015) Improved methods for classification, prediction, and design of antimicrobial peptides. Methods Mol Biol 1268:43–66. [PubMed: 25555720]
20. Wang G (ed) (2010) Antimicrobial Peptides: Discovery, Design and Novel Therapeutic Strategies, CABI, England. 2nd edition published in 2017.
21. Gudmundsson GH, Agerberth B, Odeberg J, Bergman T, Olsson B, Salcedo R (1996) The human gene FALL39 and processing of the cathelin precursor to the antibacterial peptide LL-37 in granulocytes. Eur J Biochem. 238:325–32. [PubMed: 8681941]
22. Sørensen O, Arnljots K, Cowland JB, Bainton DF, Borregaard N (1997) The human antibacterial cathelicidin, hCAP-18, is synthesized in myelocytes and metamyelocytes and localized to specific granules in neutrophils. Blood. 90:2796–803. [PubMed: 9326247]
23. Sørensen OE, Gram L, Johnsen AH, Andersson E, Bangsbøll S, Tjabringa GS, Hiemstra PS, Malm J, Egesten A, Borregaard N (2003) Processing of seminal plasma hCAP-18 to ALL-38 by gastricsin: a novel mechanism of generating antimicrobial peptides in vagina. J Biol Chem. 278(31):28540–6. [PubMed: 12759353]
24. de Jong A, van Heel AJ, Kok J, Kuipers OP (2010) BAGEL2: mining for bacteriocins in genomic data. Nucleic Acids Res 38:W647–W651. [PubMed: 20462861]
25. Blin K, Kazempour D, Wohlleben W, Weber T (2014) Improved lanthipeptide detection and prediction for antiSMASH. PLoS One. 9(2):e89420. [PubMed: 24586765]
26. Yount NY, Weaver DC, de Anda J, Lee EY, Lee MW, Wong GCL, Yeaman MR (2020) Discovery of Novel Type II Bacteriocins Using a New High-Dimensional Bioinformatic Algorithm. Front Immunol. 11:1873. [PubMed: 33013838]
27. Fjell CD, Hancock RE, Cherkasov A (2007) AMPer: a database and an automated discovery tool for antimicrobial peptides. Bioinformatics. 23:1148–55. [PubMed: 17341497]

28. Dos Santos-Silva CA, Zupin L, Oliveira-Lima M, Vilela LMB, Bezerra-Neto JP, Ferreira-Neto JR, Ferreira JDC, de Oliveira-Silva RL, Pires CJ, Aburjaile FF, de Oliveira MF, Kido EA, Crovella S, Benko-Iseppon AM (2020) Plant Antimicrobial Peptides: State of the Art, In Silico Prediction and Perspectives in the Omics Era. Bioinform Biol Insights. 14:1177932220952739.
29. Jia HP, Mills JN, Barahmand-Pour F, Nishimura D, Mallampali RK, Wang G, Wiles K, Tack BF, Bevins CL, McCray PB Jr. (1999) Molecular cloning and characterization of rat genes encoding homologues of human beta-defensins. Infect Immun. 67:4827–33. [PubMed: 10456937]
30. Wang CK, Kaas Q, Chiche L, Craik DJ (2008) CyBase: a database of cyclic protein sequences and structures, with applications in protein discovery and engineering. Nucleic Acids Res. 36:D206–10. [PubMed: 17986451]
31. Yount NY, Andrés MT, Fierro JF, Yeaman MR (2007) The gamma-core motif correlates with antimicrobial activity in cysteine-containing kaliocin-1 originating from transferrins. Biochim Biophys Acta. 1768(11):2862–72. [PubMed: 17916323]
32. Wang G (2013) Database-Guided Discovery of Potent Peptides to Combat HIV-1 or Superbugs. Pharmaceuticals (Basel). 6(6):728–58. [PubMed: 24276259]
33. Pirtskhalava M, et al. (2021) DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. Nucleic Acids Res 49: D288–D297. [PubMed: 33151284]
34. Seebah S et al. (2007) Defensins knowledgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. Nucleic Acids Res 35: D265–268. [PubMed: 17090586]
35. Di Luca M, et al. (2015) BaAMPs: the database of biofilm-active antimicrobial peptides. Biofouling 31, 193–199. [PubMed: 25760404]
36. Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I (2010) BACTIBASE second release: a database and tool platform for bacteriocin characterization. BMC Microbiol 10: 22. [PubMed: 20105292]
37. Novković M, Simunić J, Bojović V, Tossi A, Juretić D (2012) DADP: the database of anuran defense peptides. Bioinformatics 28: 1406–1407. [PubMed: 22467909]
38. Kang X, et al. DRAMP 2.0, an updated data repository of antimicrobial peptides. Sci Data 6: 148 (2019). [PubMed: 31409791]
39. Whitmore L, Wallace BA (2004) The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols. Nucleic Acids Res 32: D593–594. [PubMed: 14681489]
40. Zhao X, Wu H, Lu H, Li G, Huang Q (2013) LAMP: A Database Linking Antimicrobial Peptides. PLoS One 8: e66557. [PubMed: 23825543]
41. Piotto SP, Sessa L, Concilio S, Iannelli P (2012) YADAMP: yet another database of antimicrobial peptides. Int J Antimicrob Agents 39: 346–351. [PubMed: 22325123]
42. Hammami R, Ben Hamida J, Vergoten G, Fliss I (2009) PhytAMP: a database dedicated to antimicrobial plant peptides. Nucleic Acids Res. 37(Database issue):D963–8. [PubMed: 18836196]
43. Gómez EA, Giraldo P, Orduz S (2017) InverPep: A database of invertebrate antimicrobial peptides. J Glob Antimicrob Resist. 8:13–17. [PubMed: 27888793]
44. Qureshi A, Thakur N, Kumar M (2013) HIPdb: a database of experimentally validated HIV inhibiting peptides. PLoS One 8: e54908. [PubMed: 23359817]
45. Li J, Qu X, He X, Duan L, Wu G, Bi D, Deng Z, Liu W, Ou HY. (2012) ThioFinder: a web-based tool for the identification of thiopeptide gene clusters in DNA sequences. PLoS One. 7(9):e45878. [PubMed: 23029291]
46. Wu H, Lu H, Huang J, et al. (2012) EnzyBase: a novel database for enzybiotic studies.[J]. Bmc Microbiology 12(1):54. [PubMed: 22489867]
47. Mehta D, Anand P, Kumar V, Joshi A, Mathur D, Singh S, Tuknait A, Chaudhary K, Gautam SK, Gautam A, Varshney GC, Raghava GP. (2014) ParaPep: a web resource for experimentally validated antiparasitic peptide sequences and their structures. Database (Oxford). 2014:bau051. [PubMed: 24923818]

