



# ADAPTABLE: a comprehensive web platform of antimicrobial peptides tailored to the user's research

Francisco Ramos-Martín, Thibault Annaval, Sébastien Buchoux, Catherine Sarazin, Nicola D'Amelio 

**Antimicrobial peptides (AMPs) are part of the innate immune response to pathogens in all of the kingdoms of life. They have received significant attention because of their extraordinary variety of activities, in particular, as candidate drugs against the threat of super-bacteria. A systematic study of the relation between the sequence and the mechanism of action is urgently needed, given the thousands of sequences already in multiple web resources. ADAPTABLE web platform (<http://gec.u-picardie.fr/adaptable>) introduces the concept of “property alignment” to create families of property and sequence-related peptides (SR families). This feature provides the researcher with a tool to select those AMPs meaningful to their research from among more than 40,000 nonredundant sequences. Selectable properties include the target organism and experimental activity concentration, allowing selection of peptides with multiple simultaneous actions. This is made possible by ADAPTABLE because it not only merges sequences of AMP databases but also merges their data, thereby standardizing values and handling non-proteinogenic amino acids. In this unified platform, SR families allow the creation of peptide scaffolds based on common traits in peptides with similar activity, independently of their source.**

DOI [10.26508/lsa.201900512](https://doi.org/10.26508/lsa.201900512) | Received 1 August 2019 | Revised 7 November 2019 | Accepted 8 November 2019 | Published online 18 November 2019

## Introduction

Antimicrobial peptides (AMPs) are a class of molecules that have attracted significant attention for their antibacterial properties (Wang et al, 2015). As part of organisms' innate immunity, AMPs have been found virtually in all life kingdoms, including marine and terrestrial animals, bacteria, and plants (Zasloff, 2002; Goyal & Mattoo, 2016; Wang et al, 2017). However, particular attention has been devoted to those produced by animal venoms and frog and toad skin secretions (Xu & Lai, 2015). They display activity against a wide range of targets: bacteria, viruses, fungi, parasites, insects, and cancer cells (Tamamura et al, 1998; Albiol Matanic & Castilla, 2004; Hoskin & Ayyalusamy, 2008; Joo et al, 2012; Lacerda et al, 2014; Yang et al, 2014; Field et al, 2015). They also modulate inflammation

processes and cell–cell communications (Oyinloye et al, 2015). Thousands of AMPs are already known (Aguilera-Mendoza et al, 2015), but the number is destined to grow as hundreds of new peptides are discovered or synthetically produced each year (Wang et al, 2015). It is apparent that fast throughput methods are needed to enable their mechanism of action to be determined.

### A novel tool in the era of antimicrobial resistance

ADAPTABLE (Antimicrobial PeptiDe scAffold by Property alignment). A web platform for cLustering and dESign; <http://gec.u-picardie.fr/adaptable>) was created to provide a tool for all scientists in the field of AMPs who are interested in developing new drugs against a well-defined target. For example, according to the World Health Organization, there are 12 microorganisms considered the greatest danger towards human health (Vogel, 2017) because of their resistance to known antibiotics. For each microorganism, ADAPTABLE is able to generate SR families containing peptides known to be active against it and classify them, thus highlighting essential traits. Alternatively, each of these peptides can be used as a bait to generate their own SR families by scanning the ADAPTABLE database featuring more than 40,000 entries. Although sequence alignment has been widely used among AMPs with strictly related biological source, we believe that the comparison of evolutionarily distant AMPs with similar activity can highlight the features that are key to their mechanism of action. This unbiased search can be used to spot existing, but untested, sequence-related peptides of similar structure and mechanism of action that also have the properties of more-promising drug candidates (e.g., less hemolytic).

### A different classification strategy based on property alignment

Expression of many similar AMPs by the same cell type, organism, or genus has resulted in AMPs to be classified in families based on sequence alignment of peptides with closely related biological origin (Zouhir et al, 2010; Tassanakajon et al, 2015; Xu & Lai, 2015). However, common features can be found in families of all origins, such as (i) the recurrent presence of positively charged residues, thought to draw the peptide towards the negatively charged membranes of bacteria and cancer cells, and (ii) a significant fraction of

hydrophobic residues facilitating the interaction with the lipid bilayer. This suggests that the clustering of sequences with specific activities (what we call “property alignment”) independent of the evolutionary distance can highlight those key features underpinning mechanisms of action.

