

2023

The Antimicrobial Peptide Database is 20 Years Old: Recent Developments and Future Directions

Guangshun Wang

Tell us how you used this information in this [short survey](#).

Follow this and additional works at: https://digitalcommons.unmc.edu/com_pathmicro_articles



Part of the [Medical Microbiology Commons](#), and the [Pathology Commons](#)

The antimicrobial peptide database is 20 years old: Recent developments and future directions

Guangshun Wang 

Department of Pathology and Microbiology, College of Medicine, University of Nebraska Medical Center, Omaha, Nebraska, USA

Correspondence

Guangshun Wang, Department of Pathology and Microbiology, College of Medicine, University of Nebraska Medical Center, DRC2 5034, 985900 Nebraska Medical Center, Omaha, NE 68198-5900 USA.

Email: gwang@unmc.edu

Funding information

National Institutes of Health

Review Editor: Nir Ben-Tal

Abstract

In 2023, the Antimicrobial Peptide Database (currently available at <https://aps.unmc.edu>) is 20-years-old. The timeline for the APD expansion in peptide entries, classification methods, search functions, post-translational modifications, binding targets, and mechanisms of action of antimicrobial peptides (AMPs) has been summarized in our previous *Protein Science* paper. This article highlights new database additions and findings. To facilitate antimicrobial development to combat drug-resistant pathogens, the APD has been re-annotating the data for antibacterial activity (active, inactive, and uncertain), toxicity (hemolytic and nonhemolytic AMPs), and salt tolerance (salt sensitive and insensitive). Comparison of the respective desired and undesired AMP groups produces new knowledge for peptide design. Our unification of AMPs from the six life kingdoms into “natural AMPs” enabled the first comparison with globular or transmembrane proteins. Due to the dominance of amphipathic helical and disulfide-linked peptides, cysteine, glycine, and lysine in natural AMPs are much more abundant than those in globular proteins. To include peptides predicted by machine learning, a new “predicted” group has been created. Remarkably, the averaged amino acid composition of predicted peptides is located between the lower bound of natural AMPs and the upper bound of synthetic peptides. Synthetic peptides in the current APD, with the highest cationic and hydrophobic amino acid percentages, are mostly designed with varying degrees of optimization. Hence, natural AMPs accumulated in the APD over 20 years have laid the foundation for machine learning prediction. We discuss future directions for peptide discovery. It is anticipated that the APD will continue to play a role in research and education.

KEYWORDS

antimicrobial peptides, data re-annotation, machine learning, natural AMPs, predicted peptides, synthetic AMPs

1 | INTRODUCTION

The antimicrobial peptide (AMP) research remains to be active and appealing at least due to the following three reasons: (1) our attempt to decipher the functional roles

of AMPs in a variety of organisms; (2) our growing interest in microbiota that can be shaped by AMPs; and (3) our desire of developing AMPs into novel antibiotics (Fjell et al., 2011; Gallo & Hooper, 2012; Hanson et al., 2023; Lazzaro et al., 2020; Salzman &

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

Bevins, 2013). The AMP research may be classified into three main phases. The first phase was started with the discovery of lysozyme in 1922 (Fleming, 1932), which is the first antimicrobial polypeptide according to Robert Lehrer (Personal communication). This early stage of AMP search overlapped with the golden era of antibiotics (1940s–1960s), characterized by the discovery of multiple nonribosomally synthesized peptide antibiotics. Gramicidin was documented as the first peptide antibiotic for clinical use (Dubos, 1939). Colistin, a cyclic lipopeptide, is the last resort of peptide antibiotic against gram-negative pathogenic bacteria (Stansly et al., 1947). Although the discovery of lysozyme did not win Alexander Fleming a Nobel prize, penicillin (discovered in 1928) did. Notably, penicillin is a derivative of the dipeptide cysteine-valine. The second phase was initiated with the discovery of cecropins, defensins, pardaxins, daptomycin, magainins, bactenecin, and histatins in the 1980s (Eliopoulos et al., 1986; Ganz et al., 1985; Gennaro et al., 1989; Oppenheim et al., 1988; Steiner et al., 1981; Thompson et al., 1986; Zasloff, 1987). The invention of two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) and its application in structural determination of small proteins and peptides in the 1970s–1980s (Jeener & Alewaeters, 2016; Wüthrich, 1983) enabled us to see how AMPs look like (Gesell et al., 1997; Wang et al., 2014). The availability of 3D structures of these cationic peptides in membrane-mimetic environments consolidated the amphipathic idea. These structures stimulated the research on the mechanisms of action of antimicrobial peptides, leading to stave-barrel, carpet, detergent, and interfacial models (Ludtke et al., 1996; Oren & Shai, 1998; Vogt & Bechinger, 1999; Wimley, 2010). The research at this phase is also characterized by the efforts in developing magainins into new antibiotics (Jacob & Zasloff, 1994) because it is believed that the rapid membrane disruption by cationic AMPs makes it difficult for pathogens to develop resistance (Zasloff, 2002). The journey in elucidating the molecular mechanism of AMP expression in drosophila resulted in the discovery of Toll and Imd pathways (Lemaitre et al., 1996). This discovery won Jules Hoffmann a Nobel prize in 2011. This milestone discovery inspired the continued search of novel functional roles of these innate immune peptides in coat color, sleep and lifespan till today (Candille et al., 2007; Hanson & Lemaitre, 2023; Toda et al., 2019). The third phase was triggered by the applications of the proteomic and genomic technologies to AMP research. We entered the omic era with the completion of the sequencing of the human genome in 2003 (Human genome project website, visited 2023). The growing research on microbiota added a new dimension to AMP research due to its connection with a variety of

human diseases (Wang et al., 2023; Wehkamp et al., 2005). AMPs are important players that shape microbiota (de San et al., 2022; Pierre et al., 2023; Snelders et al., 2021). Remarkably, a Paneth cell peptide YY was able to selectively eliminate pathogenic hyphae but not the commensal yeast form of *Candida albicans* (Pierre et al., 2023). The omic approaches have also been utilized to AMP discovery (Bishop et al., 2017; Conlon et al., 2006; Lai, 2010; Moyer et al., 2021). Our imagination has gone beyond the classic territories and reached deep in the sea and uncultivable bacteria in soil (Lee et al., 2001; Ling et al., 2015). These studies increased our chances of discovering novel peptide antibiotics to combat drug-resistant bacteria, fungi, viruses, and parasites. The omic approaches have also been applied to the study of immune regulation of host defense peptides, significantly expanding our picture on the roles of AMPs in immune response (Mansour et al., 2014; Mookherjee et al., 2007; Mookherjee et al., 2009).

In the spirit of omics, the Wang laboratory established the antimicrobial peptide database (APD) in 2003 (Wang & Wang, 2004). It is 20-year-old at the time when this manuscript was completed. This database was originally conceived as a tool for antimicrobial and anticancer drug development. Hence, the first version of the APD constructed the following database interfaces: About, Database, Prediction, Peptide design, Statistic Data, Links, and Contact info (to view a picture of the historic website, see Reference Wang, Zietz, et al., 2022). The APD provided the first peptide calculation, empirical prediction, and peptide design tools for AMPs. These database tools were welcomed as echoed by the development of the Grammar approach for peptide design in 2006 and machine learning prediction in 2007, respectively (Lata et al., 2007; Loose et al., 2006). After we demonstrated the template-based design based on our database (Wang et al., 2005), we also conducted database screening to identify novel peptides against human immunodeficiency virus type 1 (HIV-1) or methicillin-resistant *Staphylococcus aureus* (MRSA; Menousek et al., 2012; Wang et al., 2010). This database provided a statistical tool, allowing users to view the amino acid composition signature for each AMP or a family of peptides (e.g., antibacterial, antiviral, hemolytic, and anticancer). The APD first defined frequently occurring amino acids (>10%), which depend on the source, structure, and activity of AMPs. These abundant amino acids constituted a minimal set for de novo peptide design (Wang et al., 2009). The APD created a powerful search engine by stringing the filters together into a pipeline so that the users can filter the peptide information at their will. Coupled with the statistical information, these database filters laid the foundation for our development of the

