

# A review on antimicrobial peptides databases and the computational tools

Shahin Ramazi<sup>1</sup>, Neda Mohammadi<sup>2,3</sup>, Abdollah Allahverdi<sup>1</sup>, Elham Khalili<sup>4</sup> and Parviz Abdolmaleki<sup>1,\*</sup>

<sup>1</sup>Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Jalal Ale Ahmad Highway, Tehran 14115-111, Iran

<sup>2</sup>Department of Medical Biotechnology, Faculty of Allied Medical Sciences, Iran University of Medical Sciences, Hemmat Highway, Tehran 1449614535, Iran

<sup>3</sup>Institute of Pharmacology and Toxicology, University of Bonn, Biomedical Center, Venusberg Campus 1, Bonn 53127, Germany

<sup>4</sup>Department of Plant Biology, Faculty of Biological Sciences, Tarbiat Modares University, Jalal Ale Ahmad Highway, Tehran 14115-111, Iran

\*Corresponding author: Tel: +98-21-82883404; Fax: +98-21-82884717; Email: [parviz@modares.ac.ir](mailto:parviz@modares.ac.ir)

Citation details: Ramazi, S., Mohammadi, N., Allahverdi, A. *et al.* A review on antimicrobial peptides databases and the computational tools. *Database* (2022) Vol. 2022: article ID baac011; DOI: <https://doi.org/10.1093/database/baac011>

## Abstract

Antimicrobial Peptides (AMPs) have been considered as potential alternatives for infection therapeutics since antibiotic resistance has been raised as a global problem. The AMPs are a group of natural peptides that play a crucial role in the immune system in various organisms. AMPs have features such as a short length and efficiency against microbes. Importantly, they have represented low toxicity in mammals which makes them potential candidates for peptide-based drugs. Nevertheless, the discovery of AMPs is accompanied by several issues which are associated with labour-intensive and time-consuming wet-lab experiments. During the last decades, numerous studies have been conducted on the investigation of AMPs, either natural or synthetic type, and relevant data are recently available in many databases. Through the advancement of computational methods, a great number of AMP data are obtained from publicly accessible databanks, which are valuable resources for mining patterns to design new models for AMP prediction. However, due to the current flaws in assessing computational methods, more interrogations are warranted for accurate evaluation/analysis. Considering the diversity of AMPs and newly reported ones, an improvement in Machine Learning algorithms are crucial. In this review, we aim to provide valuable information about different types of AMPs, their mechanism of action and a landscape of current databases and computational tools as resources to collect AMPs and beneficial tools for the prediction and design of a computational model for new active AMPs.

## Introduction

### Antimicrobial peptides (AMPs)

In the past few decades, antibiotics have been used to defeat infectious diseases and most of them were discovered during the 1940s to 1960s (1). However, the increased usage of conventional antibiotics has resulted in significant rates of resistance in microorganisms, raising concerns about the spread of infectious diseases. Antimicrobial resistance (AMR) could be detected using a variety of experimental and technological methods, such as phenotypic and molecular-based techniques, as well as the more recently developed sequencing whole-genome sequencing (WGS) and whole-genome metasequencing (WGM), MALDI-TOF MS and Infrared (IR) spectroscopy (2). AMR has become an increasingly urgent challenge in healthcare (3), with antimicrobial-resistant infections estimated to increase to 10 million cases annually by 2050 (4). It has recently been estimated that at least 700 000 people die from antimicrobial-resistant (AMR) infections every year (5, 6). In light of the 2020 COVID-19 pandemic (4, 7), antibiotic-resistant issues are even more exacerbated due to increased different antibiotics prescribed to

COVID-19 patients (8). In addition, the existing repertoire of antibiotics does not offer solutions for multidrug-resistant bacteria, so-called superbugs. Therefore, the global health burden has led to an urgent demand for the expansion of new classes of antibiotics. A highly promising approach to overcome this problem is the development of AMP-based drugs.

AMPs are naturally present in the innate immune system and have broad-spectrum antimicrobial properties aiding in the defense against invading microorganisms (9). They are short cationic peptides of up to 100 amino acids (9), with an alpha-helical secondary structure and amphiphilic surface properties, which are considered essential for establishing antimicrobial activity (10). AMPs' main mechanism of action (MOA) is the disruption of the target microorganism's cell membrane, through hydrophobic or electrostatic interactions, causing lysis of the cell (11). AMPs pose several advantages over conventional antibiotics including the rapid killing of bacteria owing to their broad-spectrum activity, antimicrobial immunomodulatory effects and the less likelihood of AMR (11, 12).

The initial study on AMPs was performed by Dubos *et al.* in 1939 over the soil bacillus strain named gramicidin which is suitable for the typical treatment of wounds and ulcers (13). Recently, AMPs have drawn much attention due to their biological and biomedical applications especially in designing various types of APM-based drugs (14, 15). Over 5000 AMPs have been so far identified or synthesized in a wide variety of organisms ranging from prokaryotes (e.g. archaea and bacteria) to eukaryotes [e.g. yeasts, fungi, viruses, parasites, protozoa, insects, plants and animals (invertebrates and vertebrates)] (13). For example, more than 300 different AMPs exist in the skin of frogs, which is a crucial part of the innate immunity against a wide range of microbes including viruses, bacteria and fungi (13). The gene expression of AMPs is correlated with their activity and maintained at an optimal level. While some AMPs have tissue-specific expression patterns like human  $\beta$ -defensin 1 (hBD-1) or mouse  $\beta$ -defensin 1 (mBD-1), their dysregulations are attributed to the pathological state (16). It has been shown that  $\beta$ -defensins are upregulated in pneumonia (17) and cystic fibrosis (18), while the expression of hBD-2 and hBD-3 is decreased in atopic dermatitis (19).

Besides their antimicrobial activity and immune regulatory roles, AMPs have antiparasitic, antiviral, anti-biofilm, anti-inflammatory, anticancer, insecticidal, wound-healing and/or chemotactic properties which make them interesting candidates for novel therapeutic strategies (20–22). Therefore, AMPs are capable of targeting different types of diseases such as infectious diseases, diabetes, cancer, cardiovascular disease and Alzheimer's disease (23–25). Although the antimicrobial mechanisms of AMPs remain poorly understood, it has been known that AMPs act through the destruction of cell membranes, interference with DNA, RNA, disruption of enzymatic/protein activity, interference with cell division and the inhibition of cell wall synthesis (26). In particular, buforin II is a histone-derived AMP that is mainly found in frogs that destroy *Escherichia coli* (*E. coli*) by binding their DNA and RNA, but without bacteria membrane permeabilization. In addition, it has been shown that  $\beta$ -defensin 4,  $\alpha$ -defensin 1 and PR-39 play key roles by targeting the intracellular bacterial components in humans (27). Nisin and lysozyme as encouraging examples of AMPs which are firstly isolated from *Lactococcus lactis* subsp, human tissues and body fluids, respectively (28). Later, a great number of membrane-lytic peptides were extracted from amphibians, insects and mammals in the 1980s (24). For instance, melittin, mastoparans, cecropins, defensins and magainins are isolated from bee wasp venom, insects, mammalian neutrophils and frog skin, respectively (29). Since then, compelling evidence has demonstrated that there is an interspecies variation in either the sequence or the structure of AMPs isolated from different organisms (30).

Despite the capacity of AMPs as promising alternatives to conventional antibiotics, the number of issues related to the production of AMPs has limited AMPs' applications in clinics (13). These difficulties are assigned to the high toxicity, reduced activity due to the extreme environmental conditions (susceptibility to proteases and extreme pH), lack of specificity, folding problems in large AMPs, bacterial resistance and highly expensive production costs (13). In general, AMPs are short in length, highly selective, efficacious and generally well-tolerated (31–33). Nevertheless, large-scale detection of AMPs is costly and challenging. In current years, computational methods have attracted considerable attention to AMP

prediction (34). To resolve this issue, many computational methods have been recently developed to predict and design putative AMPs *in silico*; databases and computational methods as common tools for the prediction of AMPs contain a great number of AMPs. The Antimicrobial Peptide Database (APD3), is the commonly used AMP database, which covers more than 2600 AMPs (35). It is difficult to classify natural AMPs due to their diversity (36). AMPs are particularly categorized based on their source, activity, structure, sequence, biosynthesis or functions (20, 21, 37). In the following section, we provide a scheme for both the structure and function of AMPs.

### Functions, structure and major activities of AMPs

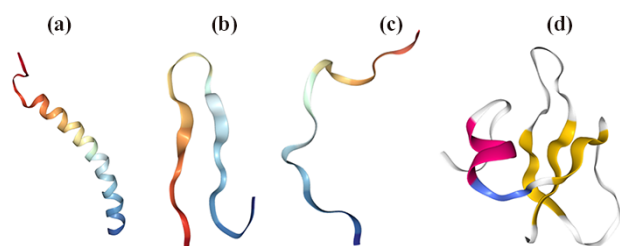
AMPs are diverse and distinct molecules that are distinguished by their chemical structures and amino acid composition. Most AMPs are less than 50 amino acids, with net cationic charge ranging from +2 to +9 and amphiphilic with the molecular weight of <10 kDa containing hydrophobic residues (34, 38). On the other hand, most of these cationic peptides are considered as a heterogeneous group with a length between 12 and 48 residues of amino acid and hydrophobic characteristics to form amphipathic-helix in solvents as fluoro-ethanol that mimics cell membrane (39). Notably, a direct correlation between the charge of AMPs and their antimicrobial activities has been indicated; **an increase in the charge of peptides leads to an improvement in the activity of peptides**. As an illustration, an increase in the charge of magainin 2 from +3 to +5 enhances the antibacterial activity against both Gram-positive and Gram-negative bacteria. In vice versa, these AMPs do not adversely affect eukaryotic membranes. For example, an increase to +6 or +7 leads to an increase in hemolytic activity and the loss of antimicrobial activity (37). 这最后几句话不是很懂

All AMPs were considered cationic in the late 90s, but later with the discovery of negatively charged AMPs in 1997, this view was changed (13). Some natural peptides are negatively charged, such as maximin H5, dermcidin and enkephalin, which are enhanced by their activities by combination with zinc or highly cationic peptides (13, 40). AMPs are phospholipid-rich and could act through a lipid bilayer in a detergent-like manner, solubilizing it into micelles and/or allowing it to penetrate by forming pores. Both interactions generate transient membrane permeation and cytoplasmic leakage depending on the AMP concentration which might cause cell death (41). In addition, AMPs interfere with the intracellular activities of the bacteria by the inhibition of intracellular activities, such as cell division and biosynthesis of proteins, nucleic acids and components of the cell wall (41).

AMPs with the structural and functional variety are obtained from three sources: natural sources like microbes, plants, animals and insects, which are synthesized by ribosomal or nonribosomal approach, recombinantly expressed in microorganisms and chemically synthesized sources (42). While ribosomal AMPs are produced by all the species of life such as mammals, birds, amphibians, insects, plants or particular microorganisms, nonribosomal AMPs are mainly synthesized by bacteria and fungi (43, 44). In terms of chemically synthesized peptides, firstly Bruce Merrifield introduced solid-phase peptide synthesis (SPPS) in 1963 (45). Currently, the chemical synthesis of peptides has been significantly

developed owing to the reasonable production price compared to recombinant production (46). Chemically synthesized AMPs are advantageous in comparison to other methods since there is a possibility to produce unnatural amino acids, D-amino acids and other building blocks with a noticeable quantity and quality (47).