Several methodologies and software already exist for the classification and prediction of AMP activities (Brahmachary et al, 2004; Fjell et al, 2007; Wang et al, 2011; Ng et al, 2015; Bhadra et al, 2018). Most provide tools for sequence similarity searches, and only few have algorithms for predicting the activity of any given peptide sequence (e.g., APD [Wang, 2014], CAMPR3 [Waghu et al, 2016], AVPPred [Thakur et al, 2012], YADAMP [Piotto et al, 2012], and ToxinPred [Gupta et al, 2013]). However, none of these combine a systematic global classification with “adaptability” (hence, the name ADAPTABLE) to the focus of the user’s research.

### A different approach tailored to researchers’ aims

One important difference of ADAPTABLE with respect to other alignment tools is that it provides the facility to study peptides active against a specific pathogen, even when a researcher does not already have a specific peptide of interest. ADAPTABLE generates many different SR families with activity against a specific pathogen, optionally including one or more peptides provided by the user. The members of different SR families are significantly different in sequence and possibly represent different mechanisms of action. Each SR family is represented by the peptide used as bait for its creation, also called “father.” The researcher can then study each father to get insights into the molecular causes of the activity and/or concentrate on the SR family that his own peptide has been assigned to.

### A comprehensive and standardized database

Whereas the power of the algorithms relies on large number of entries, the concept of property alignment requires standardization. In other words, the development of ADAPTABLE web platform has implied the creation of a comprehensive and standardized database that addresses a further fundamental problem in the study of AMPs: information scattering across multiple web resources. Existing databases have previously been merged (Aguilera-Mendoza et al, 2015), but previous standardization only collected and merged sequence information. In contrast, ADAPTABLE merges information for more than 60 properties from 25 databases specialized in both medical and agricultural fields where available and creates a unified, nonredundant entry. This way, each sequence is associated to all available data on its structure (predicted or experimental), activity (e.g., anticancer, antibacterial, antibiofilm, antiviral, antifungal, or antiparasitic), targets (e.g., lung cancer, HeLa cells, HIV virus, cell membrane target, *Klebsiella pneumoniae*, or *Plasmodium falciparum*), and other information (e.g., experimental validation, taxonomy, bibliography, or DSSP—“define secondary structure of proteins” [Kabsch & Sander, 1983] data).

Merging of data from different sources requires standardization. This task can be challenging because each existing database follows its own format and definitions. For instance, activities are reported in many different units or measured by different tests. Our

algorithm converts all activities to micromolar ( $\mu\text{M}$ ) concentration by calculating the molecular weight of the peptide, even for non-proteinogenic amino acids. Even if different activity values are not always directly comparable because of different experimental conditions or activity tests, an upper threshold value allows the filtering out of less active peptides, while the “activity test” toggle allows restriction to a single test.

To our knowledge, ADAPTABLE is the only tool able to standardize activities and non-proteinogenic amino acids, including modified and non-natural amino acids. Non-proteinogenic amino acids lack a standardized one-letter code, resulting in confusion when comparing sequences across different databases. This ambiguity often leads to redundant entries referring to the same sequence being nonuniformly annotated. Furthermore, some non-proteinogenic amino acids are named by different synonyms rather than a unified nomenclature that, if it existed, would allow unambiguous identification. ADAPTABLE addresses both issues by interpreting the different nomenclatures and providing a single name based on the PubChem database (Kim et al, 2019) and a single one-letter symbol. For non-proteinogenic amino acids, the closest proteinogenic homologue is also identified, where possible.

ADAPTABLE takes the standardization process one step forward, thanks to its inclusion of data from a specialized microbiology database (“The Microbe directory” [Shaaban et al, 2018]). This allows the inclusion of potentially missing information such as the full names of organisms, their nature (i.e., Gram positive or negative bacteria, fungi, or virus, among others), and some of their properties (i.e., ability to form biofilms). The same approach is followed to complement and store structural information, either predicted or experimental, leveraging ADAPTABLE’s integration with the Protein Data Bank (Berman et al, 2000) and the usage of PSSpred (Yan et al, 2013) and I-TASSER (Roy et al, 2010; Yang et al, 2015; Yang & Zhang, 2015).