48. Jhong JH, Chi YH, Li WC et al. (2019) dbAMP: an integrated resource for exploring antimicrobial peptides with functional activities and physicochemical properties on transcriptome and proteome data, Nucleic Acids Res 47:D285–D297. [PubMed: 30380085]
49. Usmani SS, Kumar R, Kumar V, Singh S, Raghava GPS. (2018) AntiTbPdb: a knowledgebase of anti-tubercular peptides. Database (Oxford). 2018:bay025.
50. Brahmachary M, Krishnan SP, Koh JL, Khan AM, Seah SH, Tan TW, Brusic V, Bajic VB. ANTIMIC: a database of antimicrobial sequences. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D586–9. [PubMed: 14681487]
51. Wang G (2015) Database resources dedicated to antimicrobial peptides. In "Antimicrobial Resistance and Food Safety" (editors: Chen C, Yan X, and Jackson CR), Academic Press, pp. 365–384.
52. Xiao X, Wang P, Lin WZ, Jia JH, Chou KC (2013) iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. Anal Biochem 436:168–177. [PubMed: 23395824]
53. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. (2016) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res. 44(D1):D1094–7. [PubMed: 26467475]
54. Wang P, Hu L, Liu G, Jiang N, Chen X, Xu J, Zheng W, Li L, Tan M, Chen Z, Song H, Cai YD, Chou KC. (2011) Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. PLoS One. 6(4):e18476. [PubMed: 21533231]
55. Torrent M, Di Tommaso P, Pulido D, Nogués MV, Notredame C, Boix E, Andreu D. (2012) AMPA: an automated web server for prediction of protein antimicrobial regions. Bioinformatics. 28(1):130–1. [PubMed: 22053077]
56. Fernandes FC, Rigden DJ, Franco OL. (2012) Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application. Biopolymers. 98(4):280–7. [PubMed: 23193592]
57. Mooney C, Haslam NJ, Holton TA, Pollastri G, Shields DC. (2013) PeptideLocator: prediction of bioactive peptides in protein sequences. Bioinformatics. 29(9):1120–6. [PubMed: 23505299]
58. Ng XY, Rosdi BA, Shahrudin S. (2015) Prediction of antimicrobial peptides based on sequence alignment and support vector machine-pairwise algorithm utilizing LZ-complexity. Biomed Res Int. 2015:212715. [PubMed: 25802839]
59. Lee HT, Lee CC, Yang JR et al. (2015) A large-scale structural classification of antimicrobial peptides, Biomed Res Int. 2015:475062. [PubMed: 26000295]
60. Lin W, Xu D (2016) Imbalanced multi-label learning for identifying antimicrobial peptides and their functional types. Bioinformatics. 32(24):3745–3752. [PubMed: 27565585]
61. Meher PK, Sahu TK, Saini V, Rao AR (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. [PubMed: 28205576]
62. Bhadra P, Yan J, Li J, Fong S, Siu SWI (2018) AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. Sci Rep. 8:1697. [PubMed: 29374199]
63. Veltri D, Kamath U, Shehu A (2018) Deep learning improves antimicrobial peptide recognition. Bioinformatics. 34:2740–2747. [PubMed: 29590297]
64. Agrawal P, Raghava GPS (2018) Prediction of Antimicrobial Potential of a Chemically Modified Peptide From Its Tertiary Structure. Front Microbiol. 9:2551. [PubMed: 30416494]
65. Jhong JH, Chi YH, Li WC et al. (2019) dbAMP: an integrated resource for exploring antimicrobial peptides with functional activities and physicochemical properties on transcriptome and proteome data, Nucleic Acids Res 47:D285–D297. [PubMed: 30380085]
66. Gull S, Shamim N, Minhas F. (2019) AMAP: Hierarchical multi-label prediction of biologically active and antimicrobial peptides. Comput Biol Med. 107:172–181. [PubMed: 30831306]
67. Feng P, Wang Z, Yu X (2019) Predicting Antimicrobial Peptides by Using Increment of Diversity with Quadratic Discriminant Analysis Method, IEEE/ACM Trans Comput Biol Bioinform 16:1309–1312. [PubMed: 28212093]

68. Chung CR, Kuo TR, Wu LC et al. (2019) Characterization and identification of antimicrobial peptides with different functional activities, Brief Bioinform :bbz043. doi: 10.1093/bib/bbz043. [PubMed: 31155657]
69. Gull S, Minhas FUAA (2020) AMP0: Species-Specific Prediction of Anti-microbial Peptides using Zero and Few Shot Learning. IEEE/ACM Trans Comput Biol Bioinform. PP. doi: 10.1109/TCBB.2020.2999399.
70. Tripathi V, Tripathi P (2020) Detecting antimicrobial peptides by exploring the mutual information of their sequences. J Biomol Struct Dyn. 38:5037–5043. [PubMed: 31760879]
71. Yan J, Bhadra P, Li A, Sethiya P, Qin L, Tai HK, Wong KH, Siu SWI (2020) Deep-AmPEP30: Improve Short Antimicrobial Peptides Prediction with Deep Learning. Mol Ther Nucleic Acids. 20:882–894. [PubMed: 32464552]
72. Fu H, Cao Z, Li M et al. (2020) ACEP: improving antimicrobial peptides recognition through automatic feature fusion and amino acid embedding, BMC Genomics 21:597. [PubMed: 32859150]
73. Kavousi K, Bagheri M, Behrouzi S et al. (2020) IAMPE: NMR-Assisted Computational Prediction of Antimicrobial Peptides, J Chem Inf Model 60:4691–4701. [PubMed: 32946226]
74. Santos-Junior CD, Pan S, Zhao XM et al. (2020) Macrel: antimicrobial peptide screening in genomes and metagenomes, PeerJ 8:e10555. [PubMed: 33384902]
75. Youmans M, Spainhour JCG, Qiu P (2020) Classification of Antibacterial Peptides Using Long Short-Term Memory Recurrent Neural Networks, IEEE/ACM Trans Comput Biol Bioinform 17:1134–1140. [PubMed: 30843849]
76. Fingerhut L, Miller DJ, Strugnell JM et al. (2020) ampir: an R package for fast genome-wide prediction of antimicrobial peptides, Bioinformatics. 36:5262–5263
77. Lawrence TJ, Carper DL, Spangler MK, et al. amPEPpy 1.0: A portable and accurate antimicrobial peptide prediction tool, Bioinformatics 2020. doi: 10.1093/bioinformatics/btaa917.
78. Lertampaiporn S, Vorapreeda T, Hongsthong A, et al. (2021) Ensemble-AMPPred: Robust AMP Prediction and Recognition Using the Ensemble Learning Method with a New Hybrid Feature for Differentiating AMPs. Genes (Basel) 12: 137. [PubMed: 33494403]
79. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res. 28(1):235–42. [PubMed: 10592235]
80. MacDougall A, et al. UniProt Consortium. (2020) UniRule: a unified rule resource for automatic annotation in the UniProt Knowledgebase. Bioinformatics. 36(17):4643–4648. [PubMed: 32399560]
81. Porto WF, Pires ÁS, Franco OL (2017) Antimicrobial activity predictors benchmarking analysis using shuffled and designed synthetic peptides. J Theor Biol. 426:96–103. [PubMed: 28536036]
82. Othman M, Ratna S, Tewari A, et al. (2017) Classification and Prediction of Antimicrobial Peptides Using N-gram Representation and Machine Learning. Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. Boston, Massachusetts, USA: Association for Computing Machinery, 605.
83. Mooney C, Haslam NJ, Pollastri G, et al. (2012) Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity, PLoS One 7:e45012. [PubMed: 23056189]
84. Burdukiewicz M, Sidorczuk K, Rafacz D et al. (2020) Proteomic Screening for Prediction and Design of Antimicrobial Peptides with AmpGram, Int J Mol Sci 21.
85. Kaplan N, Morpurgo N, Linial M (2007) Novel families of toxin-like peptides in insects and mammals: a computational approach, J Mol Biol 369:553–566. [PubMed: 17433819]
86. Kavousi K, Bagheri M, Behrouzi S, et al. (2020) IAMPE: NMR-Assisted Computational Prediction of Antimicrobial Peptides, J Chem Inf Model 60:4691–4701. [PubMed: 32946226]
87. Youmans M, Spainhour JCG, Qiu P. (2020) Classification of Antibacterial Peptides Using Long Short-Term Memory Recurrent Neural Networks, IEEE/ACM Trans Comput Biol Bioinform 17:1134–1140. [PubMed: 30843849]
88. Muller AT, Kaymaz AC, Gabernet G, et al. (2016) Sparse Neural Network Models of Antimicrobial Peptide-Activity Relationships, Mol Inform 35:606–614. [PubMed: 27870247]

89. Schneider P, Muller AT, Gabernet G, et al. (2017) Hybrid Network Model for “Deep Learning” of Chemical Data: Application to Antimicrobial Peptides, Mol Inform 36.