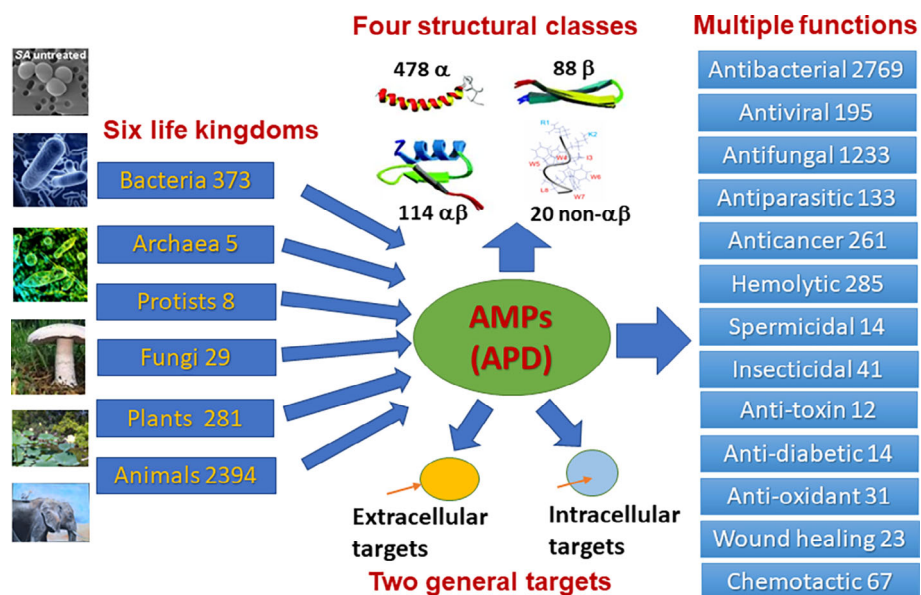
database filtering technology for designing novel antimicrobials from the beginning (*ab initio*; Mishra & Wang, 2012a). Our database filtering idea had been extended from *in silico* to *in vitro* and *in vivo*, leading to the discovery of the importance of low cationicity for systemic efficacy (Mishra et al., 2019).

The subsequent two database versions significantly expanded peptide entries, search functions, chemical modification types, 3D structures, binding targets, and mechanisms of action (Wang et al., 2009, 2016). Our database took the lead in systematically annotating over 26 searchable peptide functions (e.g., antibacterial, anti-diabetic, wound healing, anti-toxin, anticancer, antiviral) for two decades. Some of these peptide functions are depicted in Figure 1. In addition, the APD has annotated over 26 types of post-translational modifications of AMPs in a systematic manner (searchable using the XX code; Wang, Zietz, et al., 2022; Wang et al., 2016). A complete list of these functions and chemical modifications can be found in our previous publication (Wang, Zietz, et al., 2022). The APD considered the effects of chemical modifications on peptide properties such as net charge.

The data scope of the APD evolved with time. While the APD2 collected all peptides reported as AMPs in the literature, including those without activity data (Wang et al., 2009), the APD3 established a set of criteria for data registration by focusing on natural AMPs with known amino acid sequences and activity (minimal inhibitory concentration MIC <100 μ M) (Wang et al., 2016). The APD3 also unified the peptide classification methods based on biological sources (six life kingdoms), 3D structures (α , β , $\alpha\beta$, and non- $\alpha\beta$ structures), and peptide chain bonding patterns (UCLL, UCSS, UCSB, and UCBB) (Wang, 2022). The six life kingdoms in the APD include

bacteria, archaea, protists, fungi, plants, and animals. Currently known AMPs for each kingdom are summarized in Figure 1. The four unified structural classes are α -helical peptides (α class), β -sheet peptides (β class), peptides with both α and β structures ($\alpha\beta$ class), and AMPs with neither α nor β structures (non- $\alpha\beta$ class). Representative structures for each class are depicted in Figure 1. Due to limited known 3D structures (\sim 13%) for natural AMPs determined by multidimensional NMR spectroscopy and/or x-ray crystallography, the APD also adopted a universal classification scheme based on their covalent bonding patterns. Linear peptides (L) such as human cathelicidin LL-37 and magainins are represented with UCLL (searchable in the name field of the APD). Some linear peptides require two independent chains (LL) to be active. Sidechain-sidechain linked peptides such as defensins (via disulfide bonds) and lantibiotics (via thioether bridges) are searchable using UCSS. A third class (UCSB) has a chemical bond from the sidechain of one amino acid to the C or N-terminus (backbone) of the peptide. Finally, the UCBB peptides contain a chemical bond that connects the N and C-termini of the peptide. In addition, the APD provides a platform for understanding the design principles of natural AMPs (Decker et al., 2022; Lakshmaiah Narayana et al., 2020; Wang, 2020a; Wang, 2020b). Peptides with different activities, structures, and mechanisms of action can all be grouped and analyzed. For instance, the amino acid signatures for AMPs against gram-negative (more lysine) and gram-positive pathogens (more leucine) clearly differ (Wang, 2020a). While leucine is dominant in helical AMPs against gram-positive bacteria, alanine, glycine, and lysine are abundant in helical peptides against gram-negative pathogens (Wang, 2020a). For 1000 amphibian

FIGURE 1 Current statistics for natural antimicrobial peptides in the antimicrobial peptide database (APD; <https://aps.unmc.edu>). Peptide counts as of July 2023 are indicated for the six life kingdoms, four structural classes and multiple functions (activities) first classified in this database. Select images are either personal pictures or were taken from the internet. Three-dimensional structures for AMPs were determined by the author or downloaded from the Protein Database Bank (LL-37: 2K6O; RTD-1, 2LYF; MGD-1: 1FJN, and WW295: 6NM3).



AMPs, there is a linear increase in cationicity (mainly due to lysine) but a decrease in hydrophobic ratio (primarily due to leucine) with peptide length (Wang, 2020b). Remarkably, the averaged arginine content shows a linear correlation with hydrophobic ratio for all the AMPs in the APD (Lakshmaiah Narayana et al., 2020).

Our database expansion, reconfiguration, and information re-annotation continue. This article reports our recent new developments since 2021. These include (1) the unification of the AMPs from the six life kingdoms under the same searchable umbrella “natural AMPs”; (2) the definition of a new “predicted” group for AMPs predicted by machine learning and other methods (e.g., sequence alignment) followed by activity validation; (3) establishment of a systematic scheme for a complete record of reported antimicrobial activity; (4) definition and classification of hemolytic and non-hemolytic AMPs; and (5) definition and systematic annotation of salt effects on peptide activity. Our successful establishment of these new features in the APD not only further enhanced this database capability but also strengthened the notion that our database platform is user-friendly, flexible, and expandable. These re-annotations enabled us to compare natural AMPs with globular proteins and painted the amino acid landscapes for natural, predicted, and synthetic AMPs. Our results shine a novel light on natural AMPs, peptide prediction, and design. Our core data set for natural AMPs accumulated over 20 years has laid the foundation for both prediction and design of novel peptide antibiotics. Finally, we discuss future directions for AMP discovery.

2 | NEW FEATURES AND RESULTS

2.1 | A look at data growth and peptide parameter space from a new angle

Since the third version (the APD3), natural AMPs have been classified into six life kingdoms: bacteria, archaea, protists, fungi, plants, and animals (current statistics in Figure 1; Wang et al., 2016). The increase of peptide count in each kingdom every 2 years is plotted in Figure 2. In this plot, archaea, protists, and fungi were all located in the valley due to small numbers in the APD (Figure 1). Also, the peptide counts from bacteria and plant kingdoms were below 400. Therefore, the only kingdom that showed a clear increase in peptide number at this scale is animal (72.8% in the APD). Further analysis of the count of animal AMPs every 2 years uncovered a linear correlation (slope 130.88 and R squared 2020: 0.9737). Plant AMPs increased linearly (slope 19.327 and R^2 : 0.8563) despite at a much slower pace.

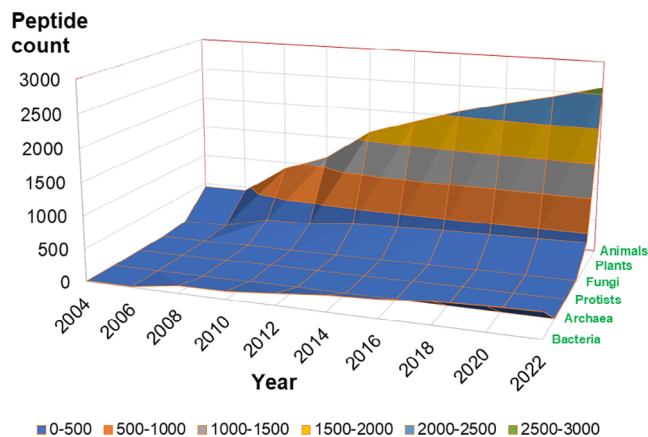


FIGURE 2 3D plot of the increase of antimicrobial peptides from bacteria, archaea, protists, fungi, plants, and animals over 20 years from 2003 to 2023 in the APD. Animal data constitute the colored mountain, while AMPs from bacteria, archaea, protists, fungi, and plants form the valley.

