

Article

# Structural and Functional Enrichment Analyses for Antimicrobial Peptides

Sheng C. Lo<sup>1</sup>, Zhong-Ru Xie<sup>2</sup> and Kuan Y. Chang<sup>1,\*</sup>

<sup>1</sup> Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung 202, Taiwan

<sup>2</sup> School of Electrical and Computer Engineering, University of Georgia, Athens 30602, USA

\* Correspondence: kchang@ntou.edu.tw; Tel.: +886-2-2462-2192

**Abstract:** Whether there is any inclination between structures and functions of antimicrobial peptides (AMPs) is a mystery yet to be unraveled. AMPs have various structures associated with many different antimicrobial functions, including antibacterial, anticancer, antifungal, antiparasitic and antiviral activities. However, none has yet reported any antimicrobial functional tendency within a specific category of protein/peptide structures nor any structural tendency of a specific antimicrobial function with respect to AMPs. Here we examine the relationships between structures categorized by three structural classification methods (CATH, SCOP and TM) and seven antimicrobial functions with respect to AMPs using an enrichment analysis. The results show that antifungal activities of AMPs were tightly related to two-layer sandwich structure of CATH, knottin fold of SCOP, and the first structural cluster of TM, especially to knottin, whose robust association sustained through the AMPs with a low sequence identity. In addition, another significant mutual enrichment was observed between the third cluster of TM and anti-gram-positive-bacterial/anti-gram-negative-bacterial activities. The findings of the structure-function inclination further our understanding of AMPs and could help us design or discover new therapeutic potential AMPs.

**Keywords:** structure-function relationships; enrichment analysis; antifungal activities; knottin; two-layer sandwich architecture

---

## 1. Introduction

Antimicrobial peptides (AMPs) or proteins have a broad spectrum of biological activities and various structures. On one hand, these biological activities of AMPs include antibacterial, anti-fungal, antiparasitic, antiviral, and anticancer activities. On the other hand, various structures of natural AMPs such as cathelicidin, defensin, and transferrin have been resolved [1]. However, little is known about propensities between antimicrobial activities and AMP structures.

Attention has been paid to AMP structure-function relationships. One well-studied example is defensins found in fungi, plants, and animals. Defensins have multiple intramolecular disulfide bonds between at least six conserved cysteine residues, which maintain their structures against chemical and proteolytic degradation and are crucial to their antimicrobial functions. Moreover, positively charged residues along with amphipathic characters within the structures have also been linked to their functions such as antifungal activities [2,3]. Although defensins exhibit a broad spectrum of antimicrobial activities including antibacterial, antifungal, and antiviral activities, antifungal activities are mostly observed, especially for plant defensins [2,3]. However, none demonstrates any functional tendency toward any structural groups of AMPs like defensins nor does any structural tendency toward any antimicrobial activities of AMPs.

To establish the associations between antimicrobial activities and AMP structures, we perform large-scale systematic enrichment analyses based on three kinds of structural classification (CATH (Class, Architecture, Topology, Homology) [4], SCOP (Structural Classification of Proteins) [5], and TM (template modeling) [6]. This study broadens our understanding of AMP structure-function

relationships, which could benefit for finding or designing therapeutic potential AMPs as peptide antibiotics.

## 2. Results

### 2.1. AMP Structure-Function Enrichment Analysis

Table 1 shows the enrichment relationships between antimicrobial functions and AMP structures under three kinds of structural classification (CATH, SCOP, and TM). Up to the second levels of CATH and SCOP are all we could examine, for any higher level fails to reach statistical significance due to limited sample size. The most noticeable mutual enrichments happen between antifungal activities and the SCOP Small Proteins Class. Even under low sequence similarity with  $\leq 40\%$  sequence identity, the mutual relationships remain. Furthermore, antifungal activities are persistently enriched in the SCOP Knottin Fold of Small Proteins (Fig. 1), under  $\leq 50\%$  sequence identity. More details about the enrichment results can be found at Supplementary Table 1-6.