90. Su X, Xu J, Yin Y, et al. (2019) Antimicrobial peptide identification using multi-scale convolutional network, BMC Bioinformatics 20:730. [PubMed: 31870282]

91. Tripathi S, et al. (2015) Antiviral Activity of the Human Cathelicidin, LL-37, and Derived Peptides on Seasonal and Pandemic Influenza A Viruses. PLoS One 10: e0124706. [PubMed: 25909853]

92. Barlow PG, et al. (2011) Antiviral activity and increased host defense against influenza infection elicited by the human cathelicidin LL-37. PLoS One 6: e25333. [PubMed: 22031815]

93. Wang G (2012) Natural antimicrobial peptides as promising anti-HIV candidates. Curr Top Pept Protein Res. 13:93–110. [PubMed: 26834391]

94. Wohlford-Lenane CL, et al. (2009) Rhesus theta-defensin prevents death in a mouse model of severe acute respiratory syndrome coronavirus pulmonary disease. J Virol 83: 11385–11390. [PubMed: 19710146]

95. He M, Zhang H, Li Y, Wang G, Tang B, Zhao J, Huang Y, Zheng J. (2018) Cathelicidin-Derived Antimicrobial Peptides Inhibit Zika Virus Through Direct Inactivation and Interferon Pathway. Front Immunol. 9:722. [PubMed: 29706959]

96. Yu Y, et al. (2020) Engineered Human Cathelicidin Antimicrobial Peptides Inhibit Ebola Virus Infection. iScience 23: 100999. [PubMed: 32252021]

97. Thakur N, Qureshi A, Kumar M (2012) AVPpred: collection and prediction of highly effective antiviral peptides. Nucleic Acids Res 40: W199–204. [PubMed: 22638580]

98. Qureshi A, Thakur N, Tandon H, Kumar M (2014) AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. Nucleic Acids Res 42: D1147–1153. [PubMed: 24285301]

99. Chowdhury AS, et al. (2020) Better understanding and prediction of antiviral peptides through primary and secondary structure feature importance. Sci Rep 10: 19260. [PubMed: 33159146]

100. Tyagi A, et al. (2019) PlantAFP: a curated database of plant-origin antifungal peptides. Amino Acids 51, 1561–1568. [PubMed: 31612325]

101. Agrawal P, et al. (2018) In Silico Approach for Prediction of Antifungal Peptides. Front Microbiol 9: 323. [PubMed: 29535692]

102. Manavalan B, et al. (2018) AIPpred: Sequence-Based Prediction of Anti-inflammatory Peptides Using Random Forest. Front Pharmacol 9: 276. [PubMed: 29636690]

103. Manavalan B, et al. (2018) PIP-EL: A New Ensemble Learning Method for Improved Proinflammatory Peptide Predictions. Front Immunol 9: 1783. [PubMed: 30108593]

104. Usmani SS, Bhalla S, Raghava GPS (2018) Prediction of Antitubercular Peptides From Sequence Information Using Ensemble Classifier and Hybrid Features. Front Pharmacol 9: 954. [PubMed: 30210341]

105. Gupta K, Singh S, van Hoek ML (2015) Short, Synthetic Cationic Peptides Have Antibacterial Activity against Mycobacterium smegmatis by Forming Pores in Membrane and Synergizing with Antibiotics. Antibiotics (Basel) 4: 358–378. [PubMed: 27025629]

106. Torres-Juarez F, et al. (2015) LL-37 immunomodulatory activity during *Mycobacterium tuberculosis* infection in macrophages. Infect Immun 83: 4495–4503. [PubMed: 26351280]

107. Rao Muvva J, et al. (2019) Polarization of Human Monocyte-Derived Cells With Vitamin D Promotes Control of *Mycobacterium tuberculosis* Infection. Front Immunol 10: 3157. [PubMed: 32038652]

108. Rivas-Santiago B, et al. (2013) Activity of LL-37, CRAMP and antimicrobial peptide-derived compounds E2, E6 and CP26 against Mycobacterium tuberculosis. Int J Antimicrob Agents 41: 143–148. [PubMed: 23141114]

109. Corrales-Garcia L et al. (2013) Bacterial expression and antibiotic activities of recombinant variants of human beta-defensins on pathogenic bacteria and *M. tuberculosis*. Protein Expr Purif 89: 33–43. [PubMed: 23459290]

110. Wong GC, O’Toole GA (2011) All together now: Integrating biofilm research across disciplines. MRS Bull 36: 339–342. [PubMed: 24465088]