Likewise, bacterial AMPs grew in proportion to a year over 20 years (slope 22.54 and R^2 : 0.9788). Hence, natural AMPs, including AMPs from bacteria, plants, and animals, all increased linearly in the APD in the past 20 years. Such linear relationships explain in part why the amino acid signatures for AMPs from bacteria, amphibians, insects, and plants remained the same from 2008 to 2020 (Wang, Zietz, et al., 2022).

Figure 3 shows the dot plots of net charge or hydrophobic percentage (Pho) as a function of peptide length for all the peptides in the APD. In the case of net charge, the dots were mostly condensed below 50 amino acids and scattered between +30 to −12. Interestingly, short peptides had a low and narrower net charge range (Figure 3A). In contrast, they displayed a wide hydrophobic range. Peptides with a Pho greater than 80% were limited to those with less than 20 residues. The hydrophobic content then shrank to ~30% with an increase in peptide length (>100 aa; Figure 3B). Examples for AMPs with 0% hydrophobic residues are special peptides. SAAP fraction 3, a surfactant-associated anionic peptide, consists of a string of aspartic acids (Brogden et al., 1996), while shepherins comprised primarily glycine and histidine residues (Park et al., 2000). These are amino acid-rich peptides where at least one amino acid is greater than 25% in the sequence (Decker et al., 2022). Only two cyclic peptides, bacericin, and lugdunin, are entirely hydrophobic (100%; Niggemann et al., 2014; Zipperer et al., 2016). When the peptide length was increased from 0 to 50, there was a clear increase in net charge (Figure 3A) but decrease in hydrophobic percentage (Figure 3B). In the extreme of a low net charge, some peptides such as amphibian temporins (Conlon

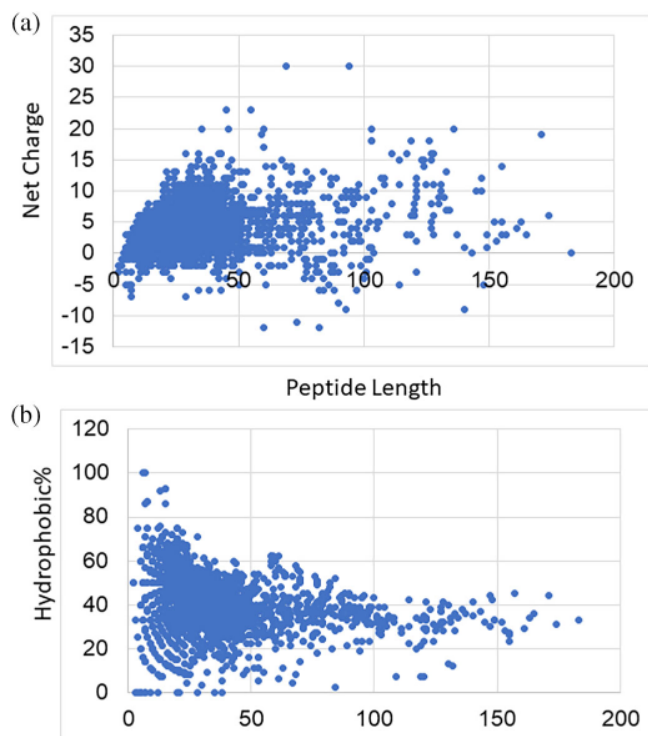


FIGURE 3 Scatter dot plot of net charge (A) and hydrophobic percentage (B) for all the antimicrobial peptides in the APD as a function of peptide length. The upper length limit (200 amino acids) was defined by the APD to include human antimicrobial proteins. This plot reveals the expansion in net charge and shrinkage of hydrophobic content with the increase in peptide length.

et al., 2009; Mangoni, 2006) tend to have a high hydrophobic percentage. These hydrophobic peptides, including the two extremes (bacteridin and lugdunin), are usually active against gram-positive bacteria such as *S. aureus* but not gram-negative bacteria such as *Escherichia coli*.

2.2 | Data re-annotation for antimicrobial development

2.2.1 | Systematic annotations of antimicrobial activity data

Initially, the APD included very limited activity data by considering the challenging nature in comparing the activity of AMPs conducted in different laboratories using different methods and conditions. So, AMPs were qualitatively labeled as antimicrobial (antibacterial, antiviral, antifungal, and anticancer) in the first version of the APD (Wang & Wang, 2004). With the increased research on microbiota that requires selective peptide antibiotics,

it became necessary to more completely annotate the activity of AMPs. This is because microbiota is so elegant that even one invading pathogen can tilt its balance. Hence, a peptide in principle can be tailored to correct the imbalance if it can perform a targeted elimination (Kim et al., 2019; Li et al., 2021). In 2021, the APD created a complete activity annotation system based on our previous definition for active and inactive peptides (Wang, Zietz, et al., 2022). This annotation system enables a more complete record of the reported activity data for each AMP. In this system, peptide activities were classified into three categories: active peptides (MIC < 100 μ M), inactive peptides (MIC > 100 μ M), and uncertain (activity greater than an MIC value less than 100 μ M) (summarized in Table 1). Active AMPs were annotated in the “Additional Information” using the abbreviated form of scientific names (e.g., *E. coli*, *S. aureus*, and *C. albicans*). When *E. coli* was entered into “Additional Information” followed by “search, we obtained 1834 AMPs active against *E. coli*. In contrast, peptides that did not kill a specific microbe (MIC > 100 μ M) are annotated in a different format so that this set of inactive peptides could be searched (e.g., *E.coli*, *S.aureus*, and *C.albicans*, no space). For example, we obtained 218 peptides when we searched the database using *E.coli* (Table 1). Those AMPs with uncertain activity (e.g., 100 μ M > MIC > any value) were annotated in a third format (e.g., *E.coli*, *S.aureus*, and *C.albicans*). In this way, all the three types of peptide activities could be searched in the APD via the “Additional information” field. Table 1 tabulated the three types of peptide activities annotated to date against *E. coli*, *S. aureus*, and *C. albicans*. Any other microbes such as *P. aeruginosa* can be searched in the same manner as long as its activity has been entered into the APD.

There are more data for *E. coli*, *S. aureus*, and *C. albicans* in the APD due to wide use of these strains in antimicrobial screening. Hence, we first compared the amino acid signature of active and inactive AMPs against *E. coli*. In the case of helical peptides, those active against *E. coli* had clearly higher contents of lysine and arginine, although glycine and valine were only slightly higher. In contrast, those helical AMPs with very weak or no activity against *E. coli* possessed higher contents of alanine, phenylalanine, isoleucine, leucine, and serine (summarized in Table 2). Clearly, cationic amino acids play an important role in killing *E. coli*. We then compared AMPs active and inactive against *S. aureus*. There appeared an opposite requirement for *S. aureus*. Active helical AMPs showed higher contents of phenylalanine, glycine, and leucine, while those *S. aureus*-inactive helical AMPs were higher in lysine and alanine (Table 2). These amino acids were derived from the amino acid plots (Figure S1).

TABLE 1 A full antibacterial activity annotation system in the APD.

Activity type	Active	Inactive	Uncertain
Definition	MIC < 100 μ M	MIC > 100 μ M	For example, MIC > 16 μ M but not > 100 μ M
Microbe annotation and search format ^a	<i>E. coli</i> , or <i>S. aureus</i>	<i>E.coli</i> , or <i>S.aureus</i>	<i>E-coli</i> , or <i>S-aureus</i>
<i>Escherichia coli</i>	1834	218	62
<i>Staphylococcus aureus</i>	1744	161	88
<i>Candida albicans</i>	806	163	31

^aData obtained from the “Additional information” field of the database (<https://aps.unmc.edu>) as of July 2023 using the defined search format. Although only three commonly used microbes were utilized here to illustrate the status of data annotations, any microbe species can be searched in the same manner.

TABLE 2 Amino acid use in desired and undesired antimicrobial peptide groups^a.

Group		Desired	Undesired	Data source
1	Peptide target	<i>Antibacterial</i>	<i>Non-antibacterial</i>	
	<i>E. coli</i>	K, R, V	A, F, I, L, S	Figure S1A
	<i>S. aureus</i>	F, G, L, P	A, K, Y	Figure S1B
2	Peptide scaffold	<i>Non-hemolytic</i>	<i>Hemolytic</i>	
	UCLL	V, W, (P), Q, (N), K, R	I, L, (F), A, G, S, (H)	Figure S2A
	UCSS	M, (W), G, N, D, H, R	I, (V), (L), F, C, (A), P, T, K	Figure S2B
3	Peptide scaffold	<i>Salt-insensitive</i>	<i>Salt-sensitive</i>	
	UCLL	F, G, P, H, R	A, S, Q, N, E, D, K	Figure S3A
	UCSS	V, L, T, S, N, H, R	I, A, P, Q, E, D, K	Figure S3B

^aSignificantly different in bold, and smaller changes in parenthesis. Data obtained from the APD (<https://aps.unmc.edu>) in July 2023. Figure S can be found in the Supporting Information section of this article.