A broad-spectrum function and the structure of APMs are excessively attributed to the post-translational modifications (PTMs) mediated by proteolytic cleavage, phosphorylation, glycosylation, amidation, halogenation, D-amino acids, disulfide bridge and cyclization (13, 48). The 3D structures of AMPs have been determined by circular dichroism spectroscopy, X-ray crystallography and nuclear magnetic resonance (NMR) (28, 49). The first structure human  $\alpha$ -defensin and neutrophil peptide 3 was characterized by X-ray crystallography in 1991, and then the structure of the human neutrophil peptide 1 was determined by NMR (50, 51). AMPs are classified based on their structure into four broad families:  $\alpha$ -helix,  $\beta$ -sheet, loop and extended. AMPs with  $\alpha$ -helix and  $\beta$ -sheet structures are the most prevalent structures in nature (Figure 1) (13). Cathelicidin LL-37, human lactoferricin, magainin and cecropin are the most studied  $\alpha$ -helical peptides (52–55). Studies showed that helical peptides are destabilized in an aqueous solution and undergo an amphipathic structure upon interaction with the biological membrane (56). Cathelicidin contains 12–80 amino acids and adopts a diversity of structures and exists in a large group of mammals such as mice, goats, sheep, horses and bovines. Lactoferrin is found in neutrophils and the secretions of the exocrine glands of mammals. Magainins are a class of helical peptides that are mainly effective against Gram-positive and Gram-negative bacteria, fungi, yeast and viruses, and isolated from the African clawed frog *Xenopus laevis* (57). Cecropins are the first discovered AMPs in eukaryotes in the silk moth which have cationic, amphipathic activities against Gram-positive and Gram-negative bacteria and fungi. In recent years, these peptides have been identified in fruit flies (*Drosophila*) and marine invertebrates such as shrimp, oysters and horseshoe crabs (37). Aurein peptides are another example of  $\alpha$ -helical AMPs and consist of more than 30 aurein peptides and five different families primarily secreted from the granular dorsal glands of the Australian Green and Golden Bell Frog *Litoria aurea* and the Southern Bell Frog *Litoria raniformis* (58).



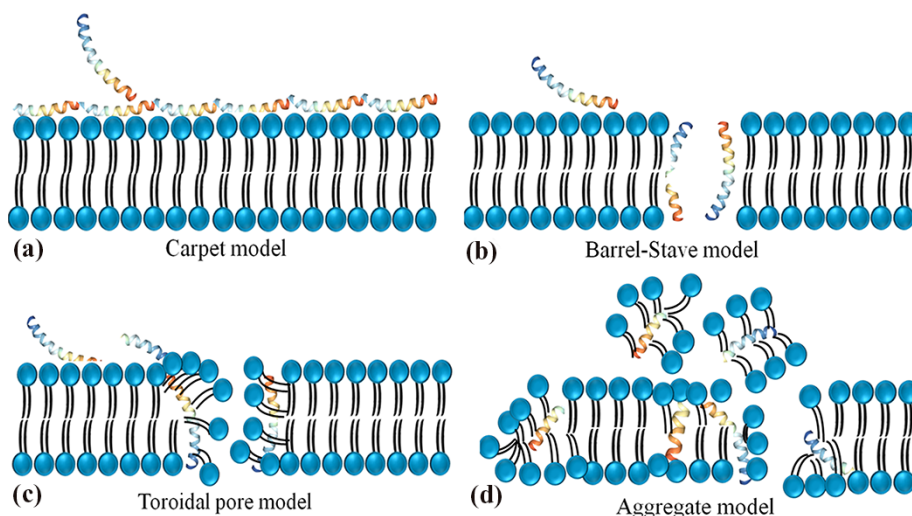
**Figure 1.** Structural diversity of AMPs based on their secondary conformations using Protein Data Bank (PDB). [(a) **2K60**: Antimicrobial Peptide, the  $\alpha$ -helical structure of cathelicidin LL37 in *Homo sapiens*. (b) **1RKK**: Antimicrobial Peptide,  $\beta$ -sheeted polyphemusin in *Limulus polyphemus*. (c) **1G89**: Antimicrobial Peptide, extended indolicidin in *Bos taurus*. (d) **1FQQ**: Antimicrobial Peptide, Antibiotic and mixed structures like human  $\beta$ -defensin-2 in *Homo sapiens*].

Most Aurein peptides are active against Gram-positive bacteria, such as *Staphylococcus aureus* and *Staphylococcus epidermidis*. Furthermore, aurein peptides 1.2, 3.2 and 3.3 show the greatest activity against more than 30 various types of cancer (37). Interestingly, Aurein peptides are rich in specific amino acids. For instance, histatin as an antimicrobial peptide isolated from human saliva is histidine-rich and defeats *Candida albicans* mushrooms (59, 60) while bactenecin Bac-5 and Bac-7 peptides are prolin-rich and possess an irregular structure (61).  $\beta$ -sheet peptides contain cysteine residues with the rigid structure stabilized with disulfide bonds and an unaltered conformation in contact with the cell membrane. Protegrins (a member of the cathelicidin family), defensins and tachyplesins have the  $\beta$ -sheet structure (37). AMPs can be found in leaves, flowers, seeds and tubers of plants. Some are cysteine-rich with multiple disulfide bonds playing key roles in high chemical, thermal and proteolytic stability. Defensins, thionins, hevein-like peptides, knottin-type peptides (linear and cyclic),  $\alpha$ -hairpinins, lipid transfer proteins and the snakins family are the examples of plant-derived AMPs (62, 63).

### Mechanism of AMPs action

Antimicrobial peptides are unique molecules and their MOA has been studied extensively since they were discovered. It is important to understand the MOA of these AMPs to facilitate further development as therapeutic agents. It was originally thought that membrane targeting was the only MOA, but there is increasing evidence now that AMPs have other modes of action. The MOA can be divided into two major classes: direct killing and immune modulation (64). As described above, AMPs have immune-modulatory and antimicrobial roles through directly targeting membrane and non-membrane regions of microbes (27). In most cases, the positive net charge of AMPs displays a significant role in antimicrobial activity via the strong interaction with negatively charged bacteria surfaces and thus disruption of the physical integrity (26). Four different models for the antimicrobial activity of AMPs have been suggested which lead to membrane disruption through permeabilization, including a barrel-stave, aggregate channel or toroidal pore and carpet models (Figure 2) (65).

In the barrel-stave model, AMPs are inserted perpendicularly in the membrane bilayer and form a pore. In this pore, the hydrophobic sides interact with the lipids of the membrane and form the interior side of the channel. In this way, they act as pore formers or act as metabolic inhibitors in bacteria (66). Compelling evidence showed that alamethicin (67), pardaxin (68) and protegrins (66) form barrel-stave channels. In the toroidal pore model, the peptides are inserted vertically into the membrane and form a curve structure and a pore through the peptides and the head phosphates of phospholipids. Some peptides are permitted to enter the cytoplasm using this model and target intracellular components (69) including magainin 2 (70), lactacin Q (70), aurein 2.2 (71) and melittin (70). In the carpet model, the AMPs cover the surface of the membrane and cause tension which leads to membrane disintegration and micelle formation. Some AMPs such as cecropin (72), indolicidin (73), aurein 1.2 (73) and LL-37 (74) form carpet models. These three models are suggested that lead to the breakdown of membrane integrity resulting in membrane dysfunction, and leakage of metabolites and ions (75).



**Figure 2.** Schematic representation of the potential mechanism of membrane disruption and/or translocation by antimicrobial peptides. **(a) Carpet model:** Another face of the membrane is covered by AMPs to form a ‘carpet’ and the membrane undergoes some perturbation and deformation. **(b) Barrel-Stave model:** AMPs interact laterally and form transmembrane pores. **(c) Toroidal pore model:** AMPs penetrate the bilayer membrane and form a toroid of high curvature. **(d) Aggregate model.**

This membrane permeabilization is also contributed to the subsequent translocation of AMPs into the intracellular region and blocks critical cellular processes such as protein/ nucleic acid synthesis, enzymatic/protein activity, protein folding, intracellular pathways and/or cell wall synthesis (75).

Furthermore, AMPs are mainly produced by some immune cells such as neutrophils and macrophages and exert immunomodulatory activities such as the recruitment and activation of immune cells, initiation of adaptive immunity, reduction of inflammation (27), chemo attraction of immune cells, induction of chemokine, cytokine, and histamine production and secretion, wound healing stimulation, angiogenesis and adjuvant city (76).

The rest of the paper is structured as follows. The five most studied AMPs are described in the section ‘Classification of AMPs based on biological functions’. In the next section, major AMP databases will be reviewed. Afterward, computational methods for the prediction of AMPs will be described in detail. Finally, available tools for AMP prediction will be reviewed in the section ‘Tools for AMP prediction’.

## Classification of AMPs based on biological functions

3000 synthetic and natural AMPs have been identified and 7 have received approval from the U.S. Food and Drug Administration (FDA) (77). In humans, AMPs are mostly present in lymphocytes and epithelial surfaces of different organs including the eye, skin, lung, intestines, etc (22). A great number of AMPs (e.g. alpha-defensins, lysozyme, etc) are produced in paneth cells, primary secretory epithelial cells in the small intestine, thereby, controlling the number of bacteria in the small intestine. Defensins, lysozyme and cathelicidins in the tear fluid protect the eyes from infections (22).

According to AMPs’ biological functions, AMPs could be divided into various groups such as antibacterial peptides (ABPs), antiviral peptides (AVPs), antifungal peptides (AFPs), anticancer peptides (ACPs) and antiparasitic peptides (APPs). Categories of AMPs discussed based on the

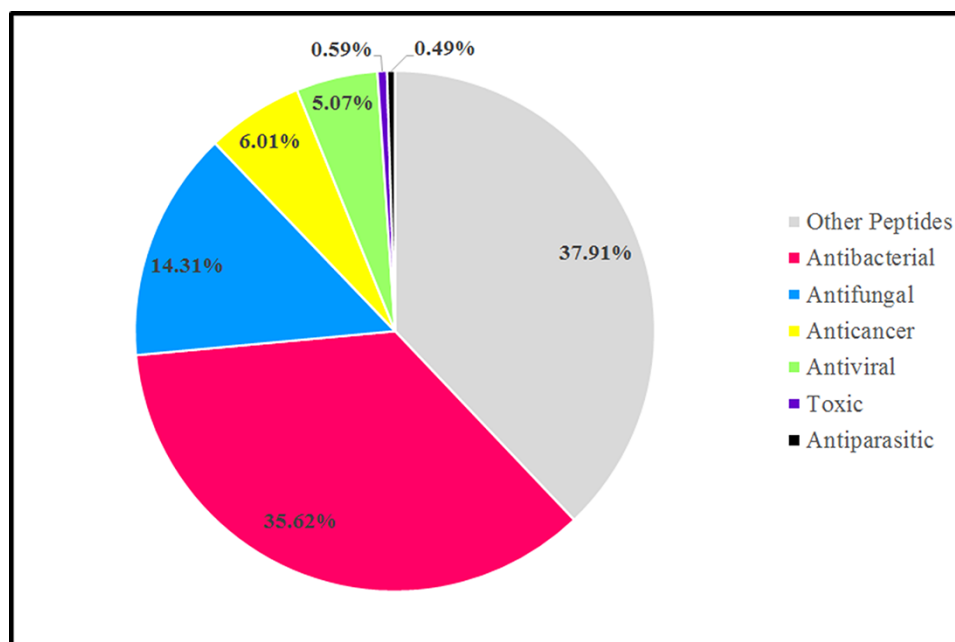
Databank antimicrobial peptides (dbAMP). In the dbAMP database, ABPs comprised the largest proportion, approximately 35.62%, followed by AFPs, which account for 14.31%, and, ACPs, AVPs, toxic peptides and APPs account for about 6.01%, 5.07%, 0.59% and 0.49%, respectively. In addition, the dbAMP database contains 37.91% of various other peptides, which are known as disease-associated peptides and new mechanism-associated peptides (Figure 3) (78).