## Results and Discussion

### Family generator tool

#### Defining the subset of peptides

ADAPTABLE offers the possibility to generate SR families using subsets of peptides with user-defined characteristics (“Family Generator” page shown in Fig 1; for details, see the case example n.1, “Designing new peptides active towards a specific organism and highlighting motifs”, in Supplemental Data 1, “ADAPTABLE Tutorial”). This tool requires a “calculation label” and a “username” that allow the user to privately analyze (“Family Analyzer” page) or download the results (“Download Results” page) at the end of the calculation. Email is used to notify the user about the start and the end of the run, together with the “Calculation label”.

The subset of peptides can be defined by their name, sequence pattern, and target organism. The expandable “Advanced” and “Peptide properties” sections provide further tuning of the required peptide properties. By selecting from among the more than 60 parameters, the user can choose the target organism, activities, or chemical or physical properties, and also stipulate other parameters (source, taxonomy, posttranscriptional modifications, N-terminal or C-terminal

## Family generator

In this page you can create a family of antimicrobial peptides featuring user-selected properties. Please take a look at the [Frequently Asked Questions \(FAQ\)](#) and the [tutorial](#).

Note: To know more about each field simply hover your mouse over it, in some fields the following options are available.

**Calculation label (mandatory)**

**Username (mandatory)**   **Email notification (mandatory)**

**Append User peptides**    **Create the family of peptides**

*Optional selection criteria (a run with unspecified criteria might require long calculation times):*

**Peptide name (empty if indifferent)**

**Sequence pattern**

**Target Organism**  **Activity (µM)**  **Active**

**Advanced (+) Expand Me**

**Peptide properties (+) Expand Me**

**Align method:**  Simple  DSSP  Substitution matrix  **Minimum % of similarity**  **Threshold percentage to group families**

**Simplify aminoacids:**  y  n **Experimentally validated:**  i  y  n **Include only peptides with data about:**

**Run SeqLogo:**  y  n **Generate additional graphical analysis:**  y  n

Results will be viewable in the [DOWNLOAD RESULTS](#) section. You can extract meaningful information from the experiment using [FAMILY ANALYZER](#) tool too.

### Special characters for modified aminoacids

You can click on the symbol to select it and use Ctrl+F to search your desired aminoacid

Parent aminoacid	Modified Aminoacid	Symbol
A,ALA	2-Amino-3-naphthalen-1-ylpropanoic_acid	Ⓛ
A,ALA	2-Naphthylalanine	Ⓛ
A,ALA	3-(2-Thienyl)alanine	Ⓛ
A,ALA	N-Naphthalen-1-yl-L-alanine	Ⓛ
A,ALA	(2S)-2-Amino-3-(4-phenylphenyl)propanoic_acid	Ⓛ
A,ALA	2-(9-Anthracenyl)-L-alanine	Ⓛ
A,ALA	(2R)-2-Amino-2-(aminomethyl)-3-oxobutanoic_acid	Ⓛ

**Experimentally validated:**  i  y  n

**antimicrobial:**  i  y  n

antibacterial:  i  y  n

antigram pos:  i  y  n antigram neg:  i  y  n

antifungal:  i  y  n antiyeast:  i  y  n

antiviral:  i  y  n virus\_name:  activity viral (µM):  activity viral test:

**antibiofilm:**  i  y  n

**antiprotozoal:**  i  y  n

antiparasitic:  i  y  n antiplasmodial:  i  y  n antitrypanosomic:  i  y  n antileishmania:  i  y  n

**insecticidal:**  i  y  n

**anticancer:**  i  y  n

antitumor:  i  y  n cell line:  tissue:  cancer type:

anticancer activity (µM):  anticancer activity test:

antiangiogenic:  i  y  n

**toxic:**  i  y  n

cytotoxic:  i  y  n

hemolytic:  i  y  n hemolytic activity (µM):  hemolytic activity test:  RBC source:

**Sequence length range**

**Stereoisomerism**

**ID**

**source**

**taxonomy**

**Family**

**gene**

**N terminus**

**C terminus**

**PTM**

**target**

**pdb**

**DSSP**

**PMID**

modifications, solubility, etc.). For example, switching on the “Experimental structure” option will filter out all peptides whose structure has not been experimentally obtained.