These plots support the contention that hydrophobic amino acids are critical in inhibiting gram-positive MRSA. This database finding is consistent with the observation from structure–activity relationship (SAR) studies (Li et al., 2006; Wang et al., 2018). Interestingly, helical AMPs active against *C. albicans* resembled those against *E. coli* since they also had higher contents of lysine and arginine. Such a feature differed from those *C. albicans*-inactive peptides with higher percentages of hydrophobic leucine and alanine.

We then compared active and inactive AMPs with beta structures. A consensus picture emerged irrespective of the pathogen type: *E. coli*, *S. aureus*, or *C. albicans*. Such a similarity for β -sheet AMPs may be attributed to the broad-spectrum activity of defensin-like peptides. The active groups with a beta structure were abundant in cysteine, glycine, and arginine, three critical amino acids in defensins (10.5% R for 379 defensins). However, the three dominant amino acids in the inactive groups were cysteine, glycine, and threonine (Table 2). The low content of arginine (4.3% or less) might be one of the major reasons for the lack of activity of the inactive group since cysteine and glycine are the two common amino acids required for β -sheet structure formation.

2.2.2 | Hemolytic and nonhemolytic peptides

The development of AMPs into a new generation of antibiotics requires that the peptide do a targeted pathogen elimination without harming the host (Bobde et al., 2021; Hancock et al., 2021). Hemolysis is frequently utilized to gauge peptide cytotoxicity. Hemolytic peptides were annotated in the first version of the APD (Wang & Wang, 2004). Our statistical analysis at that time revealed a higher hydrophobic content for hemolytic peptides than other antimicrobial groups. To better understand the differences between hemolytic and nonhemolytic AMPs at the amino acid level, the APD is in the process of re-annotating the peptides by defining criteria for “hemolytic” and “non-hemolytic” as well. Hemolytic peptides show a clear dose-dependent hemolysis with a 50% hemolytic concentration (HC₅₀) less than 100 μ M, while nonhemolytic peptides show no sign of hemolysis at least at 10-fold of MIC and have a HC₅₀ value greater than 100 μ M. In many cases, however, a detailed hemolytic curve is not provided in published papers and only HC₅₀ may be available for this estimation. Like antimicrobial activity, the APD is fully aware of the challenge in the

heterogeneity of hemolytic data, which can be influenced by numerous factors, including protocols, blood cell types (e.g., human and nonhuman), container (Eppendorf tubes or microplate/type), blood and peptide freshness, cell concentration, and incubation time. Preferred amino acids for hemolytic and nonhemolytic AMPs currently annotated in the APD are provided in Table 2 as well. These amino acids were derived from the amino acid plots of these two groups (Figure S2). It is evident that hemolytic peptides from the linear class (UCLL) possessed higher contents of hydrophobic amino acids, including isoleucine, leucine, phenylalanine, and alanine. In contrast, the non-hemolytic group showed higher contents of lysine, arginine, tryptophan, valine, glutamine, and asparagine. We also compared the two sidechain-linked AMP (UCSS) groups with hemolytic and non-hemolytic activities. Again, there were more hydrophobic residues I and F in the hemolytic group than the those in the nonhemolytic group. The nonhemolytic group had a substantial percentage of glycine, asparagine, and arginine. It appeared that arginine was higher in both nonhemolytic groups irrespective of the structural scaffolds. These comparisons may inspire the design of nonhemolytic peptides.

2.2.3 | Consensus amino acids in salt-sensitive and insensitive antimicrobial peptides

It is known that some AMPs could lose antimicrobial activity in the presence of physiological salts (Huang et al., 2007; Krishnakumari et al., 2013). There are some studies in the literature aiming at designing AMPs resistant to salts (Chu et al., 2020; Dou et al., 2017; Mishra et al., 2019). However, a global picture for salt resistance of AMPs was lacking. The APD started to pay attention to this information before 2018, especially after the successful extension of our *in silico* filters to *in vitro* filters (Mishra et al., 2019). The *in vitro* filters include salt tolerance, pH sensitivity, and serum binding, which are known hurdles that might cause a drop or even loss of peptide activity. While data are accumulating, we documented herein the amino acid differences between the salt-sensitive and insensitive groups of AMPs currently annotated in the APD. Salt-sensitive AMPs changed their MIC value four-fold or more in the presence of 150 mM NaCl, while the MIC values of salt-insensitive peptides remained constant (not more than two-fold change in MIC). We narrowed down our database search to AMPs less than 50 amino acids and further split them into two structural classes: UCLL and UCSS based on our universal peptide classification (Wang, 2022). The results are

included in Table 2. The UCLL class consisted of linear AMPs (22 salt-insensitive vs. 9 salt-sensitive), which might form amphipathic helices upon interaction with bacterial membranes. Typical examples in the UCSS class are defensins (20 salt-sensitive vs. 6 salt-insensitive). In the case of the linear group, both arginine and histidine were higher in the salt-resistant group (i.e., salt-insensitive in Table 2) than in the salt-sensitive group. Likewise, these two amino acids were also more abundant in the salt-resistant sidechain-linked group than the salt-sensitive group. In addition, the salt-sensitive AMPs shared a consensus set of abundant amino acids (A, Q, E, D, and K). These amino acids were derived from the amino acid signature plots for the respective groups (Figure S3). It appears that there was a consensus in arginine preference irrespective of the peptide structural class. Such a discovery may inspire the design of salt-resistant AMPs.

2.3 | New light on natural antimicrobial peptides by comparison with globular proteins and transmembrane proteins

The first version of the APD enabled statistical analysis of AMPs. When the percentages of the 20 amino acids with different antimicrobial activities were plotted, it became clear that some amino acids (e.g., cationic K and R) are more abundant than other amino acids (e.g., acidic D and E). This explains why most of the AMPs in the APD are cationic (94.4% with net charge ≥ 0). However, we have never compared the amino acid composition of natural AMPs with that of globular or transmembrane proteins. To facilitate this comparison, we made a new addition to the APD by manually annotating relevant peptide entries with a searchable indicator “natural AMPs,” which includes all peptides from the six life kingdoms (Figure 1; Wang et al., 2016). This led to a total of 3090 natural AMPs when we conducted this search in the name field using “natural AMPs.” This list did not include 208 natural peptides without activity data (annotated as DXWZ), which were collected into the APD prior to the definition of data registration criteria. We then compared the amino acid compositions of natural AMPs with globular and transmembrane proteins (not shown; Gromiha et al., 2005). Lysine, glycine, and cysteine of natural AMPs on average were higher than those in either globular or transmembrane proteins (which are similar). These three residues resulted from a combination of dominant amino acids in α -helical (K and G) and β -sheet AMPs (G and C) in this database (Mishra & Wang, 2012a; Wang & Wang, 2004). To gain additional