## Antibacterial peptides (ABPs)

ABPs are cationic AMPs, which have been well documented for their role in the development of antibacterial drugs. Despite eukaryotic membranes with zwitterionic lipids, bacterial membrane lipids such as phosphatidylglycerol (PG), cardiolipin (CL) or phosphatidylserine (PS) have negatively charged residues with lipids bearing phospholipid head groups which predispose their membrane to the disruption (34). These AMPs frequently fold into amphiphilic  $\alpha$  helices exposing both hydrophobic and hydrophilic surfaces (79). Nisin (as an ABP) and vancomycin (as an antibiotic) function through the blockage of cell wall synthesis (13). They are accumulated on the negatively charged outer membrane of the Gram-negative bacteria or the cell wall of the Gram-positive bacteria leading to the formation of membrane-spanning pores, inhibition of cell wall biosynthesis, and thereby, disruption of membrane integrity (80). Once the bacteria are penetrated, they interact with intracellular components to destroy them. Some of ABPs with low concentrations and without interacting with the membrane, lead to the death of bacteria by inhibiting many significant pathways inside the cell such as DNA replication and protein syntheses such as buforin II, drosocin, pyrrocoricin and apidaecin (13).

The primary link of peptides with the bacterial membrane happens via electrostatic interactions between the cationic peptide and anionic lipopolysaccharides (LPS) in the outer membrane leading to membrane disorder (49). Bacteriocins are the subset of ABPs which are mainly classified into two categories: lantibiotics and non-lantibiotics. Lantibiotics





**Figure 3.** Graphic representation sources of AMPs in the antimicrobial peptide database (dbAMP 2.0) in 2022 (78).

comprise the nonnatural amino acid lanthionine (37). Nisin and mersacidin belong to the family of antibiotics which are produced by *Lactococcus lactis* and *Bacillus* sp, respectively. It has been shown that nisin and mersacidin are active against antibiotic-resistant Gram-positive bacteria (37).

### Antiviral peptides (AVPs)

Viral diseases are the foremost cause of illness and mortality worldwide and more than 200 viruses are accounted for a variety of human diseases (81) like influenza (IAV), West Nile Virus (WNV), Epstein-Barr virus (EBV), cytomegalovirus (CMV), respiratory syncytial virus (RSV), hepatitis B and C viruses (HBV and HCV, respectively), herpes simplex virus (HSV), human immunodeficiency virus (HIV), rabies virus and Ebola virus (81–83). So far, 1.5 million deaths due to HIV, 400 million cases due to HBV or HCV, 80% of liver cancer deaths related to hepatitis viruses, 500 000 cervical cancer cases, and 250 000 deaths associated with HPV have been reported by the World Health Organization (WHO) (82). Among over 60 antiviral drugs approved by the FDA, around half of them have been used to treat HIV-1 and the rest are used to treat HBV, HSV, varicella-zoster virus (VZV), CMV, IAV and HCV (83). Therefore, the low access to therapeutic possibilities for many viral infections emphasizes the efforts to develop new and more effective antiviral drugs. In recent years, 15 peptide-based drugs which are in different stages of clinical trials have provided new opportunities to combat wide-spectrum viruses. However, none of the ACPs has cationic properties and are widely used (13, 81). AVPs have cationic features with virucidal activity: They mainly interfere with the virus replication by targeting DNA or RNA after binding to their envelopes and causing membrane instability (13, 81). As a result, antiviral drugs have mostly two types of mechanisms of action, i.e. virus targeting and host targeting drugs that can inhibit various transcriptional

and replication-related enzymes and lead to destroy a viral pathogen or inactivate their infectiveness (13).

### Antifungal peptides (AFPs)

Fungi are considered as a unique kingdom with different characteristics than other eukaryotic kingdoms such as plants, animals and are more complex than viruses or bacteria. They are used in the industry to produce peptides, vitamins, antibiotics, organic acids, enzymes, etc. (84). The cell wall of fungi is composed of chitin, 1,3- $\beta$ - and 1,6- $\beta$ - glucans, proteins and other polymers of the complex cellular organizations (85). Over 400 species are accountable for various infectious diseases in humans (86). In most cases, fungal infections may lead to serious problems in people who are very sensitive, such as mainly immunocompromised, elderly, and transplanted subjects, cancer patients and premature infants, and the elderly with significant associated morbidity/mortality (85, 87). Recent reports suggest that current antifungal drugs have caused a significant rise in drug-resistant strains and their incidence is on the increase. Therefore, is need alternative antifungal drugs that are capable of overcoming resistance mechanisms (88, 89).

AFPs have provided a great extent of advantages by being effective against multiple targets and developing less resistance (88, 89). AFPs have been extracted from many natural resources such as plants (85), amphibians (90), bacteria (91), fungi (92), marines (93) and insects (94). Most AFPs have a length of ~50 amino acid residues in linear or cyclic structures with hydrophobic or amphipathic properties, cationic and cysteine-rich proteins (CRPs) (95, 96). AFPs play key roles in many action mechanisms of cells such as inhibition of DNA, RNA and protein synthesis, binding to DNA or RNA, membrane permeabilization, inhibition of cell wall synthesis and enzyme activity, induction of apoptosis and repression of protein folding (95). Lytic peptides of fungi bind to the

membrane surface and can destroy the cell membrane with or without crossing the membrane (13). There is no direct correlation between the structure of AFPs and the type of target cell. For example, AFPs have members from various structure classes like  $\alpha$ -helical (D-V13K and P18), extended (indolicidin) and  $\beta$ -sheet (defensins) (13).

### Anticancer peptides (ACPs)

Despite a wide variety of cancer treatment methods, this disease is one of the most common causes of death worldwide (97). A common method for treating cancer is chemotherapy, which damages both cancerous and normal cells by inhibiting DNA replication. On the other hand, chemotherapy drugs cause chemical resistance, which results in a low success rate and an increased risk of recurrence (98, 99). Besides, there is a well-known reciprocal relationship between infection and cancer which is associated with the weak immune system to provide a proper situation for cancer and infection (100–102). In recent years, some AMPs have exhibited antitumor activity called anticancer peptides (ACPs), acting as mitogens and signaling molecules. ACPs are described as promising chemotherapeutic drugs in the future, particularly owing to the low resistance, minimal side effects, high specificity and proper solubility (103). ACPs are functionally categorized into two classes: ACPs with dual activity against cancer cells and bacteria, but not normal cells, and ACPs with cytotoxic function against microbial infections, cancer cells and also normal cells (104, 105). These peptides are typically less than 50 amino acids and possess high hydrophobicity and positive net charge (106).

The physicochemical properties of cancer cells provide the basis for the function of ACPs. Generally, eukaryotic cells are bilayered membranes containing asymmetric zwitterionic phospholipid composition (107). The double-layer membrane has phosphatidylcholine (PC) and sphingomyelin (SM), phosphatidylethanolamine (PE) and phosphatidylserine (PS) (107). Despite the healthy cells, PS, a phospholipid with a negative net charge, is translocated from the inner to the outer membrane in cancer cells (108). Owing to the highly cationic and amphipathic features, AMPs target cancer cells exist. Thus, ACPs are attached according to their cationic and amphipathic characteristics by electrostatic interactions with a negative net charge in the outer membrane of cancer cells. These anionic molecules can affect with the utmost selectivity and toxicity through the destabilization of the membrane integrity (108). In addition to the membranolytic mechanism, ACPs promote necrosis or apoptosis in cancer cells by inducing mitochondria-derived pathways (105, 108–110). Hence, the negative charge of the cancer cell membrane is an important factor to promote the ACPs' electrostatic interaction (111).

### Antiparasitic peptides (APPs)

Parasitic diseases like malaria, leishmaniasis, trypanosomiasis, schistosomiasis and chagas have imposed a great burden on humans, by mostly affecting the poor population. The lack of suitable vaccines and drugs without causing resistance necessitates the development of new drugs (13). APPs are short in length ( $\sim 5$ –30 amino acids) which target Protozoa through plasma membrane disruption and consist of a smaller group of AMPs compared to the other four AMP classes (13).

Magainins and cecropins are the first APPs, reported 20 years ago, which are active against *Paramecium caudatum* (13). APPs can be isolated from the host including mosquitoes and other invertebrates (112–115). It has been reported that APPs have great potential for treating diseases including protozoan parasites (114). Despite the multicellularity of some parasitic microorganisms, antiparasitic peptides act in the same way as other AMPs, directly targeting and killing cells by destabilizing cell membranes (13).

## Major AMPs databases

A multitude of evidence has shown that AMPs have remarkable antimicrobial effects, particularly against the increasing number of resistant microbes. However, many of them are not approved by FDA and fail before or during clinical trials (37). To meet this need, several databases have provided more classified information for the effective design and construction of AMPs. Databases enable users to search and mine extensive information on the peptide structure, chemical modifications, bioactivities and classification. Tables 1 and 2 present a list of databases. The AMPs databases are classified into two main groups: general databases and specific databases. The general databases contain the whole types of AMPs irrespective of a given peptide family while specific databases cover information related to a certain class of AMPs (e.g. only defensins or cyclotides) or hold a supergroup of AMPs (e.g. only plant peptides or only cyclic peptides). Currently, there is not a universal database with all AMP data, the information is divided into several databases (Tables 1 and 2), and there exists an overlap as well between AMP databases; nonetheless, each database contains some exclusive sequences (63).

### General AMPs databases

In the subsequent section, the six comprehensive general databases are defined briefly. In addition, Table 1 reviews the current main public AMPs general databases.

Data bank antimicrobial peptides (dbAMP) database contains various information about different types of AMPs in 3044 organisms. Newly, the dbAMP database contains 2 262 antimicrobial proteins and more than 26 440 unique entries, including experimentally verified AMPs and putative AMPs along with their functional activities, which expanded using protein databases of UniProt, NCBI, Protein Data Bank, and eight public AMP databases. In this study, for large-scale detection of AMPs using transcriptome data, all amino acid sequences of AMPs were converted into DNA sequences to create an efficient pipeline using the Docker container for discovering AMPs from Next-Generation Sequencing (NGS) data using the Bowtie2 program. Users can submit large-scale data from NGS reads or peptides identified via MS/MS to the dbAMP. In addition, the system could identify known AMPs with their functional types and predict new AMPs by the constructed model (78).