**Table 1.** Enriched functions and structures of antimicrobial peptides.

| Type           | Structure                               |                       | Sequence Identity Threshold (%) |    |    |    |    |    |    | F(x) |    |
|----------------|---|-----------------------|---------------------------------|----|----|----|----|----|----|------|----|
|                | 1 <sup>st</sup> level                   | 2 <sup>nd</sup> level | 100                             | 90 | 80 | 70 | 60 | 50 | 40 |      |    |
| CATH           | $\alpha \beta$                          | *                     | ↔                               |    |    |    |    |    |    | F    |    |
|                |   | 2-Layer Sandwich      | ↔                               | ↔  | ↔  |    |    |    |    | F    |    |
|                |   | Roll                  | ←                               |    |    |    |    |    |    | G    |    |
|                |   | Roll                  | ←                               |    |    |    |    |    |    | G-   |    |
|                | Mainly $\alpha$                         | *                     | ←                               | ←  | ←  | ←  | ←  |    |    |      | G  |
|                |   | *                     | ←                               |    |    |    |    |    |    |      | G+ |
|                |   | Up-down Bundle        | ←                               |    |    |    |    |    |    |      | G  |
|                |   | Orthogonal Bundle     | ←                               |    |    |    |    |    |    |      | G  |
|                | Mainly $\beta$                          | *                     | →                               |    |    |    |    |    |    |      | P  |
|                |   | Beta Barrel           | ←                               |    |    |    |    |    |    |      | G- |
|                | Few 2 <sup>nd</sup> -ary Structures     | *                     | ↔                               |    |    |    |    |    |    |      | F  |
|                |   | Irregular             | →                               |    |    |    |    |    |    |      | F  |
| SCOP           | $\alpha$ and $\beta$ ( $\alpha+\beta$ ) | *                     | ←                               |    |    |    |    |    |    | G    |    |
|                |   | *                     | ←                               |    |    |    |    |    |    | G+   |    |
|                |   | IL8-like              | ←                               |    |    |    |    |    |    |      | G- |
|                | Peptides                                | *                     | →                               |    |    |    |    |    |    |      | G  |
|                |   | *                     | →                               |    |    |    |    |    |    |      | G+ |
|                |   | RP71935               | ↔                               |    |    |    |    |    |    |      | V  |
| Small proteins | *                                       | ↔                     | ↔                               | ↔  | ↔  | ↔  | ↔  | ↔  | ↔  | F    |    |
|                | Knottins                                | ↔                     | ↔                               | ↔  | ↔  | ↔  | ↔  |    |    | F    |    |
|                | Crambin-like                            | ↔                     | ↔                               |    |    |    |    |    |    | C    |    |
|                | Defensin-like                           | ←                     |                                 |    |    |    |    |    |    | G-   |    |
| TM             | TM1                                     | ↔                     | ↔                               | ↔  | →  | →  |    |    |    | F    |    |
|                | TM2                                     | ←                     |                                 |    |    |    |    |    |    | G-   |    |
|                | TM3                                     | ↔                     |                                 |    |    |    |    |    |    | G    |    |
|                | TM3                                     | ↔                     | →                               | ↔  | →  |    |    |    |    | G+   |    |
|                | TM3                                     | ↔                     | ↔                               | ↔  | →  |    |    |    |    | G-   |    |
|                | TM4                                     | →                     | →                               | →  | →  |    |    |    |    | V    |    |
|                | TM5                                     | ↔                     | ↔                               | →  | →  |    |    |    |    | V    |    |

Abbreviations: C: Anti-cancer; F: Anti-fungal; G: Anti-bacterial; G+: Anti-gram+; G-: Anti-gram-; P: Anti-parasitic; V: Anti-viral; F(x): Function; →: Enriched function; ←: Enriched structure; ↔: Structure-function mutual enrichment; \*: All the structures within the classification. Benjamini-Hochberg correction [7] was applied to adjust P values at  $Q < 0.05$ .