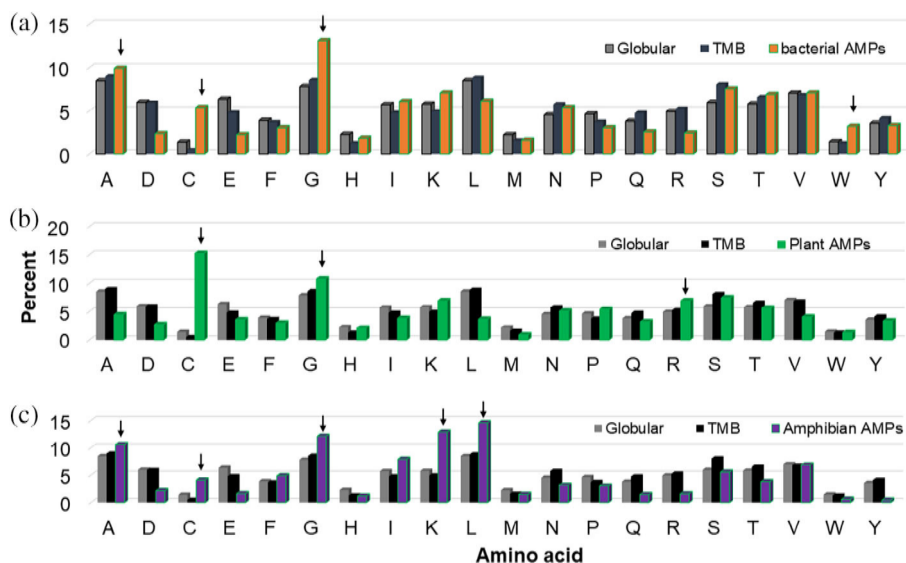


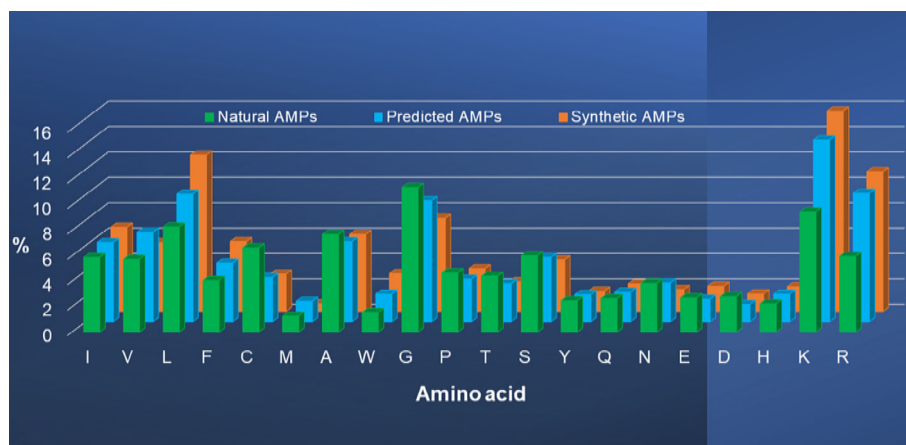
FIGURE 4 Comparison of amino acids in globular proteins (gray), transmembrane proteins (TMB, black), and antimicrobial peptides from (A) bacteria (373 entries, gold), (B) plants (281, green), and (C) amphibians (1067, purple) from the APD. Abundant amino acids of AMPs are pointed by arrows. This plot indicates that AMPs were rather different from both globular and transmembrane proteins, which have comparable amino acid compositions.

understanding, we also compared the amino acid composition of globular proteins with those of AMPs from bacteria, plants, and amphibians. In the case of bacterial AMPs (Figure 4A), cysteine and glycine were clearly higher, while leucine, arginine, aspartic, and glutamic acids were much lower. Cysteine in plant AMPs was extremely high due to the fact that these peptides frequently form disulfide bonds (Figure 4B). In addition, lysine, arginine, and glycine were slightly higher while aspartic and glutamic acids were lower. In the case of amphibian AMPs, residues alanine, glycine, leucine, and lysine became dominant (Figure 4C). These are the known frequently occurring amino acids we initially discovered in the APD2 paper (Wang et al., 2009). Beyond these amino acids, threonine was also clearly lower, but isoleucine was higher. In addition, tyrosine, arginine, glutamine, glutamic, and aspartic acids were extremely low compared to those in globular proteins. It appeared that these amino acids were less preferred in nature's design of amphibian AMPs. Thus, those more abundant amino acid signatures differed: A, C, G, and W for bacterial AMPs, C, G, and R for plant AMPs, and A, C, G, K, and L for amphibians (Figure 4 arrows). While C and G were higher in all the three AMP groups than that in globular proteins, D and E were consistently lower. Also, the use of tryptophan differed: more in AMPs from bacteria than from amphibian. Our new results obtained herein provided additional insight into AMP design in different life kingdoms or classes since amino acid compositions play an important role in determining peptide structure and activity spectrum (Maasch et al., 2023; Mishra & Wang, 2012a; Mishra & Wang, 2012b; Wang, Vaisman, & van Hoek, 2022).

2.4 | A new peptide group is created in the APD for machine learning-predicted antimicrobial peptides: Amino acid landscapes of natural, predicted, and synthetic AMPs

Since the first single-label prediction of AMPs by machine learning in 2007 (Lata et al., 2007), the APD has also been utilized for multi-label prediction (Xiao et al., 2013) based on a variety of peptide functions/activities annotated therein (Figure 1). Machine learning prediction is becoming popular these days (Maasch et al., 2023; Wang, Vaisman, & van Hoek, 2022; Xiao et al., 2013). It opens a new avenue for AMP discovery. Even encrypted antimicrobial peptides from extinct paleoproteomes were predicted by this technology in 2023 (Maasch et al., 2023). Hence, the APD cannot ignore new peptides from these predictions, especially those predicted from genomes. As we do not have enough knowledge to judge to what extent these predicted sequences correspond to natural AMPs, we decided to create a new group "predicted" for these peptides if their antimicrobial activity had been proved experimentally (i.e., MIC <100 μ M). Note that the predicted AMPs in the APD also include those sequences predicted by other methods, such as sequence alignment. For instance, FALL-39, initially included in the LL-37 entry (AP00310), now occupies an independent entry (AP03566) as a classic member of the predicted family based on the highly conserved precursor sequences of cathelicidins (Agerberth et al., 1995). FALL-39 differs from LL-37 by two residues and ALL-38 by only one residue. Both LL-37 (other tissues such as blood neutrophils and skin) and

FIGURE 5 Amino acid landscapes of 3090 natural AMPs (green), 190 predicted AMPs (blue), and 121 synthetic peptides (gold) registered in the APD as of April 2023. These peptide classes can be searched in the name field of the APD by entering “natural AMPs,” “predicted,” and “synthetic,” respectively. This plot indicates that natural AMPs and synthetic peptides in the APD set the lower and upper parameter bounds for shallow and deep learning predictions.



ALL-38 (from human reproductive system) are isolated AMPs (Gudmundsson et al., 1996; Sørensen et al., 2003). Likewise, one of the two predicted forms of murine cathelicidin differs from the isolated one by only one residue (Pestonjamas et al., 2001). These results indicate that predictions based on sequence alignment can reach a remarkably high accuracy for cathelicidins (95%–97%). At the moment, we do not have data for a similar accuracy comparison of machine learning-predicted peptides with natural AMPs. By the time this manuscript was completed, the APD has registered 194 peptides predicted by different methods, mostly AI predicted.

We then asked how the predicted peptides were deviated from natural AMPs. Figure 5 compares the averaged amino acid composition profile of natural AMPs with that of predicted peptides in the APD. The overall trend was remarkably similar, indicating heavy dependence of the positive training data of natural AMPs we obtained in the past 20 years. However, residues K and R were much more abundant in predicted AMPs (blue), while residues D and E were lower. Meanwhile, hydrophobic residues F, I, L, and V were higher, but A and C were lower than those in natural AMPs. Such a combined tilt toward larger hydrophobic and basic amino acids by machine learning might have increased the success of predicted AMPs. A lower C implied an emphasis of the current predictions on linear sequences.

Synthetic peptides are human-made AMPs to understand structure–activity relationships or to improve antimicrobial therapeutic potential. They form a separate group in the APD from the birth of the database in 2003. Due to our previous focus on natural AMPs, not all synthetic peptides have been collected yet. Among the 121 synthetic AMPs in the APD, many were designed using amino acids known to be important for peptide activity. In addition, some synthetic AMPs have been optimized for high activity and low toxicity. Library screen and SAR studies are two common *in vitro*

evolution methods to make the candidate peptide more potent and less toxic (Blondelle et al., 1995; Deslouches et al., 2005; Jacob & Zasloff, 1994; Lakshmaiah Narayana et al., 2020; Loose et al., 2006; Mishra et al., 2019). For instance, combi-1 and combi-2 are among the best peptides discovered from combinatorial libraries (Blondelle et al., 1995), while WLBU2 is obtained from a *de novo* design based on valine and arginine (Deslouches et al., 2005). Likewise, the two best peptides D28 and D51 from the grammar approach are found in the APD (Loose et al., 2006). Horine and verine are optimal candidates identified from a family of database-designed peptides (Lakshmaiah Narayana et al., 2020). We were curious how the plot for these synthetic peptides looked like. Remarkably, residues I, L, F, W, K, and R were highest, while residues A, G, N, S, and T were lowest (gold in Figure 5) compared to both predicted and natural AMPs. Hence, a select set of basic and hydrophobic amino acids was elevated to the highest level in the synthetic peptides in the current APD. Again, residue C was lowest in synthetic peptides, implying a preference of linear peptides. M was lowest to avoid loss of activity from auto-oxidation.

It is interesting that the amino acid composition of the predicted AMPs on average lay between those of natural and synthetic AMPs (Figure 5). Since both natural and synthetic peptides in the APD had usually been included in machine learning predictions, they might have set the lower and upper parameter bounds for such predictions. The picture is clear that natural AMPs laid the foundation for both synthetic and predicted peptides, which had higher cationic (K and R) and hydrophobic amino acids (I, L, F, and W) for potency. This was achieved by decreasing small amino acids (A, G, and S). W is low in natural AMPs (Figure 5) but has been widely incorporated into synthetic AMPs, especially short ones (Blondelle et al., 1995; Chan et al., 2006; Cherkasov et al., 2009; Lakshmaiah Narayana et al., 2020). We

conclude that the classic amphipathic concept is deeply rooted in the prediction and design of AMPs.