Database of antimicrobial activity and structure of peptides (DBAASP) contains over 15 700 entries (8000 more than the previous version), including  $\sim 14$  500 monomers and nearly 400 homo- and hetero-multimers. Of the monomeric AMPs,  $\sim 12$  000 are synthetic, about 2700 are ribosomally synthesized, and about 170 are non-ribosomally synthesized. DBAASP is freely accessible and contains

**Table 1.** A description of existing antimicrobial general databases

Database Name	Number of covered classes and AMPs	General Statistics			
		Size	Type of Database <sup>a</sup>	Type of Data	Years
dbAMP	26 major Functional activity classes in 3044 organism	~26 440	Exp. and Pred. Secondary	Natural and Synthetic	2022
DBAASP	Antibacterial, antifungal, antiviral, anticancer, and antitumor in seven organisms and cancer cells and mammalian cells	~15 700	Exp. and Pred. Secondary	Natural, synthetic, and patent	2021
LAMP	8 major functional classes and 38 functional activities	~ 23 250	Exp. and Pred. Secondary	Natural, synthetic, and patent	2020
DRAMP	Antimicrobial, antifungal, antiviral, anticancer, antitumor, antiprotozoal, and insecticidal	~ 22 250	Exp. and Pred. Secondary	Natural, synthetic, patent, and AMPs in drug development	2019
dbAMP	26 major Functional activity classes in 3044 organism	~26 440	Exp. and Pred. Secondary	Natural and Synthetic	2018
InverPep	Invertebrates <i>phyla Arthropoda, Mollusca, Nematoda, Annelida, Echinodermata, Platyhelminthes, Placozoa</i> , the Hydridae family ( <i>Cnidaria</i> ) and the subphylum Tunicata ( <i>Chordata</i> )	~770	Exp. Primary	Natural	2017
CAMP	Antibacterial, antifungal and/or antiviral	~8160 sequences, and 757 structures	Exp. and Pred. Secondary	Natural	2016
MEGARes	Antimicrobial compounds, e.g. drugs, biocides, multi-compound and metals	~8000	Exp. Primary	Predicted and patented Natural	2016
ADAM	archaea, bacteria, plants and animals	~7000	Exp. Primary	Natural	2015
APD	Antibacterial	~1230	Exp. and Pred. Primary	Natural and patent	2008
Defensins Knowledge-base	Defensin, antimicrobial	~360	Exp. Primary	Natural	2007

<sup>a</sup>Data in the database is collected as experimental and/or predicted, which was respectively displayed as the abbreviation 'Exp. and Pred'. Nonetheless, the primary database is created by manually experimental data and the secondary database is created by using an integration of some other databases. **DBAASP**: Database of Antimicrobial Activity and Structure of Peptides; **LAMP**: Linking antimicrobial peptide; **DRAMP**: Data Repository of Antimicrobial Peptides; **dbAMP**: Database antimicrobial peptides; **CAMP**: Collection of antimicrobial peptides; **ADAM**: A Database of Antimicrobial peptides, **APD**: Antimicrobial Peptide Database.

**Table 2.** A description of existing antimicrobial-specific databases

Database	Specific Statistics			Type of Data	Years	Web site
	Type of AMPs	Size	Type of Database			
BaAMPs	Anti-biofilm peptides	~237	Exp. Primary	Natural BaAMPs	2015	<a href="http://baamps.it/">http://baamps.it/</a>
CancerPPD	Anticancer peptides	~3490	Exp. and Pred. Secondary	Natural, and Predicted ACPs	2015	<a href="http://crdd.osdd.net/raghava/cancerppd/">http://crdd.osdd.net/raghava/cancerppd/</a>
ParaPep	Antiparasitic peptides	~860	Exp. and Pred. Secondary	Natural, and Predicted APPs	2014	<a href="http://webs.iitd.edu.in/raghava/parapep/peptide.php">http://webs.iitd.edu.in/raghava/parapep/peptide.php</a>
YADAMP	Antibacterial peptides	~2525	Exp. and Pred. Secondary	Natural, and Predicted ABPs	2012	<a href="http://yadamp.unisa.it/">http://yadamp.unisa.it/</a>
DADP	Amphibian peptides	~2570	Exp. Primary	Natural	2012	<a href="http://split4.pmfst.hr/dadp/">http://split4.pmfst.hr/dadp/</a>
THIOBASE	Bacterial thiopeptides	~100	Exp. and Pred. Secondary	Natural, and Predicted thiopeptides	2012	<a href="http://db-mml.sjtu.edu.cn/THIOBASE/">http://db-mml.sjtu.edu.cn/THIOBASE/</a>
BACTIBASE	Bacteriocins	~177	Exp. and Pred. Secondary	Natural, and Predicted ABPs	2010	<a href="http://bactibase.ammamillab.org/main.php">http://bactibase.ammamillab.org/main.php</a>
Cybase	Cyclotides	~1270	Exp. and Pred. Secondary	Natural, and Predicted AMPs	2008	<a href="http://www.cybase.org.au/">http://www.cybase.org.au/</a>
Defensins Knowledgebase	Defensins	~300	Exp. and Pred. Secondary	Natural AMPs	2007	<a href="http://defensins.bii.a-star.edu.sg/">http://defensins.bii.a-star.edu.sg/</a>
peptaibol	Peptaibols	~317	Exp. Primary	Natural AMPs	2004	<a href="http://peptaibol.cryst.bbk.ac.uk/home.shtml">http://peptaibol.cryst.bbk.ac.uk/home.shtml</a>

**BaAMPs:** Biofilm-active AMPs database; **ParaPep:** Parasites Peptides; **YADAMP:** Yet another database of antimicrobial peptides; **DADP:** Database of anuran defense peptides; **THIOBASE:** A Database of Thiopeptides Featured in Genetics and Chemistry; **BACTIBASE:** A database dedicated to bacteriocins; **Cybase:** Cyclic protein database; **Defensins:** A manual database on the defensins family of antimicrobial peptides; **Peptaibol:** A database for sequences and structures of naturally occurring peptaibols.

information about amino acid sequences, chemical structure, target species, the target object of the cell and peptide antimicrobial/hemolytic/cytotoxic activities of peptides. The user can search for peptides based on structural characteristics, complexity type, source, synthesis type (ribosomal, nonribosomal and synthetic) and target species. Importantly, DBAASP provides a prediction tool for the *in silico* design of new AMPs (116).

Linking antimicrobial peptide database (LAMP) is an online resource for studying experimentally observed AMPs. LAMP contains natural, synthetic and predicted AMPs and is a useful resource for the discovery and design of AMPs as new antimicrobial agents. LAMP comprises three catalogs of AMPs by data sources: experimental, predicted and patent. AMPs in LAMP are short in length, less than 100 amino acids. Currently, LAMP2, an updated version of LAMP has been created which contains more than 23 250 unique AMP sequences and expands to link 16 public AMP databases. LAMP2 covers ~7, 800 natural AMPs and ~15, 400 synthetic peptides (117).

Data repository of antimicrobial peptides (DRAMP) is another AMPs database and contains useful data about the sequence, structure, antimicrobial activity, physicochemical, patent, clinical and reference information of AMPs. Now, the DRAMP comprises ~22 250 entries, more than 5890 general AMPs (containing natural and synthetic AMPs), ~16 110 patent AMPs and 77 peptides in drug development. DRAMP database contains various information about computational methods obtained from data mining tools and introduces the new design for the development and optimization of AMP-based drugs (5).

A database of invertebrate antimicrobial peptides (InverPep) is a database of AMPs belonging to invertebrates.

InverPep contains more than 770 experimentally validated AMPs which were manually collected from other databases and scientific literature. Notably, this database contains 33 AMPs that are not reported in other databases. Most AMPs in InverPep are 10 and 50 amino acids in size and positively charged that have 30–50% hydrophobic amino acids. AMP peptides in InverPep have information about their source, physicochemical properties, secondary structure, biological activity and also links to the external literature (118).

Collection of antimicrobial peptides (CAMP) is a comprehensive database of sequences, structures and family-specific signatures of prokaryotic and eukaryotic AMPs. Currently, CAMP encompasses more than 8160 sequences, 757 structures, ~2080 patent AMPs and 114 family-specific signatures of prokaryotic and eukaryotic AMPs. Also, it has provided the tools for sequence alignment, pattern creation and AMP identification (119).

### Specific AMPs databases

Many databases have been created based on certain types, specific sources or certain characteristics of AMPs to search AMPs based on specific classes. Table 2 reviews the current main public AMPs specific databases.

### A brief history of machine learning techniques on AMPs

For the identification of AMPs, high-throughput experimental methods are labor-intensive and time-consuming. Therefore, machine learning (ML) methods and powerful tools to predict AMPs are urgently needed. The advent of high-throughput screening coupled with decades of experimental data allowed for the duration of large annotated datasets (120). In the



last 10–15 years, the focus of ML has shifted to an intensely data-driven approach. Significant advancements in computational power and easy-to-use statistical learning tools have made supervised ML a viable strategy for leveraging large datasets for the high-throughput and high-accuracy classification of AMPs. Typical readouts from biophysical assays on AMPs include calculations of minimum inhibitory concentrations, minimum bactericidal concentrations and binding affinities. These quantities, coupled with sequence information about AMPs, allow for the training of various supervised learning models using peptide sequence information as an input. Before this era, methods for *de novo* AMP discovery relied on long-standing bioinformatics methods, including sequence alignment and homology modeling for the prediction of biological activity. Now, the convergence of innovations in ML models, the presence of modern computational tools and the availability of high-quality datasets have enabled the ML-aided design of AMP.

In one of the first applications of ML to AMPs, Lata *et al.* (121) developed a Quantitative Structure–Active Relationship (QSAR) AMP classification tool based on artificial neural network (ANN), support vector machine (SVM) and quantitative matrix models based on unique motifs found in the C- and N-terminal residues of known AMPs. In 2009, Chersakov *et al.* used high-throughput screening methods to train an ANN model on the measured antimicrobial efficacies of thousands of nine-residue peptides to discover potent antimicrobials that were potent against multi-drug-resistant bacteria (122). Fjell *et al.* (2008) published a study using hidden Markov models (HMMs) to screen for AMPs in the bovine genome, which led to the discovery of a previously unknown AMP and confirmed the absence of  $\alpha$ -defensins (123). In a similar vein, this group later developed an ANN model in 2009 to screen a larger number of synthetic AMP candidates, characterizing 18 sequences with high antimicrobial efficacy against multi-drug-resistant bacteria (123). Wang *et al.* used a combination of sequence alignment and feature selection methods to design a computational model to more accurately classify AMPs (124). Similarly, Torrent *et al.* (125) trained an eight-descriptor SVM to classify AMPs with 75–90% accuracy while taking into account new factors like peptide aggregation. Maccari *et al.* used random forest (RF) models to design and validate the antimicrobial activity of two natural peptides and one peptide with nonnatural amino acids (126). Giguere *et al.* used a kernel method based on graph theory to train a 100 peptide dataset based on multiple measures of bioactivity to predict novel candidates (127). Most recently, Schneider *et al.* reported the first application of unsupervised–supervised two-step models to classify AMPs. They used self-organizing maps to apply nonlinear dimensionality reduction to the training data, which were then used as an input for a supervised neural network model. Together, these studies highlight a diversity of methods and approaches that have been used to classify and design AMPs with great success (128). Xiao *et al.* designed a two-level classifier to first classify peptide sequences as an AMP, and then sub-classify them into 10 functional AMP categories (129).

## Recent application of machine learning methods for predicting AMPs

Researchers have recently coupled the use of AMP databases as a unique method for AMP prediction with experimental

validations to obtain more efficient AMP-based drugs (130). Most of the learning methods discussed in these databases were based on supervised learning and validated datasets of AMPs (34). To this end, ML algorithms including SVM, NN, RF, fuzzy k-nearest neighbor (fuzzy k-NN), HMM, discriminant analysis (DA) and logistic regression (LR) are proposed to identify AMPs (131, 132). These methods had played a key role in the AMP research, nevertheless, they have problems. Models can only tell whether a new sequence is AMP and also, short peptides tend to be harder to find in the database because AMPs are usually only 10–50 amino acids long. Blast search and gene ontology methods are often ineffective in predicting AMPs. Several general databases contain imbalanced datasets of AMPs activities that are not distributed equally (133). The standard ML algorithms often cannot achieve ideal performance when trained on unbalanced AMPs data sets. To solve this problem, existing classifier learning algorithms can be adapted to strengthen learning in the minority class or used to artificially sample the class distribution. Therefore, can achieve strong classifiers via a combination of both approaches. Thus, for classifying AMPs' functions used a multilabel classification (MLC). During the past 2 decades, the topic of learning from multilabel datasets (MLDs) has been intensively discussed. For example, ML-SMOTE is a new synthetic minority over-sampling technique, designed for processing and identifying AMPs' functional families based on imbalanced and multilabel datasets (133). A good prediction method combines good unbiased training data, a discriminative feature subset and a suitable learning algorithm. Every computational method for predicting a specific type of AMP based on the sequence information requires the following steps.