The mutual enrichments involved with antifungal activities also occur to TM and CATH structures: TM cluster 1 and the CATH two-layer Sandwich Architecture under highly similar AMP sequences ( $\leq 80\%$  sequence identity). Under highly similar sequences, some other than antifungal activities also demonstrate a structural tendency or vice versa. These structure-function mutual tendencies include the following: TM structural cluster 3 and anti-gram-negative bacterial activities; TM cluster 3 and anti-gram-positive bacterial activities; TM cluster 5 and antiviral activities. In addition, weak mutual enrichments are observed between anticancer activities and the SCOP Crambin-like Fold.

As to one-direction enrichment, TM cluster 4 possesses enriched antiviral activities, but antiviral activities would not prefer to be TM cluster 4; antibacterial activities favor mainly alpha proteins of CATH, but alpha proteins need not possess antibacterial activities.

There are several functional enrichments or structural enrichments for the representative AMPs at 100% sequence identity, but not at other thresholds. For example, anti-gram-negative-bacterial activities are enriched in defensin-like structures for the original AMPs; mainly beta proteins of CATH have more antiparasitic activities than expected.

## 2.2. AMP Structure-Function Enrichments Associated with Sequences

Figure 1 illustrates the sequence motifs and associated Pfams for the AMP structure-function enrichments. According to biological activities of AMPs, two kinds of structure-function enrichments remain mutually significant at 80% sequence identity: one associated with antifungal activities and the other with anti-gram-negative/anti-gram-positive activities.

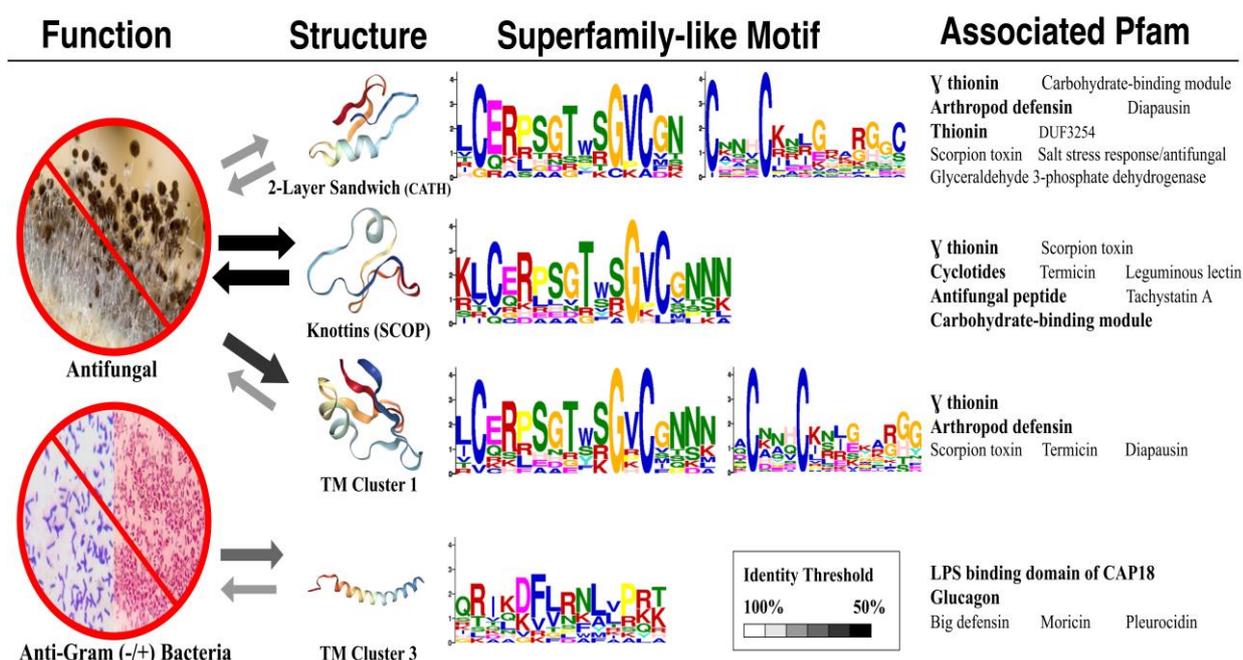


Fig. 1. AMP structure-function enrichments associated with sequence motifs and Pfams.