3 | CONCLUSION AND FUTURE DIRECTIONS

The APD is a founding database for the AMP field first constructed based on the LAMP software bundle, including Red Hat Linux operating system using the freeware Apache web server, MySQL relational database management system, and PHP script language (Wang & Wang, 2004). It built numerous database features for the first time, including unified classifications of AMPs in numerous ways (Figure 1) (Wang, Zietz, et al., 2022). Consequently, the APD is widely utilized in research and education (Wang et al., 2016). To facilitate education, the APD has put information in order by establishing multiple educational web pages, including peptide discovery timeline, glossary, nomenclature, classification, structural determination, tools, and web links. As an educational event, a hybrid symposium entitled “*Antimicrobial Peptides: Yesterday, Today and Tomorrow*” will be held in Omaha, Nebraska, USA on October 6, 2023.

To further advance AMP research, this study has unified the AMPs from bacteria, archaea, protists, fungi, plants, and animals under a common umbrella “natural AMPs” (searchable). Our new annotation facilitated a comparison of natural AMPs with globular proteins, providing novel insight into the amino acid use in natural AMPs in general and in bacteria, plants, and amphibian AMPs in particular. Different from globular proteins (usually > 100 amino acids), which automatically fold under physiological conditions, many AMPs (usually < 50 amino acids) are forced to a structural scaffold via disulfide bond formation. A majority of linear AMPs are disordered in aqueous solution and only fold into an amphipathic helical structure upon binding to membranes. The amphipathic concept has been dominating the field and widely utilized as a general principle to design and improve AMPs for decades. However, some linear peptides (e.g., proline-rich AMPs) may also adopt a non-helical structure after association with ribosomes (Gagnon et al., 2016).

This study also expanded the scope of AMPs in the APD by creating a new group predicted by machine learning followed by antimicrobial validation. However, peptides predicted by other methods such as sequence alignment are also included in this group. The similarity in amino acid landscapes (Figure 5) underscores that the natural AMPs accumulated in the APD over 20 years have laid the foundation for machine learning prediction. It is interesting to note that, man-made peptides, be they

predicted or designed, are deviated from the baseline set by natural AMPs due to higher cationic and hydrophobic amino acid incorporation to enhance antimicrobial potency. Synthetic peptides in the APD, despite a small set, have been designed and optimized to varying degrees, thereby setting the parameter upper bound for peptide prediction.

Our data re-annotation has generated new knowledge for AMPs. Cationic amino acids are important for antimicrobial activity of helical peptides against gram-negative *E. coli*, while leucine, phenylalanine, and glycine are more abundant in helical peptides active against gram-positive *S. aureus*, including MRSA. This is in line with our previous discovery that high hydrophobicity is important for anti-MRSA peptides (Mishra et al., 2019). Along this line, there are extremes where some anti-MRSA AMPs are entirely hydrophobic, indicating charged amino acids are not a must for killing MRSA. These hydrophobic extremes, together with hydrophilic extremes (e.g., poly-aspartic AMPs), provide exceptions to the amphipathic concept. Arginine plays a critical role in β -sheet AMPs broadly active against *E. coli*, *S. aureus*, and *C. albicans*. Among the multiple arginines in a central peptide of human LL-37, only the sidechain of R23 can more effectively interact with anionic lipid headgroup (Wang, 2007). Arginine is important for antiviral activity as well (Wang et al., 2010; Wang & Wang, 2004). This study suggests that arginine might be useful in conferring salt resistance to AMPs (Table 2). In certain cases, arginine may also make the peptide more hemolytic to human erythrocytes as it is more hydrophobic than lysine (Mishra, Lushnikova, et al., 2017). While the first version of the APD uncovered higher hydrophobic contents for hemolytic AMPs (Wang & Wang, 2004), our data re-annotation here reveals that high hydrophobicity for hemolytic peptides results from amino acids F, I, and L. Hence, this study has enriched our knowledge on the design of novel AMPs.

Before closing, it is also useful to point out some possible directions for AMP discovery. In the APD3 paper, we projected the number of natural AMPs at least in the million range if each of the 1.3 million named species can produce at least one such peptide (Wang et al., 2016). After re-annotation, the APD contained 3090 “natural AMPs” (with activity data) out of a total of 3623 peptides as of April 2023. Therefore, the AMP field has a long way to go to reach 1 million natural AMPs. Future studies may explore new AMPs in those under-represented life kingdoms, including bacteria, archaea, fungi, protists, and plants (Figure 2 valley). Indeed, scientists are discovering new AMPs from bacteria in a variety of environments, be they deep in the sea, buried in soil, or hidden in animal guts (Hansen et al., 2020; Hatzioanou

et al., 2017; Ling et al., 2015; Ma et al., 2022). We anticipate the application of various methods ranging from the classic isolation and purification to machine learning in future peptide discovery. With the discovery of additional AMP members from those under-represented kingdoms and further development of artificial intelligence algorithms, we anticipate a great advance in peptide prediction and design toward future antibiotics, food preservatives, and “green” pesticides. To restore imbalanced microbiota, species-specific AMPs are of top interest. This is a challenging task since it requires the use of many more relevant bacterial strains in antimicrobial assays to define the activity spectrum of a particular candidate before use. Our new data annotation scheme (Table 1) has reached a more complete record of antimicrobial activity data. This will facilitate activity spectrum matching for potential therapeutics. Based on the current knowledge, bacteriocins (i.e., bacterial AMPs) constitute excellent candidates to correct the imbalance of microbiota because of their frequently observed species-specific antimicrobial activity. Therefore, we anticipate that future AMP research will pay more attention to bacteriocins with a variety of structural scaffolds. Indeed, several AMPs currently in clinic use or for food preservation are made by microbes (Mishra, Reiling, et al., 2017).

In summary, the APD has been evolving for 20 years with continued updates and refinement. This study has further expanded the landscape of AMPs registered in the APD and improved its capability for antimicrobial development. With further developments in the future, the APD will continue to serve the antimicrobial research community.

AUTHOR CONTRIBUTIONS

Guangshun Wang: Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (equal); project administration (equal); supervision (equal); validation (equal); writing – original draft (equal).

ACKNOWLEDGMENTS

The author appreciates previous grants as well as the current National Institutes of Health grants GM138552 and AI175209. The author is grateful to the department of Pathology and Microbiology as well as UNMC IT for years of support. I thank my first graduate student Zhe Wang who did a wonderful job in creating the original APD. We appreciate all authors for peptide deposition and all users for additions, corrections, and suggestions.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

ORCID

Guangshun Wang  <https://orcid.org/0000-0002-4841-7927>

REFERENCES

- Agerberth B, Gunne H, Odeberg J, Kogner P, Boman HG, Gudmundsson GH. FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. *Proc Natl Acad Sci USA*. 1995;92:195–9.
- Bishop BM, Juba ML, Russo PS, Devine M, Barksdale SM, Scott S, et al. Discovery of novel antimicrobial peptides from *Varanus komodoensis* (komodo dragon) by large-scale analyses and De novo-assisted sequencing using electron-transfer dissociation mass spectrometry. *J Proteome Res*. 2017;16:1470–82.
- Blondelle SE, Takahashi E, Dinh KT, Houghten RA. The antimicrobial activity of hexapeptides derived from synthetic combinatorial libraries. *J Appl Bacteriol*. 1995;78(1):39–46.
- Bobde SS, Alsaab FM, Wang G, van Hoek ML. *Ab initio* designed antimicrobial peptides against gram-negative bacteria. *Front Microbiol*. 2021;12:715246.
- Brogden KA, de Lucca AJ, Bland J, Elliott S. Isolation of an ovine pulmonary surfactant-associated anionic peptide bactericidal for *Pasteurella haemolytica*. *Proc Natl Acad Sci USA*. 1996;93:412–6.
- Candille SI, Kaelin CB, Cattanaach BM, Yu B, Thompson DA, Nix MA, et al. A-defensin mutation causes black coat color in domestic dogs. *Science*. 2007;318:1418–23.
- Chan DI, Prenner EJ, Vogel HJ. Tryptophan- and arginine-rich antimicrobial peptides: structures and mechanisms of action. *Biochim Biophys Acta*. 2006;1758:1184–202.
- Cherkasov A, Hilpert K, Jenssen H, Fjell CD, Waldbrook M, Mullaly SC, et al. Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. *ACS Chem Biol*. 2009;4:65–74.
- Chu HL, Chih YH, Peng KL, Wu CL, Yu HY, Cheng D, et al. Antimicrobial peptides with enhanced salt resistance and antiendotoxin properties. *Int J Mol Sci*. 2020;21(18):6810.
- Conlon JM, Al-Ghafari N, Coquet L, Leprince J, Jouenne T, Vaudry H, et al. Evidence from peptidomic analysis of skin secretions that the red-legged frogs, *Rana aurora draytonii* and *Rana aurora aurora*, are distinct species. *Peptides*. 2006;27:1305–12.
- Conlon JM, Kolodziejek J, Nowotny N. Antimicrobial peptides from the skins of north American frogs. *Biochim Biophys Acta*. 2009;1788:1556–63.
- de San NN, Asokan A, Rosa RD, Voisin SN, Travers MA, Rocha G, et al. Functional diversification of oyster big defensins generates antimicrobial specificity and synergy against members of the microbiota. *Mar Drugs*. 2022;20:745.
- Decker AP, Mechesso AF, Wang G. Expanding the landscape of amino acid-rich antimicrobial peptides: definition, deployment in nature, implications for peptide design and therapeutic potential. *Int J Mol Sci*. 2022;23:12874.
- Deslouches B, Phadke SM, Lazarevic V, Cascio M, Islam K, Montelaro RC, et al. De novo generation of cationic antimicrobial peptides: influence of length and tryptophan substitution on antimicrobial activity. *Antimicrob Agents Chemother*. 2005;49:316–22.