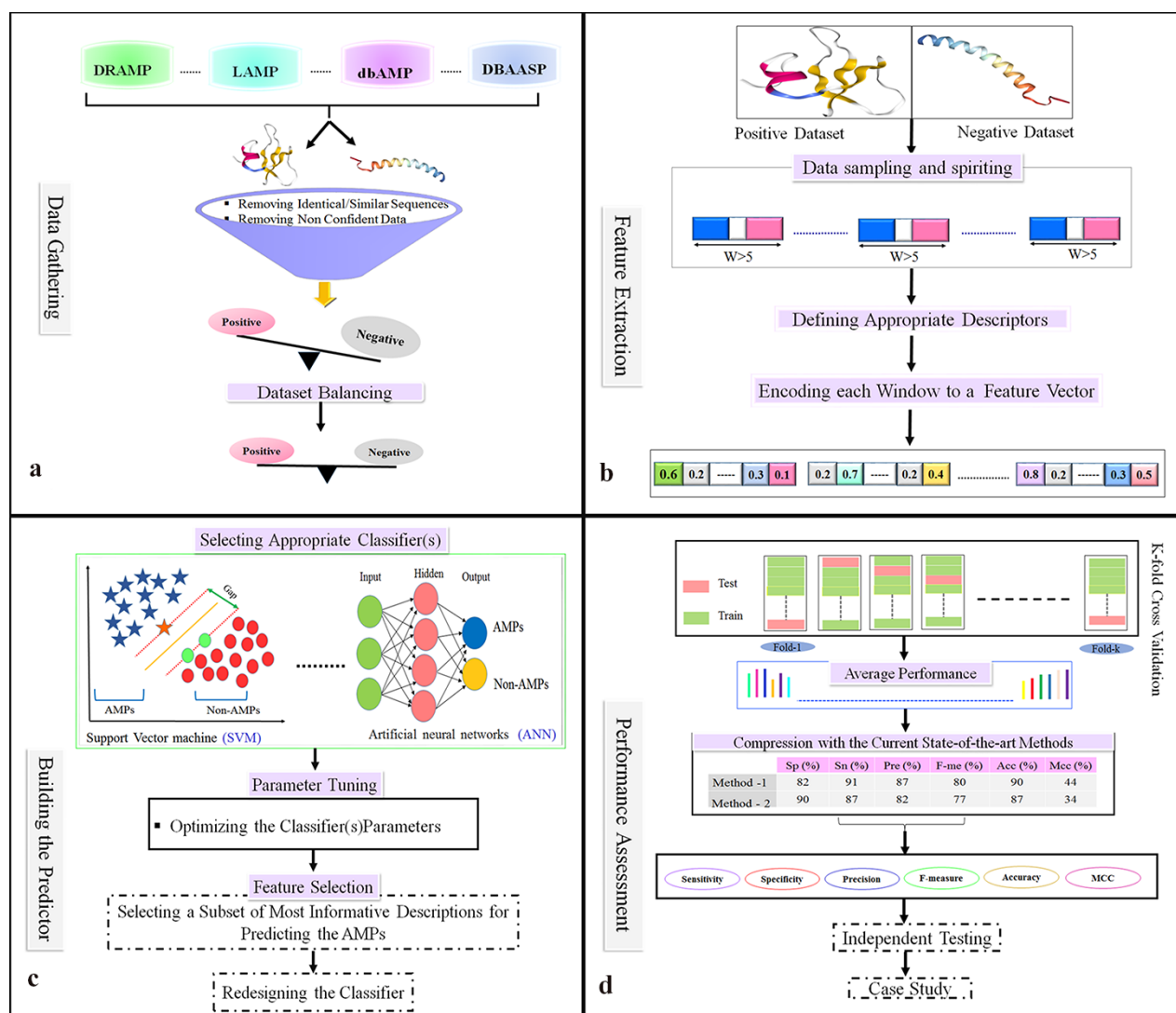
## Data gathering

Select or create a validation benchmark dataset for training and testing the predictor. To assemble a validation dataset, the first step of the AMP prediction method is to collect the data from AMP databases (Figure 4a). It is necessary to include both sequences of peptides with validated AMP activity as positive samples and sequences of peptides with no validated AMP activity as negative samples capable of training a ML algorithm for predicting AMPs. Positive samples are usually collected from the aforementioned databases (such as LAMP or DBAASP). However, the selection of negative samples is the most challenging part of the data collection process and exists main strategies for selecting the negative dataset. A random set of non-AMPs with an equal number of the positive set is selected and thus, are considered as negative samples.

## Filtering and dataset balancing

After assembling both positive and negative datasets, one main goal is reducing homology bias, removing duplicate/inconsistent samples and gaining a more reliable sample set. Depending on the study, this step may be different. Based on the literature, there are three main policies for removing inconsistent/redundant AMPs:

- 1- Removing identical AMPs and sequence less than five amino acids.
- 2- Removing similarities within AMPs in the positive and non-AMPs in the negative datasets.
- 3- Removing similarities between AMPs in the positive and non-AMPs in the negative datasets.



**Figure 4.** Flowchart of the statistical and machine learning techniques for the detection of AMPs. (a) Collection of data and creation of datasets. (b) Feature selection. (c) Designing training and testing models. (d) Evaluation of model performance.

The Cluster Database at high identity with tolerance (CD-HIT) program is used as the major tool to detect similar samples (sequences) and to reduce homology bias and redundancy. However, different studies use different thresholds of identity to consider a pair of AMP sequences to be similar/redundant. In different AMPs prediction studies, this threshold varies from 20% to 100% (134). After filtered datasets there exists an imbalanced dataset and the size of the negative dataset is greater than the size of the positive dataset. These imbalanced datasets can create biases in the learning phase when a learning method is not a sufficient algorithm. Consequently, to ensure dataset balancing, a subset of the negative dataset equal to that of the positive dataset will be chosen (Figure 4a).

### Feature extraction

Selecting suitable algorithms to learn patterns and distinguish AMPs from other sequences, feature generation, extraction, engineering and selection became essential aspects of finding good representative features or informative features that

could capture AMP patterns and increase prediction accuracy. To select features that differentiate AMPs from non-AMPs, a feature selection method was almost used. Thus, the positive or negative samples (peptide sequences), according to their biological properties, are coded into numerical feature vectors that are used for learning the proposed model (135). Then, each peptide is encoded as a numerical feature vector based on suitable biological features, such as physicochemical properties, sequence composition, and structural features (Figure 4b).

### Training the predictors

Several learning algorithms will be used to compare the effectiveness of the variable features selected. This process must take place before the final prediction model is built. As part of feature selection, a subset of the most informative/discriminative features is selected and used to train the classifier. A good classifier can be chosen according to the performance of various classification methods. As a result of parameter optimization, the algorithm is trained

on a subset of the assembled dataset (training dataset), and then it can be evaluated and compared against the current state-of-the-art methods (Figure 4c).

### Performance assessment

There are two methods for evaluating models: Independent test (Train-Test) and K-fold cross-validation. In the independent test method, a dataset is divided into two sets, a Train-Validation dataset and a Test dataset. Afterward, the Train-validation set also splits into two subsets: the train set and the validation set. A training set is used to train models, and a test set is used to evaluate the models and select the best model according to performance via evaluation of the test set. On the other hand, the validation set evaluation results differ from the train set evaluation results, it shows the model has been overfitted to the train set. Lastly, the test set should be reported, and there should not be a huge difference between the validation and test sets.

K-fold cross-validation is a standard procedure for assessing the performance of a given classifier and is used to evaluate ML models on a limited data sample. In this process, the available dataset is randomly divided into  $k$  subsets without any overlap. One of the subsets is used as a test set dataset, and the other as training for assessing the predictor. Each subset is used exactly once as the test set, and the process is repeated  $k$  times. Finally, the average performance for all  $k$  test sets is considered (Figure 4d). Cross-validation with k-fold is typically used when there is only a limited amount of Train-Validation data, and the Train-Validation method is mostly used when massive amounts of data are accessible. K-fold is used in most classified AMPs methods.

### Model evaluation criteria

Assessing the performance of the AMPs prediction methods based on the four basic parameters which are explained as follows:

1-‘True positive’ (TP): the experimentally validated AMPs that have been correctly predicted by the prediction method.

2-‘True negative’ (TN): the non- AMPs sites that have been correctly predicted.

3-‘False positive’ (FP): the non- AMPs that have been incorrectly predicted as AMPs.

4-‘False negative’ (FN): the experimentally validated AMPs that have been incorrectly predicted non- AMPs.

The classification performance is often evaluated by accuracy, sensitivity (Recall), specificity, precision, F-measure and Matthews correlation coefficient (MCC). All performance criteria for AMPs prediction are shown in Equations (1–6). Alongwith the abovementioned measures which are known as threshold-dependent measures, ROC (receiver operative characteristic) and AUC (area under the ROC-curve) are two main threshold independent evaluation measures (136, 137). The most important assessment measures based on the abovementioned parameters have been described in the following section.

**Sensitivity:** Sensitivity or recall indicates the percentage of samples that have been predicted correctly.

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (1)$$

**Specificity:** Specificity shows the percentage of negative samples that have been predicted correctly as negative samples.

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (2)$$

**Accuracy:** Accuracy is a ratio between the correctly classified data points to the total number of samples (138).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (3)$$

**Precision:** Precision or positive prediction value (PPV) is shown as the ratio of the number of correctly predicted positive samples to the total number of positive samples (138).

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (4)$$

**F-measure:** This metric represents the harmonic mean of recall and precision, and is calculated as:

$$F\text{-measure} = \frac{2TP}{2TP + FP + FN} \times 100 \quad (5)$$

**Matthews Correlation Coefficient (MCC):** MCC shows the correlation between true and predicted labels (139).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + TN)(TP + FN)(TN + FP)(TN + FN)}} \times 100 \quad (6)$$

### Tools for AMPs prediction

Considering the high cost and labor-intensive experimental identification of AMPs, many computational methods have been proposed for the prediction of AMPs and their functional types which can be useful to design new and more effective antimicrobial agents. The attention to ML has been converted to a strongly data-driven approach. As a result, with development in computational methods and tools, supervised learning can be considered as a suitable strategy for leveraging large datasets for the high-throughput and high-accuracy classification of AMPs (34). Studies have shown differences in amino acid composition (AAC), the physico-chemical property, sequence order and the pattern of terminal residues in AMPs that can affect AMP prediction. Furthermore, it has been reported that the net charge, isoelectric point, composition and tendency for the secondary structure are different in the AMPs, like antibacterial, antifungal and antiviral activities, and as a result, these features should be used for learning algorithms for AMPs prediction (140). Many of these methods have been implemented as publicly accessible tools. However, there is still a lack of efficient prediction models to identify potential peptides with specific activities. An overview of existing predictive tools supporting AMP studies is presented in Table 3. Indeed an explanation of the newest comprehensive tool was provided.