111. O’Toole GA (2003) To build a biofilm. J Bacteriol 185: 2687–2689. [PubMed: 12700246]

112. O'Toole GA (2011) Microtiter dish biofilm formation assay. J Vis Exp, (47):2437. [PubMed: 21307833]
113. de la Fuente-Nunez C, et al. (2014) Broad-spectrum anti-biofilm peptide that targets a cellular stress response. PLoS Pathog 10: e1004152. [PubMed: 24852171]
114. de la Fuente-Nunez C, et al. , (2012) Inhibition of bacterial biofilm formation and swarming motility by a small synthetic cationic peptide. Antimicrob Agents Chemother 56:2696–2704. [PubMed: 22354291]
115. Overhage J, et al. (2008) Human host defense peptide LL-37 prevents bacterial biofilm formation. Infect Immun 76: 4176–4182. [PubMed: 18591225]
116. Chung EMC, et al. (2017) Komodo dragon-inspired synthetic peptide DRGN-1 promotes wound-healing of a mixed-biofilm infected wound. NPJ Biofilms Microbiomes 3: 9. [PubMed: 28649410]
117. Duplantier AJ, van Hoek ML (2013) The Human Cathelicidin Antimicrobial Peptide LL-37 as a Potential Treatment for Polymicrobial Infected Wounds. Front Immunol 4: 143. [PubMed: 23840194]
118. Dean SN, Bishop BM, van Hoek ML (2011) Susceptibility of Pseudomonas aeruginosa Biofilm to Alpha-Helical Peptides: D-enantiomer of LL-37. Front Microbiol 2: 128. [PubMed: 21772832]
119. Dean SN, Bishop BM, van Hoek ML (2011) Natural and synthetic cathelicidin peptides with anti-microbial and anti-biofilm activity against *Staphylococcus aureus*. BMC Microbiol 11: 114. [PubMed: 21605457]
120. Amer LS, B. Bishop BM, van Hoek ML(2010) Antimicrobial and antibiofilm activity of cathelicidins and short, synthetic peptides against Francisella. Biochem Biophys Res Commun 396: 246–251. [PubMed: 20399752]
121. Sharma A, et al. (2016) dPABBs: A Novel in silico Approach for Predicting and Designing Anti-biofilm Peptides. Sci Rep 6: 21839. [PubMed: 26912180]
122. Fallah F, et al. (2020) BIPEP: Sequence-based Prediction of Biofilm Inhibitory Peptides Using a Combination of NMR and Physicochemical Descriptors. ACS Omega 5: 7290–7297. [PubMed: 32280870]
123. Gupta S, et al. (2016) Prediction of Biofilm Inhibiting Peptides: An In silico Approach. Front Microbiol 7: 949. [PubMed: 27379078]
124. Rajput A, Thakur A, Sharma S, Kumar M (2018) aBiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance. Nucleic Acids Res 46: D894–D900. [PubMed: 29156005]
125. Bishop BM, Juba ML, Devine MC, Barksdale SM, Rodriguez CA, Chung MC, Russo PS, Vliet KA, Schnur JM, van Hoek ML (2015) Bioprospecting the American alligator (Alligator mississippiensis) host defense peptidome. PLoS One. 10:e0117394. [PubMed: 25671663]
126. Cherkasov A, Hilpert K, Jenssen H, Fjell CD, Waldbrook M, Mullaly SC, Volkmer R, Hancock RE (2009) Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. ACS Chem Biol. 4(1):65–74. [PubMed: 19055425]
127. Wang X, Mishra B, Lushnikova T, Narayana JL, Wang G (2018) Amino Acid Composition Determines Peptide Activity Spectrum and Hot-Spot-Based Design of Merecidin. Adv Biosyst. 2(5):1700259. [PubMed: 30800727]
128. Gupta S, et al. (2013) In silico approach for predicting toxicity of peptides and proteins. PLoS One 8: e73957. [PubMed: 24058508]
129. Gupta S, et al. (2015) Peptide toxicity prediction. Methods Mol Biol 1268: 143–157. [PubMed: 25555724]
130. Gordon YJ, et al. (2005) Human cathelicidin (LL-37), a multifunctional peptide, is expressed by ocular surface epithelia and has potent antibacterial and antiviral activity. Curr Eye Res 30: 385–394. [PubMed: 16020269]
131. Wang W, et al. (2017) Antimicrobial peptide LL-37 promotes the viability and invasion of skin squamous cell carcinoma by upregulating YB-1. Exp Ther Med 14: 499–506. [PubMed: 28672959]

132. Oliveira-Bravo M, et al. (2016) LL-37 boosts immunosuppressive function of placenta-derived mesenchymal stromal cells. Stem Cell Res Ther 7: 189. [PubMed: 28038684]
133. Hosseini Z, et al. (2020) The Human Cathelicidin LL-37, a Defensive Peptide Against Rotavirus Infection. International Journal of Peptide Research and Therapeutics 26: 911–919.
134. Haisma EM, et al. (2014) LL-37-derived peptides eradicate multidrug-resistant Staphylococcus aureus from thermally wounded human skin equivalents. Antimicrob Agents Chemother 58: 4411–4419. [PubMed: 24841266]
135. Barksdale SM, Hrifko EJ, van Hoek ML (2017) Cathelicidin antimicrobial peptide from Alligator mississippiensis has antibacterial activity against multi-drug resistant Acinetobacter baumanii and Klebsiella pneumoniae. Dev Comp Immunol 70: 135–144. [PubMed: 28089718]
136. Barksdale SM, Hrifko EJ, Chung EM, van Hoek ML (2016) Peptides from American alligator plasma are antimicrobial against multi-drug resistant bacterial pathogens including *Acinetobacter baumannii*. BMC Microbiol 16: 189. [PubMed: 27542832]
137. Hitt SJ, Bishop BM, van Hoek ML. (2020)Komodo-dragon cathelicidin-inspired peptides are antibacterial against carbapenem-resistant Klebsiella pneumoniae. J Med Microbiol 69, 1262–1272. [PubMed: 33084564]
138. de Latour FA, et al. (2010) Antimicrobial activity of the Naja atra cathelicidin and related small peptides. Biochemical and biophysical research communications 396: 825–830. [PubMed: 20438706]
139. van Dijk A, et al. (2009) Identification of chicken cathelicidin-2 core elements involved in antibacterial and immunomodulatory activities. Mol Immunol 46: 2465–2473. [PubMed: 19524300]
140. Nizet V, et al. (2001) Innate antimicrobial peptide protects the skin from invasive bacterial infection. Nature 414: 454–457. [PubMed: 11719807]
141. Gao J, et al. (2020) Design of a Sea Snake Antimicrobial Peptide Derivative with Therapeutic Potential against Drug-Resistant Bacterial Infection. ACS Infect Dis 6: 2451–2467. [PubMed: 32786271]
142. Win TS, et al. (2017) HemoPred: a web server for predicting the hemolytic activity of peptides. Future Med Chem 9: 275–291. [PubMed: 28211294]
143. Chaudhary K, et al. (2016) A Web Server and Mobile App for Computing Hemolytic Potency of Peptides. Sci Rep 6: 22843. [PubMed: 26953092]
144. Timmons PB, Hewage CM (2020) HAPPENN is a novel tool for hemolytic activity prediction for therapeutic peptides which employs neural networks. Sci Rep 10: 10869. [PubMed: 32616760]
145. Oren Z, et al. (1999) Structure and organization of the human antimicrobial peptide LL-37 in phospholipid membranes: relevance to the molecular basis for its non-cell-selective activity. Biochem J 341 (Pt 3): 501–513. [PubMed: 10417311]
146. Ciornei CD, Sigurdardottir T, Schmidtchen A, Bodelsson M (2005) Antimicrobial and chemoattractant activity, lipopolysaccharide neutralization, cytotoxicity, and inhibition by serum of analogs of human cathelicidin LL-37. Antimicrob Agents Chemother 49: 2845–2850. [PubMed: 15980359]
147. Al-Adwani S, et al. (2020) Studies on citrullinated LL-37: detection in human airways, antibacterial effects and biophysical properties. Sci Rep 10: 2376. [PubMed: 32047184]
148. Rajasekaran G, Kim EY, Shin SY (2017) LL-37-derived membrane-active FK-13 analogs possessing cell selectivity, anti-biofilm activity and synergy with chloramphenicol and anti-inflammatory activity. Biochim Biophys Acta Biomembr 1859: 722–733. [PubMed: 28161291]
149. Luo Y, et al. (2017) The Naturally Occurring Host Defense Peptide, LL-37, and Its Truncated Mimetics KE-18 and KR-12 Have Selected Biocidal and Antibiofilm Activities Against Candida albicans, Staphylococcus aureus, and Escherichia coli In vitro. Front Microbiol 8: 544. [PubMed: 28408902]
150. Koro C, et al. (2016) Carbamylated LL-37 as a modulator of the immune response. Innate Immun 22: 218–229. [PubMed: 26878866]
151. Murakami M, et al. (2004) Postsecretory processing generates multiple cathelicidins for enhanced topical antimicrobial defense. J Immunol 172: 3070–3077. [PubMed: 14978112]

152. Chung MC, Dean SN, van Hoek ML (2015) Acyl carrier protein is a bacterial cytoplasmic target of cationic antimicrobial peptide LL-37. Biochem J 470: 243–253. [PubMed: 26188040]

153. Limoli DH, et al. (2014) Cationic antimicrobial peptides promote microbial mutagenesis and pathoadaptation in chronic infections. PLoS Pathog 10: e1004083. [PubMed: 24763694]

154. Limoli DH, Wozniak DJ (2014) Mutagenesis by host antimicrobial peptides: insights into microbial evolution during chronic infections. Microb Cell 1: 247–249. [PubMed: 28357249]

155. Oikawa K, et al. (2018) Screening of a Cell-Penetrating Peptide Library in Escherichia coli: Relationship between Cell Penetration Efficiency and Cytotoxicity. ACS Omega 3: 16489–16499.