- Dou X, Zhu X, Wang J, Dong N, Shan A. Novel design of heptad amphiphiles to enhance cell selectivity, salt resistance, antibiofilm properties and their membrane-disruptive mechanism. *J Med Chem*. 2017;60:2257–70.
- Dubos RJ. Studies on a bactericidal agent extracted from a soil bacillus: I. Preparation of the agent. Its activity in vitro. *J Exp Med*. 1939;70:1–10.
- Eliopoulos GM, Willey S, Reiszner E, Spitzer PG, Caputo G, Moellering RC Jr. In vitro and in vivo activity of LY 146032, a new cyclic lipopeptide antibiotic. *Antimicrob Agents Chemother*. 1986;30:532–5.
- Fleming A. President's address. *Proc R Soc Med*. 1932;26:71–84.
- Fjell CD, Hiss JA, Hancock RE, Schneider G. Designing antimicrobial peptides: form follows function. *Nat Rev Drug Discov*. 2011;11:37–51.
- Gagnon MG, Roy RN, Lomakin IB, Florin T, Mankin AS, Steitz TA. Structures of proline-rich peptides bound to the ribosome reveal a common mechanism of protein synthesis inhibition. *Nucleic Acids Res*. 2016;44:2439–50.
- Gallo RL, Hooper LV. Epithelial antimicrobial defence of the skin and intestine. *Nat Rev Immunol*. 2012;12:503–16.
- Ganz T, Selsted ME, Szklarek D, Harwig SS, Daher K, Bainton DF, et al. Defensins. Natural peptide antibiotics of human neutrophils. *J Clin Invest*. 1985;76:1427–35.
- Gennaro R, Skerlavaj B, Romeo D. Purification, composition, and activity of two bactenecins, antibacterial peptides of bovine neutrophils. *Infect Immun*. 1989;57:3142–6.
- Gesell J, Zasloff M, Opella SJ. Two-dimensional 1H NMR experiments show that the 23-residue magainin antibiotic peptide is an alpha-helix in dodecylphosphocholine micelles, sodium dodecylsulfate micelles, and trifluoroethanol/water solution. *J Biomol NMR*. 1997;9:127–35.
- Gromiha MM, Ahmad S, Suwa M. TMBETA-NET: discrimination and prediction of membrane spanning beta-strands in outer membrane proteins. *Nucleic Acids Res*. 2005;33:W164–7.
- Gudmundsson GH, Agerberth B, Odeberg J, Bergman T, Olsson B, Salcedo R. The human gene FALL39 and processing of the cathelin precursor to the antibacterial peptide LL-37 in granulocytes. *Eur J Biochem*. 1996;238:325–32.
- Hancock REW, Alford MA, Haney EF. Antibiofilm activity of host defence peptides: complexity provides opportunities. *Nat Rev Microbiol*. 2021;19:786–97.
- Hansen IKØ, Lövdahl T, Simonovic D, Hansen KØ, Andersen AJC, Devold H, et al. Antimicrobial activity of small synthetic peptides based on the marine peptide Turgencin a: prediction of antimicrobial peptide sequences in a natural peptide and strategy for optimization of potency. *Int J Mol Sci*. 2020;21:5460.
- Hanson MA, Grollmus L, Lemaitre B. Ecology-relevant bacteria drive the evolution of host antimicrobial peptides in *Drosophila*. *Science*. 2023;381:eadg5725.
- Hanson MA, Lemaitre B. Antimicrobial peptides do not directly contribute to aging in *Drosophila*, but improve lifespan by preventing dysbiosis. *Dis Model Mech*. 2023;16:dmm049965.
- Hatzioanou D, Gherghisan-Filip C, Saalbach G, Horn N, Wegmann U, Duncan SH, et al. Discovery of a novel lantibiotic nisin O from *Blautia obeum* A2-162, isolated from the human gastrointestinal tract. *Microbiology*. 2017;163:1292–305.
- Huang LC, Redfern RL, Narayanan S, Reins RY, McDermott AM. In vitro activity of human beta-defensin 2 against *Pseudomonas aeruginosa* in the presence of tear fluid. *Antimicrob Agents Chemother*. 2007;51:3853–60.
- Human genome project completed in 2003: <https://www.genome.gov/human-genome-project>
- Jacob L, Zasloff M. Potential therapeutic applications of magainins and other antimicrobial agents of animal origin. *Ciba Found Symp*. 1994;186:197–216.
- Jeener J, Alewaeters G. "Pulse pair technique in high resolution NMR" a reprint of the historical 1971 lecture notes on two-dimensional spectroscopy. *Prog Nucl Magn Reson Spectrosc*. 2016;94-95:75–80.
- Kim SG, Becattini S, Moody TU, Shliaha PV, Littmann ER, Seok R, et al. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant *Enterococcus*. *Nature*. 2019;572:665–9.
- Krishnakumari V, Packiyannathan KK, Nagaraj R. Human-β-defensins-1-3 and analogs do not require proton motive force for antibacterial activity against *Escherichia coli*. *FEMS Microbiol Lett*. 2013;348(1):52–7.
- Lai R. Combined peptidomics and genomics approach to the isolation of amphibian antimicrobial peptides. *Methods Mol Biol*. 2010;615:177–90.
- Lakshmaiah Narayana J, Mishra B, Lushnikova T, Wu Q, Chhonker YS, Zhang Y, et al. Two distinct amphipathic peptide antibiotics with systemic efficacy. *Proc Natl Acad Sci USA*. 2020;117:19446–54.
- Lata S, Sharma BK, Raghava GP. Analysis and prediction of antibacterial peptides. *BMC Bioinformatics*. 2007;8:263.
- Lazzaro BP, Zasloff M, Roff J. Antimicrobial peptides: application informed by evolution. *Science*. 2020;368:eaau5480.
- Lee IH, Lee YS, Kim CH, Kim CR, Hong T, Menzel L, et al. Dicynthaurin: an antimicrobial peptide from hemocytes of the solitary tunicate *Halocynthia aurantium*. *Biochim Biophys Acta*. 2001;1527:141–8.
- Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette spätzle/toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell*. 1996;86:973–83.
- Li J, Chen J, Yang G, Tao L. Sublancin protects against methicillin-resistant *Staphylococcus aureus* infection by the combined modulation of innate immune response and microbiota. *Peptides*. 2021;141:170533.
- Li X, Li Y, Han H, Miller DW, Wang G. Solution structures of human LL-37 fragments and NMR-based identification of a minimal membrane-targeting antimicrobial and anticancer region. *J Am Chem Soc*. 2006;128:5776–85.
- Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, et al. A new antibiotic kills pathogens without detectable resistance. *Nature*. 2015;517:455–9.
- Loose C, Jensen K, Rigoutsos I, Stephanopoulos G. A linguistic model for the rational design of antimicrobial peptides. *Nature*. 2006;443:867–9.
- Ludtke SJ, He K, Heller WT, Harroun TA, Yang L, Huang HW. Membrane pores induced by magainin. *Biochemistry*. 1996;35:13723–8.
- Ma Y, Guo Z, Xia B, Zhang Y, Liu X, Yu Y, et al. Identification of antimicrobial peptides from the human gut microbiome using deep learning. *Nat Biotechnol*. 2022;40:921–31.
- Maasch JRMA, Torres MDT, Melo MCR, de la Fuente-Nunez C. Molecular de-extinction of ancient antimicrobial peptides