**Table 3.** Online AMPs prediction tools

Acronym	Features	Method	Validation Method	Years	URL
Ensemble-AMPPred	517 features and a hybrid feature Amino acid composition, pseudo amino acid composition (PseAAC) in parallel and series correlation, and the details of the secondary structure conformation, composition-transition-distribution (CTD), various physical-chemical properties, antimicrobial propensity scale, and the percentage of different conformations in the peptide sequence.	Ensemble learning method	10-fold CV, Independent test	2021	<a href="http://ncrna-pred.com/Hybrid_AMPPred.htm">http://ncrna-pred.com/Hybrid_AMPPred.htm</a>
DBAASP	Physicochemical characteristics of peptides: normalized hydrophobic moment, normalized hydrophobicity, net charge, isoelectric point, penetration depth, tilt angle, disordered conformation propensity, linear moment, and propensity for <i>in vitro</i> aggregation	Cutoff discriminator	5-fold CV, Independent test	2021	<a href="http://dbaasp.org/home">http://dbaasp.org/home</a>
Deep-AmPEP30	AMPs in sequences	Deep Learning	10-fold CV, Independent test	2020	<a href="http://cbbio.online/AxPEP/">http://cbbio.online/AxPEP/</a>
AntiCP	Pseudo K-tuple RAAC ACPs in sequences Amino acid composition, dipeptide composition, terminus composition, binary profile, and hybrid features	<b>Support Vector Machine</b>	5-fold CV, Independent dataset	2020	<a href="https://webs.iitd.edu.in/raghava/antcp2/">https://webs.iitd.edu.in/raghava/antcp2/</a>
AmpGram	AMPs in sequences	<b>Random Forest</b>	5-fold CV, Independent dataset	2020	<a href="http://biongram.biotech.uni.wroc.pl/AmpGram/">http://biongram.biotech.uni.wroc.pl/AmpGram/</a>
AMPScanner	Numerical matrix from deep neural network (DNN)	Deep Learning	10-fold CV, Independent dataset	2018	<a href="https://www.dveltri.com/ascan/">https://www.dveltri.com/ascan/</a>
AntiMPmod	AMPs in structures	<b>Support Vector Machine</b>	5-fold CV Independent dataset	2018	<a href="https://webs.iitd.edu.in/raghava/antimpmod/">https://webs.iitd.edu.in/raghava/antimpmod/</a>
PseAAC	AMPs in sequences and structures	<b>Support Vector Machine</b>	10-fold CV, Independent dataset	2018	<a href="http://www.csbio.sjtu.edu.cn/bioinf/PseAAC/">http://www.csbio.sjtu.edu.cn/bioinf/PseAAC/</a>
MLAMP	PseAAC with the gray model (GM)	ML-SMOTE	Independent dataset	2016	<a href="http://www.jci-bioinfo.cn/MLAMP">http://www.jci-bioinfo.cn/MLAMP</a>
CAMP3	Sequence composition, physicochemical properties, and structural characteristics of amino acids	Support Vector Machine, Random Forests, and N-to-1 neural networks	10-fold CV, Independent dataset	2016	<a href="http://www.camp.bicnirrh.res.in/prediction.php">http://www.camp.bicnirrh.res.in/prediction.php</a>
CPPpred	cell-penetrating peptides in sequences		5-fold CV, Independent dataset	2013	<a href="http://bioware.ucd.ie/~compass/biowareweb/Server_pages/cpppred.php">http://bioware.ucd.ie/~compass/biowareweb/Server_pages/cpppred.php</a>
iAMP-2 L	Pseudo amino acid composition (PseAAC) incorporating five physicochemical properties	fuzzy K-nearest neighbor	Independent dataset	2013	<a href="http://www.jci-bioinfo.cn/iAMP-2L">http://www.jci-bioinfo.cn/iAMP-2L</a>
PeptideLocator	Bioactive peptides in sequences	Bidirectional Recursive Neural Networks	5-fold CV, Independent dataset	2013	<a href="http://bioware.ucd.ie/~compass/biowareweb/">http://bioware.ucd.ie/~compass/biowareweb/</a>
BAGEL3	Bacteriocins in DNA sequences	BLAST	—	2013	<a href="http://bagel.molgenrug.nl/">http://bagel.molgenrug.nl/</a>
CS-AMPPred	cysteine-stabilized AMPs in sequences	ORFs prediction tools Support Vector Machine	5-fold CV	2012	<a href="http://sourceforge.net/projects/csamppred/">http://sourceforge.net/projects/csamppred/</a>
AMPA	Antimicrobial index based on IC50 value	Antimicrobial propensity scale threshold	—	2011	<a href="http://tcofee.crg.cat/apps/ampa/guide.html">http://tcofee.crg.cat/apps/ampa/guide.html</a>



## Ensemble-AMPPred

In this work, several well-known single and ensemble (ML) approaches have been explored and evaluated based on balanced training datasets and two large testing datasets. They have demonstrated that the developed program with various predictive models has high performance in differentiating between AMPs and non-AMPs. The present work used a benchmark AMP dataset consisting of 920 AMPs and 920 non-AMPs in testing existing AMP prediction programs and detected the false predictive answers of each program are different. The results suggest that, due to the use of different models and features exist unpredictable answers that have a different distribution. Because of these limitations, each program should consider improvements, including minimizing false positives (FPs) and increasing predictive accuracy. Due to the mentioned points, they considered using of integrating different learning models using ensemble learning techniques. In ensemble learning techniques, using multiple predictors and ensemble methods for incorporating individual classification models (bagging and boosting) leads to a decrease in FPs and increasing prediction accuracy. In this study, AMP prediction models were developed using ensemble methods based on five different algorithms, as well as comparing four different single models (135).

Data from 15 public available bioactive peptide databases were collected by Ensembl-AMPPred, and peptides with sequences shorter than 10 amino acids were removed. CD-HIT program for reducing data redundancy with a threshold of 0.9 (90% sequence similarity) was used. Finally, 13 434 peptides were considered as positive sequence data. Using Uniprot, negative data was collected on proteins without antimicrobial activity and a secretory signal peptide position. The basic local alignment search tool (BLAST) was used to remove AMP matches and peptide sequences with lengths <10 amino acids. Furthermore, peptide sequences that had an identity greater than 25% were removed using the CD-HIT program. As a result, 37 595 peptides are considered negative. Also, the similarity between the positive and negative datasets was removed, and as a result, the sequence similarity between the training and testing datasets was calculated as 47.29%. Lastly, the training data includes 1800 sequences of peptides from the AMP dataset, and 1800 sequences from the non-AMP dataset were prepared (135). A dataset consisting of 517 peptide features was then extracted and to filter this feature vector a logistic regression was applied to create a hybrid feature vector using four preselected single features based on an equation for logistic regression. Finally, a performance comparison of eight single predictive models and five ensemble models was done and prediction accuracy was evaluated using 10-fold cross-validation. Although ensemble models have better performance than single models, these four models (RF, NN, SVM and LDA) showed the highest performance among single models. Ensemble models have better performance than single models. Nevertheless, four models (RF, NN, SVM and LDA) showed the highest performance among single models. Thus, Ensemble-AMPPred is an AMP predictor which able to high performance in differentiating between AMPs and non-AMPs in comparing other available methods (135).

## Conclusion

The AMPs constitute an important component of innate immunity and are effective against disease-causing pathogens. Multidrug-resistant bacteria (superbugs) are on the rise, making AMPs an important alternative to traditional antibiotics. However, the identification of AMPs through lab experiments is still expensive and time-consuming, and, most importantly, ineffective due to the staggering number of genomes being sequenced today. Therefore, the development of an efficient computational tool is essential to identify the best candidate AMPs with high accuracy before the *in vitro* experiments. Thus, the bioinformatics resources and the usage of computational tools to analyze AMPs data and their functional outcomes across species are crucial. There would be a significant interest in the development of computational methods for the reliable prediction of AMPs. Recently, many advanced computational methods and tools have been developed to predict AMPs, and many of them are publicly available. Therefore, in the near future, research in databases could be a key step in developing a typical new antimicrobial agent.

## Acknowledgements

The authors of the article are gratefully proud to appreciate Dr Hossein Dehghani Tafti suggestions and cooperation that allowed us to improve the quality of the manuscript.

## Funding

There is no funding available for this article.

## Conflict of interest

None declared.

## Author contributions

Shahin Ramazi devised the project, the main conceptual ideas, and the proof outline and designed the figures. Shahin Ramazi wrote the manuscript with support from Neda Mohammadi in consultation with all authors. All authors read and approved the final manuscript.

## References

1. Mishra,N.N. and Bayer,A.S. (2013) Correlation of cell membrane lipid profiles with daptomycin resistance in methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.*, 57, 1082–1085.
2. Kaprou,G.D., Bergšpica,I., Alexa,E.A. *et al.* (2021) Rapid methods for antimicrobial resistance diagnostics. *Antibiotics*, 10, 209.
3. Petrosillo,N. (2020) Infections: the emergency of the new millennium. In: *Nuclear Medicine in Infectious Diseases*. Springer Nature Switzerland AG, pp. 1–8.
4. Sohrabi,C., Alsafi,Z., O’neill,N. *et al.* (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int. J. Surg.*, 76, 71–76.
5. Kang,X., Dong,F., Shi,C. *et al.* (2019) DRAMP 2.0, an updated data repository of antimicrobial peptides. *Sci. Data*, 6, 1–10.