It is also possible to restrict the selection to those peptides tagged “experimentally validated” by their original source database (“Experimentally validated” toggle). Further restriction is possible by selecting only peptides for which (i) the target organism is described or (ii) the target organism is described and a specific value of activity has been measured.

The user can optionally include their own sequences in the calculation (“Append User peptides”). A remarkable feature of ADAPTABLE is its ability to handle non-proteinogenic amino acids in a standardized fashion. Thanks to a character picker (insert at the top of Fig 1), the user can insert non-proteinogenic amino acids in the sequence by clicking their symbol. The “Simplify amino acids” option transforms non-proteinogenic amino acids to their most similar proteinogenic counterpart, if possible.

Rather than selecting peptides based on their properties, the user can choose to create an SR family using a specific reference sequence (“Create the family of a specific peptide”). This feature can be very useful for classifying new peptides of unknown activity. If very similar sequences are found in the database, this option can even be used to help determine their biological properties. For example, if the peptide introduced by the user has generated an SR family that is 80% antibacterial and 70% anticancer, it can be hypothesized that this peptide potentially has both activities, providing an interesting hypothesis to validate experimentally.

### Choosing the alignment method

ADAPTABLE allows three kinds of alignment methods: “Substitution matrix,” “Simple,” and “DSSP.” The first (default) method uses mutation substitution matrices. The user can choose from among multiple point accepted mutation and BLOSUM matrices or the simpler unitary scoring matrix, with the possibility to edit both the minimum percentage of similarity for the generation of the SR families and the minimum percentage of common peptides to group similar SR families. The “simple” option was introduced to highlight very general properties shared by evolutionary distant peptides (i.e., the presence of amphipathic helices). The aim of the simple mode is to drastically reduce the number of amino acid types, and it will convert any amino acid to one of the nine classes: hydrophobic residues (A, V, I, L, and M) represented by A, negative residues (D and E) by D, positive residues (K and R) by K, aromatic residues (W, Y, H, and F) by F, polar residues (S, T, N, and Q) by S, and modified amino acids by  $\text{\textcircled{M}}$ . Gly, Pro, and Cys are treated individually.

Finally, the “DSSP” option aligns sequences on the basis of the secondary structure to highlight the role of well-defined three-dimensional arrangements. Seven conformations are present in the DSSP notation (Kabsch & Sander, 1983):  $\alpha$ -helix (H), 3–10 helix (G),  $\pi$  helix (I),  $\beta$ -bridge (B),  $\beta$ -strand (E), turn (T), and bend (S).

### Additional graphical analysis

ADAPTABLE optionally computes a series of properties during the generation of SR families and creates multiple visualizations of the results. Besides peptide length distribution, ADAPTABLE computes for each SR family the average presence of proteinogenic amino acids, their average number per peptide, and their percent occurrence at each position. The average presence may highlight the importance of specific amino acid types. For example, positively charged residues are commonly present in AMPs, and they are thought to drive peptides towards negatively charged bacterial membranes; the average number per peptide (e.g., two cysteines) might suggest the presence of specific interactions (e.g., disulphide bonds). The rate of occurrence of each amino acid in each position introduces the spatial information, which is also essential for the design of new active peptides. For example, it has been shown that the distribution of hydrophobic amino acids along the sequence is highly asymmetric in selective AMPs, whereas it tends to be rather constant in hemolytic peptides (which tend to insert deeper in the membrane of the host) (Juretić et al, 2011). A more detailed description about this concept is present in ADAPTABLE Tutorial (Supplemental Data 1).

### Family analyzer

The “Family Analyzer” page provides an overview of the properties of the SR families originated by the “Family Generator.” This tool provides both a summary of the SR family properties (Fig 2, top) and a full output (Fig 2, bottom). The summary page displays statistical analyses of the properties of the SR family and lists each member’s sequence with links to the ADAPTABLE and source database entries. The full output contains an interface, allowing rapid and effective navigation of a large amount of data: for example, it is possible to visualize sequence alignment with color codes for residue conservation (frequency), polarity, amino acid types, or secondary structure. Alternatively, one may choose to visualize only selected information such as the source and the gene of origin of each member of the SR family. A more detailed description of this tool is given in the Tutorial, Supplemental Data 1.