### 2.2.1. Superfamily-like sequence motifs

The sequence motifs of the structure-function enrichments associated with the antifungal activities are similar, albeit with some differences. All the motifs display multiple conserved cysteine residues. In primary motifs, a distinct glycine beside a conserved cysteine stands out although the primary motif associated with knottins is longer. Moreover, the secondary cysteine motifs are only detectable in the enrichments associated with 2-layer sandwich of CATH and TM structural cluster

1. The biological meanings of the conserved cysteine need to be investigated by further biological and computational experiments.

As for the enrichment associated with anti-gram-negative/anti-gram-positive activities, its sequence motif is rich in positively charged amino acids. These conserved arginine and lysine residues appear periodically, supporting an amphiphilic character.

### 2.2.2. Associated Pfam families or domains

Regardless of structure classification, the Pfam [8] associated with the structure-function enrichments with antifungal activities consistently include  $\gamma$  thionin, arthropod defensin, scorpion short toxin and toxin-like domain. Furthermore, the additional families associated with knottins include cyclotides and antifungal peptide families; those with other structures contain diapausin-related AMPs. As for those associated with anti-gram-negative/anti-gram-positive activities, C-terminal lipopolysaccharide binding domain of CAP18 and glucagon-related peptide hormone are frequently found.

## 3. Discussion

We demonstrate that there exist inclinations between structures and functions with respect to AMPs. The most significant AMP structure-function enrichments we observed happened to be antifungal activities associated concurrently with CATH's two-layer sandwich, SCOP's knottin, and TM's cluster 1. Surprisingly, CATH, SCOP, and TM all could detect this important antifungal association, suggesting the association was strong enough under any structural classification. Fig. 1 further shows the high AMP sequence similarities among the three associated structures, indicating that these structures were alike in the make-up despite being obtained through three different classification approaches. The motifs with conserved cysteine and glycine, which they all had, similar to the  $\gamma$ -core motif of antifungal plant defensins [2] might hold a key to antifungal activities of peptides.

However, the three structure-function associations were still different. The most significant differences were only SCOP's knottins, characterized by a distinguished "disulfide through disulfide knot", to remain significantly with antifungal activities at a low sequence identity. Peptide knots used to be thought rare, but now have been found widely distributed in eukaryotic organisms [9]. We speculate that both natural and artificial knottins could be potential pharmacological agents. Although knottins are known to possess several functions such as analgesics, anthelmintic, antimalarial, and antimicrobial activities [10], this is the first study to reveal enriched antifungal activities within knottins and enriched knottins among the AMPs with antifungal activities.

As for the TM cluster 3 enriched with anti-gram-negative/anti-gram-positive activities, the positively-charged amphipathic AMPs, such as arginine, may appeal to negatively-charged lipopolysaccharide of outer membrane of gram-negative bacterial cell wall as well as negatively-charged teichoic acid of gram-positive bacterial cell wall to interact with.

As for one-direction enrichment, the result of the inclination of the TM cluster 4 toward antiviral activities, not the opposite, indicates that antiviral activities of AMPs need not to be the structures of the TM cluster 4 although several AMPs coming from the TM cluster 4 possess antiviral activities. Interestingly, the result of the inclination of antibacterial activities toward mainly alpha proteins, not the opposite, suggests that alpha helices alone are not destined for antimicrobial functions and there are other major contributing factors such as amphipathicity to antibacterial activities.