6. O'Neill, J. (2016) Tackling drug-resistant infections globally: final report and recommendations.
7. Fard, M.B., Fard, S.B., Ramazi, S. *et al.* (2021) Thrombosis in COVID-19 infection: role of platelet activation-mediated immunity. *Thromb. J.*, **19**, 1–11.
8. Chen, N., Zhou, M., Dong, X. *et al.* (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*, **395**, 507–513.
9. Pushpanathan, M., Gunasekaran, P. and Rajendhran, J. (2013) Antimicrobial peptides: versatile biological properties. *Int. J. Pept.*, **2013**, 1–15.
10. Pfalzgraff, A., Brandenburg, K. and Weindl, G. (2018) Antimicrobial peptides and their therapeutic potential for bacterial skin infections and wounds. *Front. Pharmacol.*, **9**, 281.
11. Koprivnjak, T. and Peschel, A. (2011) Bacterial resistance mechanisms against host defense peptides. *Cell. Mol. Life Sci.*, **68**, 2243–2254.
12. Park, S.-C., Park, Y. and Hahm, K.-S. (2011) The role of antimicrobial peptides in preventing multidrug-resistant bacterial infections and biofilm formation. *Int. J. Mol. Sci.*, **12**, 5971–5992.
13. Bahar, A.A. and Ren, D. (2013) Antimicrobial peptides. *Pharmaceuticals*, **6**, 1543–1575.
14. Chu, H.-L., Yip, B.-S., Chen, K.-H. *et al.* (2015) Novel antimicrobial peptides with high anticancer activity and selectivity. *PLoS One*, **10**, e0126390.
15. Brahmachary, M., Krishnan, S., Koh, J.L.Y. *et al.* (2004) ANTIMIC: a database of antimicrobial sequences. *Nucleic Acids Res.*, **32**, D586–D589.
16. Beisswenger, C. and Bals, R. (2005) Functions of antimicrobial peptides in host defense and immunity. *Curr. Protein Pept. Sci.*, **6**, 255–264.
17. Hiratsuka, T., Nakazato, M., Date, Y. *et al.* (1998) Identification of human  $\beta$ -defensin-2 in respiratory tract and plasma and its increase in bacterial pneumonia. *Biochem. Biophys. Res. Commun.*, **249**, 943–947.
18. Bals, R., Weiner, D.J., Meegalla, R.L. *et al.* (2001) Salt-independent abnormality of antimicrobial activity in cystic fibrosis airway surface fluid. *Am. J. Respir. Cell Mol. Biol.*, **25**, 21–25.
19. Ong, P.Y., Ohtake, T., Brandt, C. *et al.* (2002) Endogenous antimicrobial peptides and skin infections in atopic dermatitis. *N. Engl. J. Med.*, **347**, 1151–1160.
20. Joseph, S., Karnik, S., Nilawe, P. *et al.* (2012) ClassAMP: a prediction tool for classification of antimicrobial peptides. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **9**, 1535–1538.
21. Wang, G., Li, X. and Wang, Z. (2008) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.*, **37**, D933–D937.
22. Li, J., Koh, J.-J., Liu, S. *et al.* (2017) Membrane active antimicrobial peptides: translating mechanistic insights to design. *Front. Neurosci.*, **11**, 73.
23. Vijayakumar, S. and Lakshmi, P. (2015) ACP: a web server for prediction and design of anti-cancer peptides. *Int. J. Pept. Res. Ther.*, **21**, 99–106.
24. Wang, G. (2020) The antimicrobial peptide database provides a platform for decoding the design principles of naturally occurring antimicrobial peptides. *Protein Sci.*, **29**, 8–18.
25. Iwasaki, T., Ishibashi, J., Tanaka, H. *et al.* (2009) Selective cancer cell cytotoxicity of enantiomeric 9-mer peptides derived from beetle defensins depends on negatively charged phosphatidylserine on the cell surface. *Peptides*, **30**, 660–668.
26. Sun, J., Xia, Y., Li, D. *et al.* (2014) Relationship between peptide structure and antimicrobial activity as studied by de novo designed peptides. *Biochim. Biophys. Acta (BBA)-Biomembr.*, **1838**, 2985–2993.
27. Diamond, G., Beckloff, N., Weinberg, A., Kisich, K.O. (2009) The roles of antimicrobial peptides in innate host defense. *Curr. Pharm. Des.*, **15**, 2377–2392.
28. Porcelli, F., Ramamoorthy, A., Barany, G. *et al.* (2013) On the role of NMR spectroscopy for characterization of antimicrobial peptides. *Membr. Proteins*, **1063**, 159–180.
29. Wieprecht, T., Dathe, M., Epand, R.M. *et al.* (1997) Influence of the angle subtended by the positively charged helix face on the membrane activity of amphipathic, antibacterial peptides. *Biochemistry*, **36**, 12869–12880.
30. Giuliani, A. and Rinaldi, A.C. (2010) Antimicrobial peptides. In: *Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa NJ, p. 618.
31. Zhang, W., Xu, X., Zhang, J., Ye, T., Zhou, Q., Xu, Y. *et al.* (2022) Discovery and characterization of a New Crustin antimicrobial peptide from amphibalanus amphitrite. *Pharmaceutics*, **14**, 413.
32. Mack, A., Relman, D.A. and Choffnes, E.R. (2011) *Antibiotic Resistance: Implications for Global Health and Novel Intervention Strategies: Workshop Summary*. National Academies Press, Washington DC.
33. Khosravi, M., Kazemi Faramarzi, F., Mohammad Beigi, M. *et al.* (2013) Predicting antibacterial peptides by the concept of Chou's pseudo-amino acid composition and machine learning methods. *Protein Pept. Lett.*, **20**, 180–186.
34. Lee, E.Y., Lee, M.W., Fulan, B.M. *et al.* (2017) What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning? *Interface Focus*, **7**, 20160153.
35. Wang, G., Li, X. and Wang, Z. (2015) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.*, **44**, gkv1278.
36. Huan, Y., Kong, Q., Mou, H. *et al.* (2020) Antimicrobial peptides: classification, design, application and research progress in multiple fields. *Front. Microbiol.*, **11**, 2559.
37. Kumar, P., Kizhakkedathu, J.N. and Straus, S.K. (2018) Antimicrobial peptides: diversity, mechanism of action and strategies to improve the activity and biocompatibility in vivo. *Biomolecules*, **8**, 4.
38. Domingues, M.M., Felício, M.R. and Gonçalves, S. (2019) Antimicrobial peptides: effect on bacterial cells. In: *Atomic Force Microscopy*. Humana Press, New York, pp. 233–242.
39. Ruiz, J., Calderon, J., Rondón-Villarreal, P. *et al.* (2014) Analysis of structure and hemolytic activity relationships of antimicrobial peptides (AMPs). In: *Advances in Computational Biology*. Springer, Springer-Cham, pp. 253–258.
40. Thomas, S., Karnik, S., Barai, R.S. *et al.* (2009) CAMP: a useful resource for research on antimicrobial peptides. *Nucleic Acids Res.*, **38**, D774–D780.
41. 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**, 4310.
42. Ingham, A.B. and Moore, R.J. (2007) Recombinant production of antimicrobial peptides in heterologous microbial systems. *Biotechnol. Appl. Biochem.*, **47**, 1–9.
43. Nissen-Meyer, J. and Nes, I.F. (1997) Ribosomally synthesized antimicrobial peptides: their function, structure, biogenesis, and mechanism of action. *Arch. Microbiol.*, **167**, 67–77.
44. Papagianni, M. (2003) Ribosomally synthesized peptides with antimicrobial properties: biosynthesis, structure, function, and applications. *Biotechnol. Adv.*, **21**, 465–499.
45. Münzker, L., Oddo, A. and Hansen, P.R. (2017) Chemical synthesis of antimicrobial peptides. In: *Antimicrobial Peptides*. Humana press, New York, pp. 35–49.
46. Ntwasa, M. (2012) Cationic peptide interactions with biological macromolecules. *Binding Protein*, 139–164.
47. Breitling, F., Löffler, F., Schirwitz, C. *et al.* (2011) Alternative setups for automated peptide synthesis. *Mini Rev. Org. Chem.*, **8**, 121–131.
48. Andreu, D. and Rivas, L. (1998) Animal antimicrobial peptides: an overview. *Pept. Sci.*, **47**, 415–433.

49. Powers, J.-P.S. and Hancock, R.E. (2003) The relationship between peptide structure and antibacterial activity. *Peptides*, **24**, 1681–1691.
50. Zhang, X.L., Selsted, M.E. and Pardi, A. (1992) NMR studies of defensin antimicrobial peptides. 1. Resonance assignment and secondary structure determination of rabbit NP-2 and human HNP-1. *Biochemistry*, **31**, 11348–11356.
51. Pardi, A., Zhang, X.L., Selsted, M.E. *et al.* (1992) NMR studies of defensin antimicrobial peptides. 2. Three-dimensional structures of rabbit NP-2 and human HNP-1. *Biochemistry*, **31**, 11357–11364.
52. Ramamoorthy, A., Thennarasu, S., Lee, D.-K. *et al.* (2006) Solid-state NMR investigation of the membrane-disrupting mechanism of antimicrobial peptides MSI-78 and MSI-594 derived from magainin 2 and melittin. *Biophys. J.*, **91**, 206–216.
53. Porcelli, F., Buck-Koehntop, B.A., Thennarasu, S. *et al.* (2006) Structures of the dimeric and monomeric variants of magainin antimicrobial peptides (MSI-78 and MSI-594) in micelles and bilayers, determined by NMR spectroscopy. *Biochemistry*, **45**, 5793–5799.
54. Nijnik, A. and Hancock, R.E. (2009) The roles of cathelicidin LL-37 in immune defences and novel clinical applications. *Curr. Opin. Hematol.*, **16**, 41–47.
55. Hunter, H.N., Demcoe, A.R., Jenssen, H. *et al.* (2005) Human lactoferricin is partially folded in aqueous solution and is better stabilized in a membrane mimetic solvent. *Antimicrob. Agents Chemother.*, **49**, 3387–3395.
56. Beevers, A.J. and Dixon, A.M. (2010) Helical membrane peptides to modulate cell function. *Chem. Soc. Rev.*, **39**, 2146–2157.
57. Arias, M., McDonald, L.J., Haney, E.F. *et al.* (2014) Bovine and human lactoferricin peptides: chimeras and new cyclic analogs. *Biomaterials*, **27**, 935–948.
58. Rozek, T., Bowie, J.H., Wallace, J.C. *et al.* (2000) The antibiotic and anticancer active aurein peptides from the Australian Bell Frogs *Litoria aurea* and *Litoria raniformis*. Part 2. Sequence determination using electrospray mass spectrometry 1. *Rapid Comm. Mass Spectrom.*, **14**, 2002–2011.
59. Xu, T., Levitz, S., Diamond, R. *et al.* (1991) Anticandidal activity of major human salivary histatins. *Infect. Immun.*, **59**, 2549–2554.
60. Khurshid, Z., Najeeb, S., Mali, M. *et al.* (2017) Histatin peptides: pharmacological functions and their applications in dentistry. *Saudi Pharm. J.*, **25**, 25–31.
61. Gennaro, R., Skerlavaj, B. and Romeo, D. (1989) Purification, composition, and activity of two bactericins, antibacterial peptides of bovine neutrophils. *Infect. Immun.*, **57**, 3142–3146.
62. Tam, J.P., Wang, S., Wong, K.H. *et al.* (2015) Antimicrobial peptides from plants. *Pharmaceuticals*, **8**, 711–757.
63. Porto, W., Pires, A. and Franco, O. (2017) Computational tools for exploring sequence databases as a resource for antimicrobial peptides. *Biotechnol. Adv.*, **35**, 337–349.
64. Ulm, H., Wilmes, M., Shai, Y. *et al.* (2012) Antimicrobial host defensins—specific antibiotic activities and innate defense modulation. *Front. Immunol.*, **3**, 249.
65. Huan, Y., Kong, Q., Mou, H. *et al.* (2020) Antimicrobial peptides: classification, design, application and research progress in multiple fields. *Front. Microbiol.*, **11**, 1–21.
66. Brogden, K.A. (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.*, **3**, 238–250.
67. Wimley, W.C. (2010) Describing the mechanism of antimicrobial peptide action with the interfacial activity model. *ACS Chem. Biol.*, **5**, 905–917.
68. Shai, Y., Bach, D. and Yanovsky, A. (1990) Channel formation properties of synthetic pardaxin and analogues. *J. Biol. Chem.*, **265**, 20202–20209.
69. Uematsu, N. and Matsuzaki, K. (2000) Polar angle as a determinant of amphipathic  $\alpha$ -helix-lipid interactions: a model peptide study. *Biophys. J.*, **79**, 2075–2083.
70. Lee, T.-H., N Hall, K. and Aguilar, M.-I. (2016) Antimicrobial peptide structure and mechanism of action: a focus on the role of membrane structure. *Curr. Top Med. Chem.*, **16**, 25–39.
71. Cheng, J.T., Hale, J.D., Elliot, M. *et al.* (2009) Effect of membrane composition on antimicrobial peptides aurein 2.2 and 2.3 from Australian southern bell frogs. *Biophys. J.*, **96**, 552–565.
72. Sitaram, N. and Nagaraj, R. (1999) Interaction of antimicrobial peptides with biological and model membranes: structural and charge requirements for activity. *Biochim. Biophys. Acta (BBA)-Biomembr.*, **1462**, 29–54.
73. Rozek, A., Friedrich, C.L. and Hancock, R.E. (2000) Structure of the bovine antimicrobial peptide indolicidin bound to dodecylphosphocholine and sodium dodecyl sulfate micelles. *Biochemistry*, **39**, 15765–15774.
74. Shai, Y. (2002) From innate immunity to de-novo designed antimicrobial peptides. *Curr. Pharm. Des.*, **8**, 715–725.
75. Mahlapuu, M., Håkansson, J., Ringstad, L. *et al.* (2016) Antimicrobial peptides: an emerging category of therapeutic agents. *Front. Cell. Infect. Microbiol.*, **6**, 194.
76. Mousavizadegan, M. and Mohabatkar, H. (2018) Computational prediction of antifungal peptides via Chou's PseAAC and SVM. *J. Bioinform. Comput. Biol.*, **16**, 1850016.
77. Chen, C.H. and Lu, T.K. (2020) Development and challenges of antimicrobial peptides for therapeutic applications. *Antibiotics*, **9**, 24.
78. Jhong, J.-H., Yao, L., Pang, Y. *et al.* (2022) dbAMP 2.0: updated resource for antimicrobial peptides with an enhanced scanning method for genomic and proteomic data. *Nucleic Acids Res.*, **50**, D460–D470.
79. Hollmann, A., Martinez, M., Maturana, P. *et al.* (2018) Antimicrobial peptides: interaction with model and biological membranes and synergism with chemical antibiotics. *Front. Chem.*, **6**, 204.
80. Malanovic, N. and Lohner, K. (2016) Antimicrobial peptides targeting gram-positive bacteria. *Pharmaceuticals*, **9**, 59.
81. Thakur, N., Qureshi, A. and Kumar, M. (2012) AVPPred: collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res.*, **40**, W199–W204.
82. da Mata, É.C.G., Mourão, C.B.F., Rangel, M. *et al.* (2017) Antiviral activity of animal venom peptides and related compounds. *J. Venom. Anim. Toxins Incl. Trop. Dis.*, **23**, 3.
83. Findlay, E.G. Currie, S.M. and Davidson, D.J. (2013) Cationic host defence peptides: potential as antiviral therapeutics. *BioDrugs*, **27**, 479–493.
84. Fernández de Ullivarri, M., Arbulu, S., Garcia-Gutierrez, E. *et al.* (2020) Antifungal peptides as therapeutic agents. *Front. Cell Infect. Microbiol.*, **10**, 105.
85. Ciociola, T., Giovati, L., Conti, S. *et al.* (2016) Natural and synthetic peptides with antifungal activity. *Future Med. Chem.*, **8**, 1413–1433.
86. Lupetti, A., Danesi, R., Campa, M. *et al.* (2002) Molecular basis of resistance to azole antifungals. *Trends Mol. Med.*, **8**, 76–81.
87. Matejuk, A., Leng, Q., Begum, M. *et al.* (2010) Peptide-based antifungal therapies against emerging infections. *Drugs Future*, **35**, 197.
88. Akkam, Y. (2016) A review of antifungal peptides: basis to new era of antifungal drugs. *Jordan J. Pharm. Sci.*, **9**, 51–75.
89. Ravi, C., Jeyashree, A. and Devi, K.R. (2011) Antimicrobial peptides from insects: an overview. *Res. Biotechnol.*, **2**, 1–7.
90. Rollins-Smith, L.A., Reinert, L.K., O'Leary, C.J. *et al.* (2005) Antimicrobial peptide defenses in amphibian skin. *Integr. Comp. Biol.*, **45**, 137–142.
91. Pingitore, E.V., Salvucci, E., Sesma, F. *et al.* (2007) Different strategies for purification of antimicrobial peptides from lactic acid bacteria (LAB). *Commun. Curr. Res. Educ. Top. Trends Appl. Microbiol.*, **1**, 557–568.