156. Mishra B, Wang G (2012) *Ab initio* design of potent anti-MRSA peptides based on database filtering technology. J Am Chem Soc. 134(30):12426–9. [PubMed: 22803960]

157. Mishra B, Lakshmaiah Narayana J, Lushnikova T, Wang X, Wang G (2019) Low cationicity is important for systemic in vivo efficacy of database-derived peptides against drug-resistant Gram-positive pathogens. Proc Natl Acad Sci U S A. 116(27):13517–13522. [PubMed: 31209048]

158. Beheshtirouy S, Mirzaei F, Eyvazi S, Tarhriz V (2020) Recent Advances on Therapeutic Peptides for Breast Cancer Treatment. Curr Protein Pept Sci. doi: 10.2174/1389203721999201117123616.

159. Marqus S, Pirogova E, Piva TJ (2017) Evaluation of the use of therapeutic peptides for cancer treatment. J Biomed Sci 24: 21. [PubMed: 28320393]

160. Wang G (2020) Bioinformatic Analysis of 1000 Amphibian Antimicrobial Peptides Uncovers Multiple Length-Dependent Correlations for Peptide Design and Prediction. Antibiotics (Basel). 9(8):491.

161. Wang G, Watson KM, Peterkofsky A, Buckheit RW Jr. (2010) Identification of novel human immunodeficiency virus type 1-inhibitory peptides based on the antimicrobial peptide database. Antimicrob Agents Chemother. 54(3):1343–6. [PubMed: 20086159]

162. Menousek J, Mishra B, Hanke ML, Heim CE, Kielian T, Wang G (2012) Database screening and in vivo efficacy of antimicrobial peptides against methicillin-resistant *Staphylococcus aureus* USA300. Int J Antimicrob Agents. 39(5):402–6. [PubMed: 22445495]

163. Dong Y, Lushnikova T, Golla RM, Wang X, Wang G. (2017) Small molecule mimics of DFTamP1, a database designed anti-Staphylococcal peptide. Bioorg Med Chem. 25(3):864–869. [PubMed: 28011203]

164. Witten J, Witten Z (2019) Deep learning regression model for antimicrobial peptide design. bioRxiv. A preprint posted on July 12, 2019.

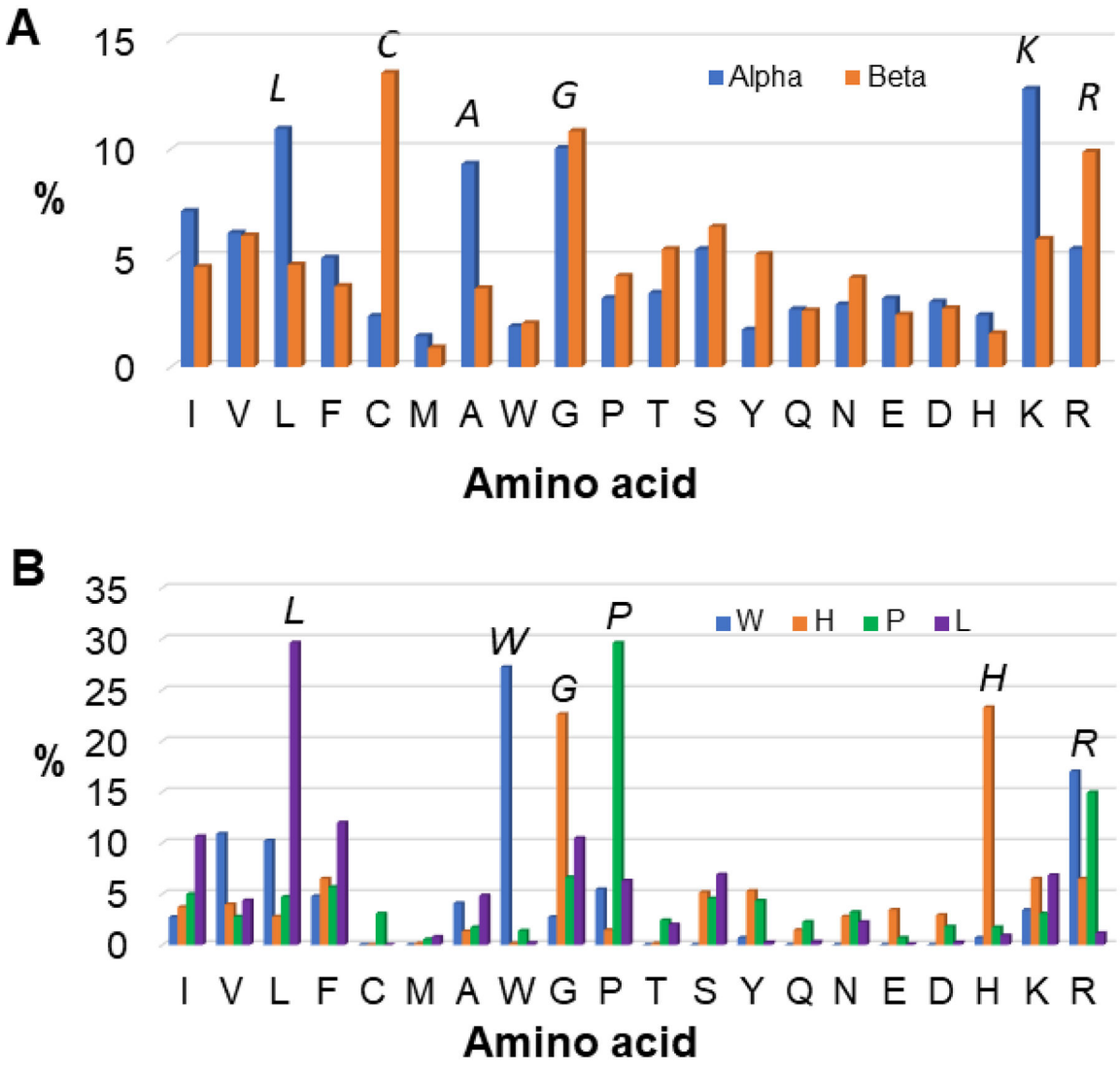

Fig. 1.

Important amino acids derived from amino acid composition profiles of classic classes of antimicrobial peptides [3]: (A) α-helical and β-sheet families and (B) amino acid-rich families, including Trp-rich, His-rich, Pro-rich, and Leu-rich AMPs. Data obtained in the APD [13] in Dec 2020.