### Downloading results

Results can be downloaded by providing the username and calculation label. Data are available for 6 mo. Once downloaded, the HTML output can be read in a browser.

### ADAPTABLE browsing tool

The ADAPTABLE browsing tool (“Browse AMPs & Families”) is composed of three subsections:

#### Figure 1. Screenshot of the ADAPTABLE “Family Generator” page.

A calculation label and username are provided to allow private access to the results. The user can optionally append new sequences to the calculation. The “Character picker” (insert above) allows the insertion of special characters used for non-proteinogenic amino acids into the sequence fields. The picker offers a direct link to the PubChem (Kim et al, 2019) database for each entry. Expanding the “Advanced” and “Peptide properties” sections (see inserts in the bottom) offers the possibility to choose from among over more than 60 parameters. In the last part of the page, the user can select the desired alignment method, the minimum similarity threshold for the creation of SR families, and the minimum percent overlap among SR families for their clustering. Additional graphical analysis is optionally offered.

## Summaries

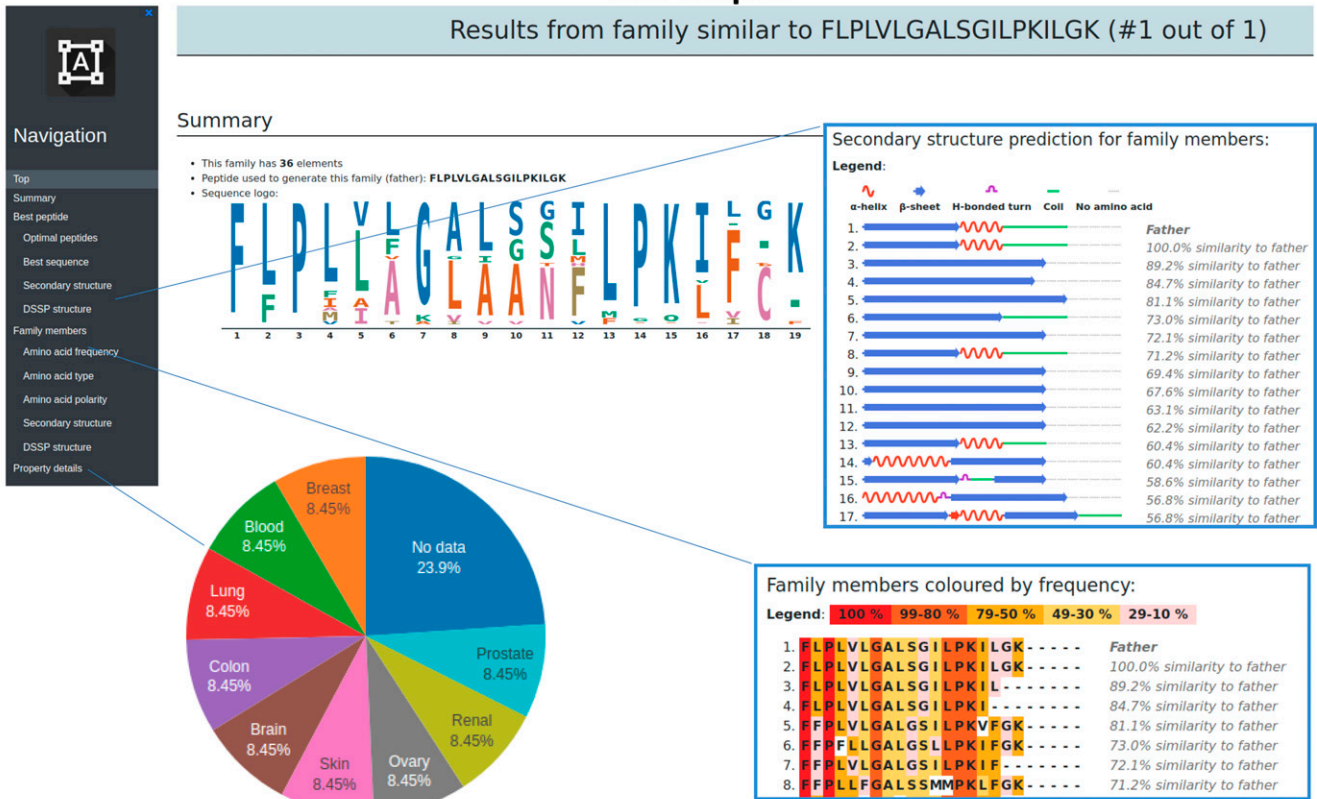
Family 1405 properties		Property	Value
<b>Activity:</b>			
antimicrobial: (data available for the 100 % of members)	ant	antimicrobial: (data available for the 100 % of members)	antimicrobial (100.0-100.0 %)
antibacterial: (data available for the 56.29 % of members)	ant	antibacterial: (data available for the 56.29 % of members)	antibacterial (56.3-100.0 %)
antigram_pos: (data available for the 50.37 % of members)	ant	antigram_pos: (data available for the 50.37 % of members)	antigram_pos (50.4-100.0 %)
antigram_neg: (data available for the 47.40 % of members)	ant	antigram_neg: (data available for the 47.40 % of members)	antigram_neg (47.4-100.0 %)
antifungal: (data available for the 36.29 % of members)	ant	antifungal: (data available for the 36.29 % of members)	antifungal (36.3-100.0 %)
antiyeast: (data available for the 19.25 % of members)	ant	antiyeast: (data available for the 19.25 % of members)	antiyeast (19.3-100.0 %)
antiviral: (data available for the 1.48 % of members)	ant	antiviral: (data available for the 1.48 % of members)	antiviral ( 1.5-100.0 %)
<b>Toxicity:</b>			
toxic: (data available for the 32.59 % of members)	ant	toxic: (32.6-100.0 %)	
cytotoxic: (data available for the 15.55 % of members)	ant	cytotoxic: (15.6-100.0 %)	
hemolytic: (data available for the 20 % of members)	ant	hemolytic: (20.0-100.0 %)	
hemolytic_activity: (data available for the 8.88 % of members)	ant	300 (0.7- 8.3 %); 120 (0.7- 8.3 %); 200 (0.7- 8.3 %)	
hemolytic_activity_test: (data available for the 8.88 % of members)	ant	h50 (3.7-41.7 %); h50 (3.0-33.3 %); h50 (2.2-25.0 %)	
<b>Source:</b>			
name: (data available for the 71.85 % of members)	ant	brevisin (45.9-63.9 %); frogs (50.4-70.1 %); amphibians (50.4-70.1 %); animals (50.4-70.1 %); temporin (13.3-18.6 %); antimicrobial (13.3-18.6 %); precursor (10.4-14.4 %); amolopin (3.7- 5.2 %)	