- enabled by machine learning. *Cell Host Microbe*. 2023;31:1260–1274.e6.
- Mangoni ML. Temporins, anti-infective peptides with expanding properties. *Cell Mol Life Sci*. 2006;63:1060–9.
- Mansour SC, Pena OM, Hancock RE. Host defense peptides: frontline immunomodulators. *Trends Immunol*. 2014;35:443–50.
- Menousek J, Mishra B, Hanke ML, Heim CE, Kielian T, Wang G. Database screening and in vivo efficacy of antimicrobial peptides against methicillin-resistant *Staphylococcus aureus* USA300. *Int J Antimicrob Agents*. 2012;39:402–6.
- Mishra B, Lakshmaiah Narayana J, Lushnikova T, Wang X, Wang G. Low cationicity is important for systemic in vivo efficacy of database-derived peptides against drug-resistant Gram-positive pathogens. *Proc Natl Acad Sci USA*. 2019;116:13517–22.
- Mishra B, Lushnikova T, Golla RM, Wang X, Wang G. Design and surface immobilization of short anti-biofilm peptides. *Acta Biomater*. 2017;49:316–28.
- Mishra B, Reiling S, Zarena D, Wang G. Host defense antimicrobial peptides as antibiotics: design and application strategies. *Curr Opin Chem Biol*. 2017;38:87–96.
- Mishra B, Wang G. Ab initio design of potent anti-MRSA peptides based on database filtering technology. *J Am Chem Soc*. 2012a;134:12426–9.
- Mishra B, Wang G. The importance of amino acid composition in natural AMPs: an evolutionary, structural, and functional perspective. *Front Immunol*. 2012b;3:221.
- Mookherjee N, Lippert DN, Hamill P, Falsafi R, Nijnik A, Kindrachuk J, et al. Intracellular receptor for human host defense peptide LL-37 in monocytes. *J Immunol*. 2009;183:2688–96.
- Mookherjee N, Rehaume LM, Hancock RE. Cathelicidins and functional analogues as antiseptic molecules. *Expert Opin Ther Targets*. 2007;11(8):993–1004.
- Moyer TB, Allen JL, Shaw LN, Hicks LM. Multiple classes of antimicrobial peptides in *Amaranthus tricolor* revealed by prediction, proteomics, and mass spectrometric characterization. *J Nat Prod*. 2021;84:444–52.
- Niggemann J, Bozko P, Bruns N, Wodtke A, Gieseler MT, Thomas K, et al. Baceridin, a cyclic hexapeptide from an epiphytic bacillus strain, inhibits the proteasome. *Chembiochem*. 2014;15:1021–9.
- Oppenheim FG, Xu T, McMillian FM, Levitz SM, Diamond RD, Offner GD, et al. Histatins, a novel family of histidine-rich proteins in human parotid secretion. Isolation, characterization, primary structure, and fungistatic effects on *Candida albicans*. *J Biol Chem*. 1988;263:7472–7.
- Oren Z, Shai Y. Mode of action of linear amphipathic alpha-helical antimicrobial peptides. *Biopolymers*. 1998;47:451–63.
- Park CJ, Park CB, Hong SS, Lee HS, Lee SY, Kim SC. Characterization and cDNA cloning of two glycine- and histidine-rich antimicrobial peptides from the roots of shepherd's purse, *capsella bursa-pastoris*. *Plant Mol Biol*. 2000;44:187–97.
- Pestonjamasp VK, Huttner KH, Gallo RL. Processing site and gene structure for the murine antimicrobial peptide CRAMP. *Peptides*. 2001;22:1643–50.
- Pierre JF, Peters BM, la Torre D, Sidebottom AM, Tao Y, Zhu X, et al. Peptide YY: a Paneth cell antimicrobial peptide that maintains *Candida* gut commensalism. *Science*. 2023;381:502–8.
- Salzman NH, Bevins CL. Dysbiosis—a consequence of Paneth cell dysfunction. *Semin Immunol*. 2013;25:334–41.
- Snelders NC, Petti GC, van den Berg GCM, Seidl MF, Thomma BPHJ. An ancient antimicrobial protein co-opted by a fungal plant pathogen for in planta mycobiome manipulation. *Proc Natl Acad Sci USA*. 2021;118:e2110968118.
- Sørensen OE, Gram L, Johnsen AH, Andersson E, Bangsbøll S, Tjabringa GS, et al. Processing of seminal plasma hCAP-18 to ALL-38 by gastricsin: a novel mechanism of generating antimicrobial peptides in vagina. *J Biol Chem*. 2003;278:28540–6.
- Stansly PG, Shepherd RG, White HJ. Polymyxin: a new chemotherapeutic agent. *Bull Johns Hopkins Hosp*. 1947;81:43–54.
- Steiner H, Hultmark D, Engström A, Bennich H, Boman HG. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature*. 1981;292:246–8.
- Thompson SA, Tachibana K, Nakanishi K, Kubota I. Melittin-like peptides from the shark-repelling defense secretion of the sole *Pardachirus pavoninus*. *Science*. 1986;233:341–3.
- Toda H, Williams JA, Gullledge M, Sehgal A. A sleep-inducing gene, *nemuri*, links sleep and immune function in *Drosophila*. *Science*. 2019;363:509–15.
- Vogt TC, Bechinger B. The interactions of histidine-containing amphipathic helical peptide antibiotics with lipid bilayers. The effects of charges and pH. *J Biol Chem*. 1999;274:29115–21.
- Wang CY, Kuang X, Wang QQ, Zhang GQ, Cheng ZS, Deng ZX, et al. GMMAD: a comprehensive database of human gut microbial metabolite associations with diseases. *BMC Genomics*. 2023;24:482.
- Wang G. Determination of solution structure and lipid micelle location of an engineered membrane peptide by using one NMR experiment and one sample. *Biochim Biophys Acta*. 2007;1768:3271–81.
- Wang G. The antimicrobial peptide database provides a platform for decoding the design principles of naturally occurring antimicrobial peptides. *Protein Sci*. 2020a;29:8–18.
- Wang G. Bioinformatic analysis of 1000 amphibian antimicrobial peptides uncovers multiple length-dependent correlations for peptide design and prediction. *Antibiotics*. 2020b;9:491.
- Wang G. Unifying the classification of antimicrobial peptides in the antimicrobial peptide database. *Methods Enzymol*. 2022;663:1–18.
- Wang G, Li X, Wang Z. APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res*. 2009;37(Database issue):D933–7.
- Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res*. 2016;44:D1087–93.
- Wang G, Li Y, Li X. Correlation of three-dimensional structures with the antibacterial activity of a group of peptides designed based on a nontoxic bacterial membrane anchor. *J Biol Chem*. 2005;280:5803–11.
- Wang G, Mishra B, Epanand RF, Epanand RM. High-quality 3D structures shine light on antibacterial, anti-biofilm and antiviral activities of human cathelicidin LL-37 and its fragments. *Biochim Biophys Acta*. 2014;1838:2160–72.
- Wang G, Vaisman II, van Hoek ML. Machine learning prediction of antimicrobial peptides. *Methods Mol Biol*. 2022;2405:1–37.
- Wang G, Watson KM, Peterkofsky A, Buckheit RW Jr. Identification of novel human immunodeficiency virus type 1-inhibitory

- peptides based on the antimicrobial peptide database. *Antimicrob Agents Chemother.* 2010;54:1343–6.
- Wang G, Zietz CM, Mudgapalli A, Wang S, Wang Z. The evolution of the antimicrobial peptide database over 18 years: milestones and new features. *Protein Sci.* 2022;31:92–106.
- Wang X, Mishra B, Lushnikova T, Narayana JL, Wang G. Amino acid composition determines peptide activity spectrum and hotspot-based design of mercedin. *Adv Biosyst.* 2018;2:1700259.
- Wang Z, Wang G. APD: the antimicrobial peptide database. *Nucleic Acids Res.* 2004;32:D590–2.
- Wehkamp J, Fellermann K, Herrlinger KR, Bevins CL, Stange EF. Mechanisms of disease: defensins in gastrointestinal diseases. *Nat Clin Pract Gastroenterol Hepatol.* 2005;2:406–15.
- Wimley WC. Describing the mechanism of antimicrobial peptide action with the interfacial activity model. *ACS Chem Biol.* 2010;5:905–17.
- Wüthrich K. Sequential individual resonance assignments in the ¹H-NMR spectra of polypeptides and proteins. *Biopolymers.* 1983;22:131–8.
- Xiao X, Wang P, Lin WZ, Jia JH, Chou KC. iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Anal Biochem.* 2013;436:168–77.
- Zasloff M. Magainins, a class of antimicrobial peptides from *Xenopus* skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc Natl Acad Sci USA.* 1987;84:5449–53.
- Zasloff M. Antimicrobial peptides of multicellular organisms. *Nature.* 2002 Jan;24(415):389–95.
- Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, et al. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature.* 2016;535:511–6.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang G. The antimicrobial peptide database is 20 years old: Recent developments and future directions. *Protein Science.* 2023;32(10):e4778. <https://doi.org/10.1002/pro.4778>