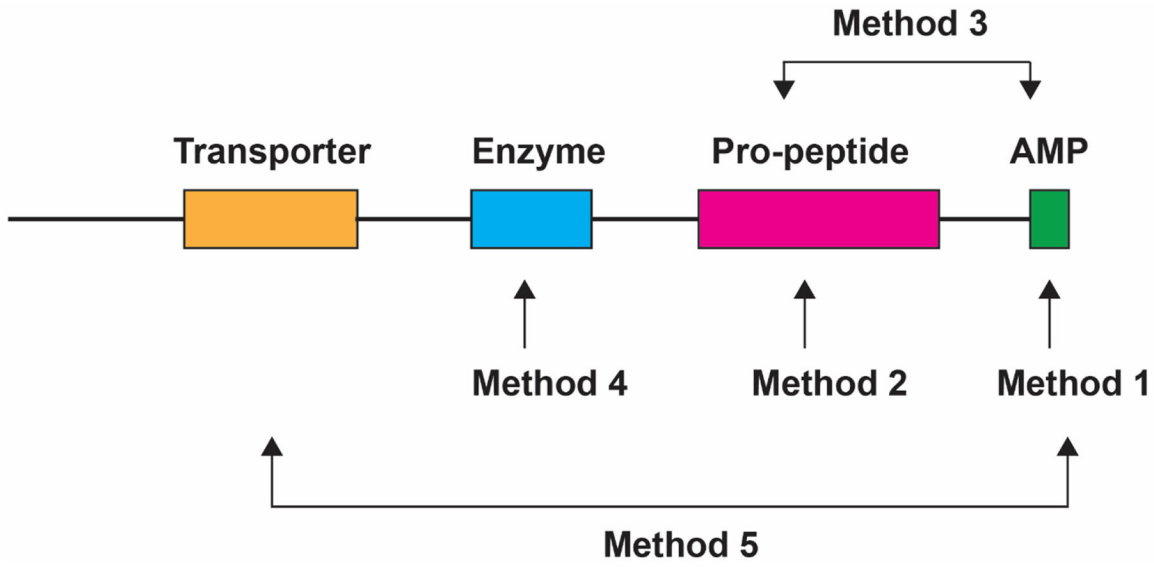

Fig. 2.
Information-content based five methods for prediction of antimicrobial peptides [20].

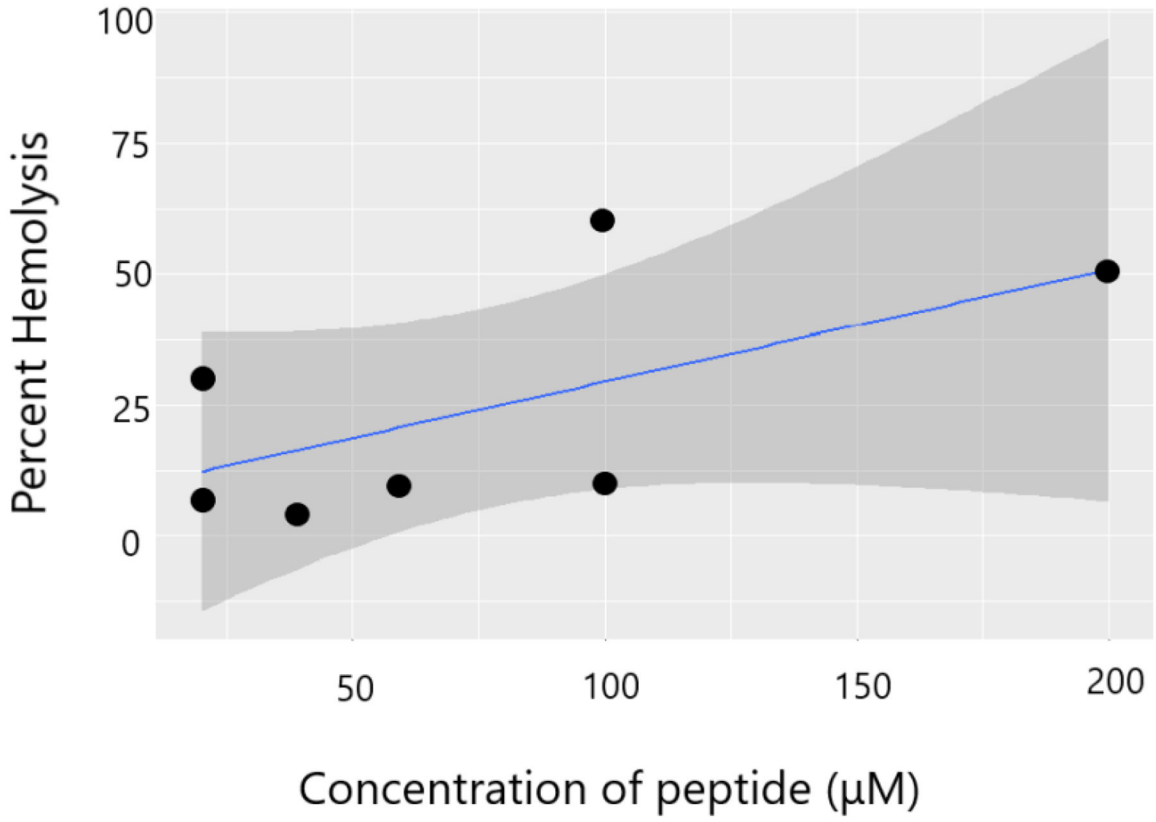

Fig. 3:

Percent hemolysis results with different amounts of LL-37 peptide against human red blood cells. The data from Table 11 were plotted. The best-fit line is y=0.2142x + 8.0017. The shaded grey area represents a 95% confidence interval.

Table 1.

Amino acid properties, frequency and peptide count in the antimicrobial peptide database (APD)

Single letter	Full name	Molecular weight	Class[a]	Peptide count	Count% (2020)	Frequency in 3257 AMPs
I	Isoleucine	113.16	phobic	2511	0.77	5.9%
V	Valine	99.13	phobic	2492	0.76	5.69%
L	Leucine	113.16	phobic	2835	**0.87**	**8.26%**
F	Phenyl alanine	147.18	phobic	2240	0.69	4.09%
C	Cysteine	103.14	phobic	1721	0.53	6.81%
M	Methionine	131.2	phobic	959	0.29	1.27%
A	Alanine	71.08	phobic	2511	0.77	7.68%
W	Tryptophan	186.21	phobic	1185	0.36	1.65%
G	Glycine	57.05	special	2950	**0.91**	**11.51%**
P	Proline	97.12	special	1958	0.60	4.67%
T	Threonine	101.11	polar	2053	0.63	4.48%
S	Serine	87.08	polar	2483	0.76	6.07%
Y	Tyrosine	163.18	polar	1266	0.39	2.49%
Q	Glutamine	128.13	polar	1352	0.42	2.59%
N	Asparagine	114.1	polar	1968	0.60	3.86%
E	Glutamate acid	129.12	acidic	1465	0.45	2.68%
D	Aspartic acid	115.09	acidic	1463	0.45	2.7%
H	Histidine	137.14	basic	1231	0.38	2.17%
K	Lysine	128.17	basic	2782	**0.85**	**9.51%**
R	Arginine	156.19	basic	1843	0.57	5.88%

[a]phobic=hydrophobic. In the APD, the hydrophobic content (Pho) is the ratio between the total hydrophobic amino acids and total amino acids in a peptide sequence [9]. Visited January 2021.

Table 2.