This study has some limitations. First, the scope of structural classification in this study is restricted to CATH's architecture and SCOP's fold. We cannot determine any structural preferences higher than architecture/fold such as topology/superfamily due to limited number of AMPs. For further structural exploration, more AMPs are required to pass statistical significances. Second, some activities of AMPs may have been misannotated or have not been validated experimentally. We not only examined all the AMP annotation carefully but also focused on mutual structure-function enrichments to eliminate minor or less relevant relationships.

Understanding AMP structure-function enrichments could lead to identifying novel AMPs or designing potent and therapeutic peptides as antibiotics alternatives. Our findings may boost future studies in different AMP areas such as *de novo* computational design of AMPs [11]. For example, we may take the advantage of the structural template of the TM cluster 3 to design novel AMPs with potent anti-gram-negative/anti-gram-positive activities.

## 4. Materials and Methods

### 4.1. Antimicrobial peptides

3,061 non-redundant experimentally-verified AMPs with length  $\geq 8$  amino acids used in this study were obtained by combining 2,774 sequences of the Collection of Anti-Microbial Peptides (CAMPR3) [12], 2,619 sequences of the Antimicrobial Peptide Database 3 (APD3) [13], and 3,152 sequences of A Database of Anti-Microbial peptides (ADAM) [1].

Each AMP was annotated with at least one and at most seven different activities: anti-bacterial, anti-gram-positive bacterial, anti-gram-negative bacterial, anti-fungal, anti-viral, anti-parasitic, and anti-cancer.

### 4.2. Antimicrobial peptide structures

By running the compiled 3,061 AMP sequences against the Protein Data Bank (PDB) [14], 445 AMP structures were identified.

The AMP structures in this study were annotated by three structural classification approaches—CATH (v4.2) [4], SCOPe (v2.07) [5], and TM. Not every AMP structure has CATH or SCOP annotations, but each AMP structure can be classified into one of 134 TM structural fold clusters.

### 4.3. Representative antimicrobial peptides

Many-against-Many sequence searching tool 2 (MMseqs2) was used to obtain AMP families by clustering highly similar sequences together and select one representative AMP sequence for each family [15]. Using sequence identity thresholds which ranged from 100% to 40%, seven representative AMPs sets were generated.

### 4.4. Structure classification methods

#### 4.4.1. CATH structural classification

There are four basic levels of CATH structural hierarchy. At the Class (C) level, proteins are divided into four categories based on the secondary structure composition: mainly alpha, mainly beta, alpha beta, and few secondary structures. Most proteins fall into the former three categories. The Architecture (A) level, the second CATH level, groups proteins based on spatial arrangements of secondary structures. For example, orthogonal bundle, 2-layer sandwich, or alpha-beta barrel. At the Topology (T) level, the third level, the order of secondary structures is used to classify protein folds with specific structural characteristics. At the Homologous superfamily (H) level, the fourth CATH level, proteins are assigned to the same superfamily if their sequences and structures are highly similar, suggesting a plausible common evolutionary ancestor [4,16].

CATH has been exploring protein structure-function relationships according to protein evolution. Functional Families (Fun-Fam), which tend to have highly similar structures and functions, are formed by clustering protein domains (sequences and structures) within homologous super families[4].

#### 4.4.2. SCOP structural classification

Five basic levels of the SCOP database are Class, Fold, Superfamily, Family, and domain. The SCOP database used to heavily rely on manual classification with the focus on protein evolution and structure similarities.

Instead, the SCOPe (SCOP-extended) database [5] which inherited the SCOP and corrected some classification errors by integrating the Astral database has become more automatic. Like the first CATH level, the SCOP Class only considers the overall secondary structure composition. There are 12 classes in the SCOPe Class while most proteins are assigned to five classes, namely all alpha, all beta, alpha and beta proteins ( $\alpha/\beta$ ) that contain  $\alpha$ -helices and  $\beta$ -folds, alpha and beta proteins ( $\alpha+\beta$ ) with separate  $\alpha$ -helices and  $\beta$ -folds, and small proteins. The SCOP Fold further divides proteins by topology and architecture. It is not until the second level that the divergence between CATH and SCOP becomes apparent. The SCOP Superfamilies are more distant related than the Families while each Superfamily contains all the domains in a fold and the proteins within the same SCOP Family which share at least 30% sequence identity often have distinct functions. There are two additional levels above domain in the SCOPe: Species and Protein domain. Species represents a distinct protein sequence derived from a specific source while protein domain gathers isoforms or similar sequences from any sources.