92. Nagaraj,G., Uma,M., Shivayogi,M. *et al.* (2001) Antimalarial activities of peptide antibiotics isolated from fungi. *Antimicrob. Agents Chemother.*, **45**, 145–149.
93. Tincu,J.A. and Taylor,S.W. (2004) Antimicrobial peptides from marine invertebrates. *Antimicrob. Agents Chemother.*, **48**, 3645–3654.
94. Mishra,B. and Wang,G. (2012) Ab initio design of potent anti-MRSA peptides based on database filtering technology. *J. Am. Chem. Soc.*, **134**, 12426–12429.
95. Bondaryk,M., Staniszewska,M., Zielińska,P. *et al.* (2017) Natural antimicrobial peptides as inspiration for design of a new generation antifungal compounds. *J. Fungi*, **3**, 46.
96. Garrigues,S., Gandía,M., Borics,A. *et al.* (2017) Mapping and identification of antifungal peptides in the putative antifungal protein AfpB from the filamentous fungus *Penicillium digitatum*. *Front. Microbiol.*, **8**, 592.
97. Siegel,R., Ma,J., Zou,Z. *et al.* (2014) Cancer statistics, 2014. *CA Cancer J. Clin.*, **64**, 9–29.
98. Huang,Y.-B., Wang,X.-F., Wang,H.-Y. *et al.* (2011) Studies on mechanism of action of anticancer peptides by modulation of hydrophobicity within a defined structural framework. *Mol. Cancer Ther.* molcanther. 0811.2010 **10**, 416–426.
99. Naumov,G.N., Townson,J.L., MacDonald,I.C. *et al.* (2003) Ineffectiveness of doxorubicin treatment on solitary dormant mammary carcinoma cells or late-developing metastases. *Breast Cancer Res. Treat.*, **82**, 199–206.
100. Parkin,D.M. (2006) The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer*, **118**, 3030–3044.
101. Attiê,R., Chinen,L.T.D., Yoshioka,E.M. *et al.* (2014) Acute bacterial infection negatively impacts cancer specific survival of colorectal cancer patients. *WJG*, **20**, 13930.
102. Vedham,V., Divi,R.L., Starks,V.L. *et al.* (2014) Multiple infections and cancer: implications in epidemiology. *Technol. Cancer Res. Treat.*, **13**, 177–194.
103. Applegate,K.T., Besson,S., Matov,A. *et al.* (2011) plusTip-Tracker: quantitative image analysis software for the measurement of microtubule dynamics. *J. Struct. Biol.*, **176**, 168–184.
104. Freire,J.M., Gaspar,D., Veiga,A.S. *et al.* (2015) Shifting gear in antimicrobial and anticancer peptides biophysical studies: from vesicles to cells. *J. Pept. Sci.*, **21**, 178–185.
105. Tyagi,A., Tuknait,A., Anand,P. *et al.* (2014) CancerPPD: a database of anticancer peptides and proteins. *Nucleic Acids Res.*, **43**, D837–D843.
106. Lei,J., Sun,L., Huang,S., Zhu,C., Li,P., He,J. *et al.* (2019) The antimicrobial peptides and their potential clinical applications. *Am J transl res*, **11**, 3919.
107. Bevers,E., Comfurius,P. and Zwaal,R. (1996) Regulatory mechanisms in maintenance and modulation of transmembrane lipid asymmetry: pathophysiological implications. *Lupus*, **5**, 480–487.
108. Gaspar,D., Veiga,A.S. and Castanho,M.A. (2013) From antimicrobial to anticancer peptides. A review. *Front. Microbiol.*, **4**.
109. Buri,M.V., Domingues,T.M., Paredes-Gamero,E.J. *et al.* (2013) Resistance to degradation and cellular distribution are important features for the antitumor activity of gomesin. *PLoS One*, **8**, e80924.
110. Ting,C.-H., Huang,H.-N., Huang,T.-C. *et al.* (2014) The mechanisms by which pardaxin, a natural cationic antimicrobial peptide, targets the endoplasmic reticulum and induces c-FOS. *Biomaterials*, **35**, 3627–3640.
111. Schweizer,F. (2009) Cationic amphiphilic peptides with cancer-selective toxicity. *Eur. J. Pharmacol.*, **625**, 190–194.
112. Lacerda,A.F., Pelegrini,P.B., de Oliveira,D.M. *et al.* (2016) Antiparasitic peptides from arthropods and their application in drug therapy. *Front. Microbiol.*, **7**, 91.
113. Torrent,M., Pulido,D., Rivas,L. *et al.* (2012) Antimicrobial peptide action on parasites. *Curr. Drug Targets*, **13**, 1138–1147.
114. Pretzel,J., Mohring,F., Rahlfs,S. *et al.* (2013) Antiparasitic peptides. In: *Yellow Biotechnology I*. Springer, Berlin, Heidelberg, pp. 157–192.
115. Mor,A. (2009) Multifunctional host defense peptides: antiparasitic activities. *FEBS J.*, **276**, 6474–6482.
116. Pirtskhalava,M., Armstrong,A.A., Grigolava,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.
117. Ye,G., Wu,H., Huang,J. *et al.* (2020) LAMP2: a major update of the database linking antimicrobial peptides. *Database*, **2020**, 1–6.
118. Gómez,E.A., Giraldo,P. and Orduz,S. (2017) InverPep: a database of invertebrate antimicrobial peptides. *J. Glob. Antimicrob. Resist.*, **8**, 13–17.
119. Waghu,F.H., Barai,R.S., Gurung,P. *et al.* (2015) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.*, **44**, gkv1051.
120. Wang,G., Li,X. and Wang,Z. (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.*, **44**, D1087–D1093.
121. Lata,S., Sharma,B. and Raghava,G.P. (2007) Analysis and prediction of antibacterial peptides. *BMC Bioinform.*, **8**, 1–10.
122. Cherkasov,A., Hilpert,K., Jenssen,H. *et al.* (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**, 65–74.
123. Fjell,C.D., Jenssen,H., Fries,P. *et al.* (2008) Identification of novel host defense peptides and the absence of  $\alpha$ -defensins in the bovine genome. *Proteins*, **73**, 420–430.
124. Wang,P., Hu,L., Liu,G. *et al.* (2011) Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. *PLoS One*, **6**, e18476.
125. Torrent,M., Andreu,D., Nogués,V.M. *et al.* (2011) Connecting peptide physicochemical and antimicrobial properties by a rational prediction model. *PLoS One*, **6**, e16968.
126. Maccari,G., Di Luca,M., Nifosi,R. *et al.* (2013) Antimicrobial peptides design by evolutionary multiobjective optimization. *PLoS Comput. Biol.*, **9**, e1003212.
127. Giguere,S., Laviolette,F., Marchand,M. *et al.* (2015) Machine learning assisted design of highly active peptides for drug discovery. *PLoS Comput. Biol.*, **11**, e1004074.
128. Schneider,P., Müller,A.T., Gabernet,G. *et al.* (2017) Hybrid network model for “deep learning” of chemical data: application to antimicrobial peptides. *Mol. Inform.*, **36**, 1600011.
129. Xiao,X., Shao,Y.-T., Cheng,X. *et al.* (2021) iAMP-CA2L: a new CNN-BiLSTM-SVM classifier based on cellular automata image for identifying antimicrobial peptides and their functional types. *Brief. Bioinf.*, **22**, bbab209.
130. Torrent,M., Victoria Noguez,M. and Boix,E. (2012) Discovering new in silico tools for antimicrobial peptide prediction. *Curr. Drug Targets*, **13**, 1148–1157.
131. Lin,Y., Cai,Y., Liu,J. *et al.* (2019) An advanced approach to identify antimicrobial peptides and their function types for penaeus through machine learning strategies. *BMC Bioinform.*, **20**, 291.
132. Veltri,D., Kamath,U. and Shehu,A. (2018) Deep learning improves antimicrobial peptide recognition. *Bioinformatics*, **34**, 2740–2747.
133. Lin,W. and Xu,D. (2016) Imbalanced multi-label learning for identifying antimicrobial peptides and their functional types. *Bioinformatics*, **32**, 3745–3752.
134. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
135. 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*, **12**, 137.
136. Khalili,E., Kouchaki,S., Ramazi,S. *et al.* (2020) Machine learning techniques for soybean charcoal rot disease prediction. *Front. Plant Sci.*, 11.
  137. Ramazi,S. and Zahiri,J. (2021) Posttranslational modifications in proteins: resources, tools and prediction methods. *Database*, **2021**, 1–20.
  138. Sokolova,M., Japkowicz,N. and Szpakowicz,S. (eds). (2006) Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In: *Australasian Joint Conference on Artificial Intelligence*. Springer, Berlin, Heidelberg, 1–2.
  139. Boughorbel,S., Jarray,F. and El-Anbari,M. (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS One*, **12**, e0177678.
  140. Chung,C.-R., Jhong,J.-H., Wang,Z. *et al.* (2020) Characterization and identification of natural antimicrobial peptides on different organisms. *Int. J. Mol. Sci.*, **21**, 986.