Web accessible databases dedicated to antimicrobial peptides[a]

Databases & Prediction algorithms	Link	Notes	Citing References
APD3	http://aps.unmc.edu/AP/main.php	Antimicrobial peptide database, with curated, experimentally verified antimicrobial peptides from bacteria, archaea, protists, fungi, plants, and animals	[11]
CAMPR3	http://www.camp3.bicnirrh.res.in/	Collection of Anti-microbial peptides	[18]
DBAASP v3	https://dbaasp.org	Database of antimicrobial activity and structure of peptides	[33]
Defensins knowledgebase	http://defensins.bii.a-star.edu.sg/	Antimicrobial peptides from the defensin family	[34]
BaAMPs	http://www.baamps.it/	Database of biofilm-active antimicrobial peptides	[35]
BACTIBASE	http://bactibase.hammamilab.org/about.php	Bacterocin type naturally occurring antimicrobial peptides.	[36]
DADP	http://split4.pmfst.hr/dadp/	Database of anuran (frog or toad) defense peptides	[37]
DRAMP	http://dramp.cpu-bioinfor.org	Database of AMPs including clinical trial data on peptides.	[38]
Peptaibol	http://peptaibol.cryst.bbk.ac.uk/introduction.htm	Database of Peptaibols, mainly antifungal peptides.	[39]
LAMP	http://biotechlab.fudan.edu.cn/database/lamp/index.php	AMPs taken from other databases	[40]
YADAMP	http://www.yadamp.unisa.it/default.aspx	Yet another database of antimicrobial peptides	[41]
PhytAMP	http://phytamp.pfba-lab-tun.org/main.php	A database dedicated to plant AMPs	[42]
InverPep	https://ciencias.medellin.unal.edu.co/gruposdeinvestigacion/prospeccionydisenobiomoleculas/InverPep/public/home_en	AMPs from invertebrates from other databases	[43]
HIPdb	http://crdd.osdd.net/servers/hipdb	Manually curated database of experimentally validated HIV inhibitory peptides	[44]
Thiobase	https://db-mml.sjtu.edu.cn/THIOBASE/	Sulfur-rich, highly modified heterocyclic peptide antibiotics	[45]
EnzyBase	http://biotechlab.fudan.edu.cn/database/EnzyBase/home.php	lysins, bacteriocins, autolysins, and lysozymes	[46]
ParaPep	http://crdd.osdd.net/raghava/parapep/	Antiparasitic peptides	[47]
dbAMP	Not accessible	AMPs	[48]
AntiTbPdb	https://webs.iiitd.edu.in/raghava/antitbpdb/	Anti-TB peptides	[49]

[a]Adapted and updated based on the APD Links [13,20].

Table 3.

Machine learning prediction of antimicrobial peptides

Tool name	URL	Algorithms	Features	Year	Ref
AntiBP	http://crdd.osdd.net/raghava/antibp2	SVM,QM,ANN	Single label	2007	[17]
CAMP	http://www.bicnirrh.res.in/antimicrobial	SVM, RF, DA	Single label	2010	[18, 53]
	http://amp.biosino.org/	BLASTP, NNA	Single label	2011	[54]
AMPA	http://tcoffee.crg.cat/apps/ampa		AMP region scan	2012	[55]
ANFIS		ANFIS	Single label	2012	[56]
Peptide Locator	http://bioware.ucd.ie/	BRNN	Single label	2013	[57]
iAMP-2L	http://www.jci-bioinfo.cn/iAMP-2L	FKNN	Two-level, Multi-label	2013	[52]
DBAASP	https://dbaasp.org/prediction/general	thresholds		2014	[33]
SVM-LZ	NG (BioMed Research International)	SVM	Single label	2015	[58]
ADAM	http://bioinformatics.cs.ntou.edu.tw/ADAM/	SVM, HMM	Single label	2015	[59]
MLAMP	http://www.jci-bioinfo.cn/MLAMP	RF – ML-SMOTE	Multi-label	2016	[60]
*i*AMPpred	http://cabgrid.res.in:8080/amppred/	SVM	Single label	2017	[61]
AmPEP	http://cbbio.cis.umac.mo/software/AmPEP/	RF	Single label	2018	[62]
AMP scanner	www.ampscanner.com	DNN	Single label, Large scale	2018	[63]
AntiMPmod	https://webs.iiitd.edu.in/raghava/antimpmod/	SVM	Single label, PTM/3D	2018	[64]
dbAMP	http://csb.cse.yzu.edu.tw/dbAMP/	RF	Single label	2019	[65]
AMAP	http://faculty.pieas.edu.pk/fayyaz/software.html#AMAP	SVM, XGBoost	Multi-label	2019	[66]
	NA	IDQD	Single label	2019	[67]
AMPfun	http://fdblab.csie.ncu.edu.tw/AMPfun/index.html	CART	Multi-label	2020	[68]
AMP0	http://ampzero.pythonanywhere.com	ZSL, FSL	Single label, Species-specific	2020	[69]
MIV-RF	NA	RF	Single label, Sequence	2020	[70]
Deep-AmPEP30	https://cbbio.cis.um.edu.mo/AxPEP	CNN	Genome search	2020	[71]
ACEP	https://github.com/Fuhaoyi/ACEP	DNN	high-throughput predictions	2020	[72]
IAMPE	http://cbb1.ut.ac.ir/	KNN, SVM, RF	Single label	2020	[73]
Macrel	https://big-data-biology.org/software/macrel.	RF	Genome search	2020	[74]
	https://github.com/mtyoumans/lstm_peptides	LSTM RNN	Single label	2020	[75]
ampir	https://github.com/legana/ampir	SVM	Genome wide	2020	[76]
amPEPpy	https://github.com/tlawrence3/amPEPpy	RF	Genome wide	2020	[77]
Ensemble-AMPPred		Ensemble model	Single label	2021	[78]

Table 4:

Prediction algorithm websites for antiviral peptides (AVPs).

Prediction algorithms	Link	Notes	Ref
AVPPred	http://crdd.osdd.net/servers/avppred/	Webserver for collecting and detecting effective AVPs	[97]
AVPdb	http://crdd.osdd.net/servers/avpdb	A database of experimentally validated anti-viral peptides.	[98]
FIRM-AVP	https://msc-viz.emsl.pnnl.gov/AVPR	“Feature-Informed Reduced Machine Learning for Antiviral Peptide Prediction”	[99]

Table 5:

Prediction algorithm websites for antifungal peptides (AFPs).

Database	Link	Notes	Ref
PlantAFP	http://bioinformatics.cimap.res.in/sharma/PlantAFP/	Plant derived peptides	[100]
AntiFP	https://webs.iiitd.edu.in/raghava/antifp/algo.php		[101]

Table 6:

Prediction algorithm websites for other specific and unique kinds of peptides.

Databases & Prediction algorithms	Link	Notes	Ref
AIPred	www.thegleelab.org/AIPpred	Anti-Inflammatory Peptides	[102]
PIP-EL	www.thegleelab.org/PIP-EL	Pro-inflammatory peptide	[103]
AntiTBpred	http://webs.iiitd.edu.in/raghava/antitbpred/	Antitubercular Peptides	[104]

Table 7:

AntiTBpred output for the activity of LL-37 against tuberculosis.