#### 4.4.3. TM structural classification

To overcome the drawback of conventional protein structural classification methods, the TM structural classification introduces the TM-scores [6]. Conventional structural classification often fails to recognize similar protein structures with different lengths, for it relies on calculating the length-sensitive root mean square deviation (RMSD). To evaluate protein structure similarity regardless of their lengths or sizes, the TM method adopts a geometric goodness of fit approach (the TM-score), which utilizes the Levitt-Gerstein weight factor that favors close residue pairs to distant ones.

The TM-score of two structures ranges from 0 to 1. When the TM-score  $\geq 0.5$ , the proteins are considered to fall into identical topological structures; otherwise, the proteins have different structures [17].

The order of the TM clusters used in this study were adopted from ADAM [1].

#### 4.5. Enrichment analysis

The enrichment analysis uses the hypergeometric test to check whether a specific property is over- /under-represented in a sample based on the hypergeometric distribution. If the number of objects with the specific property is significantly greater than expected compared to the population, it is enriched or over-represented; otherwise, it is not. To check whether a specific property is enriched in a sample, the following formula is used:

$$P = \sum_{i=m}^n \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}} \quad (1)$$

where  $N$  is the population size,  $n$  is the sample size, and  $M$  and  $m$  are the numbers of objects with the specific property in the population and in the sample, respectively. The  $P$  value which refers to the probability of at least  $m$  objects occurring in the sample must be  $<0.05$  for enrichment. Due to multiple hypotheses, the  $P$  values are adjusted by Benjamini-Hochberg correction [7].

We examined two different concepts using the enrichment analysis. First, we examined whether there are any anti-microbial functional tendencies within a specific structure, where in terms of the number of AMP sequences,  $N$  and  $M$  are the population size with any functions and with a specific function, respectively, while  $n$  and  $m$  are the sample size of the specific structure with any functions and with the specific function, respectively. Second, we examined whether there are any structural tendencies for a specific anti-microbial function, where  $N$  is the population size under a structural classification,  $M$  is the subpopulation size with a specific structure under this classification,  $n$  and  $m$  are the sample size of a specific anti-microbial function with any structures and with the specific structure, respectively.

## 5. Conclusions

Structural tendencies and functional tendencies exist with respect to AMPs. The structure-function mutual tendency happens to antifungal activities of AMPs with two-layer sandwich structure of CATH, the first structural cluster of TM, and especially knottin fold of SCOP. This study gives hints how to apply protein engineering to design AMPs as therapeutic agents by modifying their sequences and structures.

**Supplementary Materials:** Supplementary materials can be found at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1). Table S1. Enriched antimicrobial activities within a specific protein/peptide structure categorized by CATH. Table S2. Enriched CATH structure of a specific antimicrobial function with respect to AMPs. Table S3. Enriched antimicrobial activities within a specific protein/peptide structure categorized by SCOP. Table S4. Enriched SCOP structure of a specific antimicrobial function with respect to AMPs. Table S5. Enriched antimicrobial activities within a specific protein/peptide structure categorized by TM. Table S6. Enriched TM structure of a specific antimicrobial function with respect to AMPs.

**Author Contributions:** Conceptualization, K.Y.C.; methodology, S.C.L. and K.Y.C.; software, S.C.L. and K.Y.C.; formal analysis, S.C.L. and K.Y.C.; investigation, S.C.L. and K.Y.C.; data curation, S.C.L.; writing—original draft preparation, S.C.L., Z.-R.X. and K.Y.C.; writing—review and editing, Z.-R.X. and K.Y.C.; visualization, S.C.L.; supervision, K.Y.C.; project administration, K.Y.C.; funding acquisition, K.Y.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Taiwan Ocean University under Grant NTOU-107-009.