Prediction Method	ID	Score	Prediction	ID	Score	Prediction
AntiTB_MD SVM ensemble	LL37	0.78	Anti-TB peptide	HBD2	−0.30	**Non** Anti-TB peptide
AntiTB_RD SVM ensemble	LL37	−0.25	**Non** Anti-TB peptide	HBD2	−0.202	**Non** Anti-TB peptide
AntiTB_MD Hybrid method	LL37	−0.25	**Non** Anti-TB peptide	HBD2	0.053	**Non** Anti-TB peptide
AntiTB_RD Hybrid method *	LL37	0.317	Anti-TB peptide	HBD2	0.673	Anti-TB peptide

Table 8:

Prediction algorithm websites for Antibiofilm peptides.

Databases & Prediction algorithms	Link	Notes	Ref
BaAMPs	http://www.baamps.it/	Database of biofilm-active antimicrobial peptides	[35]
dPABBs	http://ab-openlab.csir.res.in/abp/antibiofilm/	Predictor of antibiofilm activity of peptides, and generates possible peptide variants and predicts their antibiofilm activity.	[121]
BIPEP	http://cbb1.ut.ac.ir/BIPClassifier/Index	Uses NMR and physicochemical descriptors	[122]
BioFIN	http://metagenomics.iiserb.ac.in/biofin/ and http://metabiosys.iiserb.ac.in/biofin/		[123]

Table 9:

Hemolytic prediction of activity for LL-37 human cathelicidin peptide.

Table 9(A): Predicted Toxicity of LL-37 on ToxinPred (validated via ExPASy ProParam tool).

Peptide Sequence	SVM score	Prediction	Hydro-phobicity	Hydropathicity	Amphi-pathicity	Hydro-philicity	Net charge	pI	Mol wt
LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES	−1.58	Non-toxin	−0.34	−0.72	1.06	0.62	+6.0	10.61	4493.32

Table 9(B): Experimental cytotoxicity activity of human cathelicidin LL-37

Peptide	Cell Line	Assay	Result	Ref
LL-37	A549	MTT	Not cytotoxic up to 50 μg/mL	[130]
Scrambled LL-37	A549	MTT	Not cytotoxic up to 50 μg/mL	[130]
LL-37	A431 squamous cell carcinoma cells	MTT	Cytotoxic at 20 μg/mL. Not toxic at 5 μg/mL.	[131]
LL-37	pMSC	MTT	No toxicity up to 10 μg/mL.	[132]
LL-37	MA-104	MTT, Neutral red	Statistically significant cytotoxicity (>10%) observed 20–50 μg/mL.	[133]
LL-37	Thermally wounded human skin equivalents (HSE)	MTT	No cytotoxicity at up to 200 μg/model	[134]

Table 10:

Hemolytic predictor websites

Name	Link	Reference
HemoPred	http://codes.bio/hemopred/	[142]
HemoPI/Hemolytik	https://webs.iiitd.edu.in/raghava/hemopi/index.php or http://crdd.osdd.net/raghava/hemopi/	[143]
HAPPENN	https://research.timmons.eu/happenn	[144]

Table 11:

Hemolytic prediction of activity for LL-37 human cathelicidin peptide

Test sequence:	**LL-37: LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES**	
Prediction results		
Program used	**Predicted result**	**Notes**
HemoPred	Hemolytic	
HemoPI PROB score	0.34 (SVM (HemoPI-1) based 0.72 (SVM (HemoPI-2) based) (Hemolytic) 0.88 SVM (HemoPI-3) based) (Hemolytic)	Note from website: PROB score is the normalized SVM score and ranges between 0 and 1, i.e. 1 very likely to be hemolytic, 0 very unlikely to be hemolytic.
HAPPENN PROB score	0.089 (Not Hemolytic)	Note from website: PROB score is the normalized sigmoid score and ranges between 0 and 1. 0 is predicted to be most likely non-hemolytic, 1 is predicted to be most likely hemolytic.

Table 12:

Summary of reported percent hemolysis results with different amounts of LL-37 peptide against human red blood cells

Hemolysis of human red blood cells	Reference
8% hemolysis at 20 μM	[147]
~30% hemolysis at 20 μM	[150]
4.47% hemolysis at 38.8 μM	[149]
~10% hemolysis at 60 μM	[146]
9% hemolysis at 100 μM	[151]
~60% hemolysis at 100 μM	[145]
~50% hemolysis at 200 μM	[148]

Table 13:

Peptide parameters for integrated prediction

Parameter of Interaction	**Commonly used parameters**	**Comments**
Antibacterial activity	MIC > 8 μg/mL is often considered "active" performed under CLSI guidelines using CA-MHB and designated concentrations of peptide. The peptide is defined as inactive in the APD with MIC > 100 μg/mL or μM.	Different methods and conditions for antimicrobial activity make it difficult to compare peptide activity. Doesn't account for peptide binding to serum proteins or being cleaved by serum factors *in vivo*. PK/PD data are lacking for AMPs and they are not addressed by this metric.
Host cell cytotoxicity	Cytotoxicity at 100 μg/mL or less; TC_{50} should be < 10–20% at the MIC, depending on the assay used.	The relationship of this value *in vitro* with *in vivo*/whole body toxicity has not been established. Often the level of LL-37 is taken as a benchmark, since it is native to the human body.
Hemolysis	Hemolysis at 100 μg/ml or HC_{50} should be < 10–20% at MIC.	The relationship of this value to *in vivo*/whole body toxicity has not been measured. Often the level of LL-37 is taken as a benchmark, since it is native to the human body.
Host cell permeability	An important parameter if the target microorganism has an intracellular step to its infectious life-cycle.	Assays to measure intracellular replication of bacteria in the presence of extracellular peptide are useful to assess this parameter [116].
Pathogen cell permeability	An important parameter if the target of the peptide at sub-MIC concentrations might be an intracellular component of the bacteria, such as target enzymes or DNA.	Assays to measure intracellular bacterial targets such as enzymes or DNA in the presence of extracellular peptide are useful to assess this parameter [152–154].
Stickiness to other proteins (Boman index)	"This function computes the potential protein interaction index proposed by Boman [3] based in the amino acid sequence of a protein. The index is equal to the sum of the solubility values for all residues in a sequence, it might give an overall estimate of the potential of a peptide to bind to membranes or other proteins as receptors, to normalize it is divided by the number of residues. A protein have high binding potential if the index value is higher than 2.48."	Initially called protein-binding potential [3], Boman index was renamed and programmed in the APD for every peptide [9]. It is also available in the calculation and prediction interface of the APD for any other peptides. This parameter is also programmed in R at https://rdrr.io/cran/Peptides/man/boman.html.
Propensity for host protease cleavage	Protease cleavage will reduce the activity and half-life of the peptide.	Can be predicted using Expasy server PeptideCutter. https://web.expasy.org/peptide_cutter/
Other Negative Effects	**Refs**	**Comments**
Carcinogenic effect	none	No reports were found on the carcinogenic effect of antimicrobial peptides. Work is being done to use AMPs to fight cancer [158–159].
Antigenicity	none	It is very difficult to raise antibodies against antimicrobial peptides. This is accomplished if at all by coupling KLH to the peptide. To our knowledge, there have been no reports of spontaneous antibody production against naturally produced AMP, which is too small.
Cell penetrating properties	[155]	Cell penetrating properties of peptides are probably a negative property on net, especially in seeking a bactericidal mechanism. Website are available to select for CPPs; this could be a counter-selection or down-selection step in an AMP design protocol unless this property is used to target intracellular pathogens.