**Acknowledgments:** We would like to thank Dr. Wu Ming-Shi for constructive comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

|      |   |
|------|---|
| ADAM | A Database of Anti-Microbial peptides   |
| AMP  | Antimicrobial peptide                   |
| CATH | Class, Architecture, Topology, Homology |
| SCOP | Structural Classification of Proteins   |
| TM   | Template Modeling                       |

## References

1. Lee, H.T.; Lee, C.C.; Yang, J.R.; Lai, J.Z.; Chang, K.Y. A large-scale structural classification of antimicrobial peptides. *Biomed Res Int* **2015**, *2015*, 475062, doi:10.1155/2015/475062.
2. Lacerda, A.F.; Vasconcelos, E.A.; Pelegrini, P.B.; Grossi de Sa, M.F. Antifungal defensins and their role in plant defense. *Front Microbiol* **2014**, *5*, 116, doi:10.3389/fmicb.2014.00116.
3. Tam, J.P.; Wang, S.; Wong, K.H.; Tan, W.L. Antimicrobial Peptides from Plants. *Pharmaceuticals (Basel)* **2015**, *8*, 711-757, doi:10.3390/ph8040711.
4. Dawson, N.L.; Lewis, T.E.; Das, S.; Lees, J.G.; Lee, D.; Ashford, P.; Orengo, C.A.; Sillitoe, I. CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res* **2017**, *45*, D289-D295, doi:10.1093/nar/gkw1098.
5. Chandonia, J.M.; Fox, N.K.; Brenner, S.E. SCOPe: classification of large macromolecular structures in the structural classification of proteins-extended database. *Nucleic Acids Res* **2019**, *47*, D475-D481, doi:10.1093/nar/gky1134.
6. Zhang, Y.; Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **2005**, *33*, 2302-2309, doi:10.1093/nar/gki524.
7. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **1995**, 289–300.

8. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A., et al. The Pfam protein families database in 2019. *Nucleic Acids Res* **2019**, *47*, D427-D432, doi:10.1093/nar/gky995.
9. Molesini, B.; Treggiari, D.; Dalbeni, A.; Minuz, P.; Pandolfini, T. Plant cystine-knot peptides: pharmacological perspectives. *Br J Clin Pharmacol* **2017**, *83*, 63-70, doi:10.1111/bcp.12932.
10. Postic, G.; Gracy, J.; Perin, C.; Chiche, L.; Gelly, J.C. KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res* **2018**, *46*, D454-D458, doi:10.1093/nar/gkx1084.
11. Schneider, G.; Clark, D.E. Automated De Novo Drug Design: Are We Nearly There Yet? *Angew Chem Int Ed Engl* **2019**, *58*, 10792-10803, doi:10.1002/anie.201814681.
12. Wagh, F.H.; Barai, R.S.; Gurung, P.; Idicula-Thomas, S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res* **2016**, *44*, D1094-1097, doi:10.1093/nar/gkv1051.
13. Wang, G.; Li, X.; Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* **2016**, *44*, D1087-1093, doi:10.1093/nar/gkv1278.
14. Burley, S.K.; Berman, H.M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J.M.; Dutta, S., et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res* **2019**, *47*, D464-D474, doi:10.1093/nar/gky1004.
15. Mirdita, M.; Steinegger, M.; Soding, J. MMseqs2 desktop and local web server app for fast ,interactive sequence searches. **2018**, 10.1101/419895, doi:10.1101/419895.
16. Michie, A.D.; Orengo, C.A.; Thornton, J.M. Analysis of Domain Structural Class Using an Automated Class Assignment Protocol. *Journal of Molecular Biology* **1996**, *262*, 168-185, doi:<https://doi.org/10.1006/jmbi.1996.0506>.
17. Xu, J.; Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **2010**, *26*, 889-895, doi:10.1093/bioinformatics/btq066.