Members of family 1405		ID
1	.....-FLPLVLGALSGILPKILGK-.....	sa12081 hemo4028
2	.....-FLPLVLGALSGILPKILGK-.....	sa22972 sa17436 DRAMP01807

## Full output

Results from family similar to FLPLVLGALSGILPKILGK (#1 out of 1)



The first one, “Browse AMPs Database,” offers the possibility to view single entries by typing part of a sequence or name; the ADAPTABLE entry consists of a table summarizing all available data on the selected peptide and providing links to the related SR families obtained in the “all\_families” built-in experiment described below (see Screenshot 2 in the Tutorial, Supplemental Data 1). In particular, the first field is a link to the SR family generated by the peptide, whereas the second field contains links to all SR families where the peptide can be found. The field “External database ID” can be used to visit the peptide entries in other databases. When available, biochemical parameters (provided by ExPASy ProtParam tool [Gasteiger et al, 2005]) and the relevant literature are accessible by links. Direct access to the Protein Data Bank is provided for the experimental tridimensional structure, whereas integration with I-TASSER (Roy et al, 2010; Yang et al, 2015; Yang & Zhang, 2015) allows structural prediction even when the experimental one is not available.

The second subsection, “Family overview,” allows finding information regarding the built-in “all\_families” experiment, created by sequence alignment of all the peptides in the database. It constitutes an unrestrained version of the general procedure for the creation of any ADAPTABLE SR family, where no restrictions are applied by the user in terms of peptide properties. As a consequence, each peptide acts as a father for the formation of a family.

The third subsection allows generation of a FASTA file with the sequences of peptides featuring user-defined properties.

### Future Directions

With ADAPTABLE, we want to provide the researcher with a tool to study the mechanism of action of AMPs. Its completeness in terms of database sources permits its applications in various research fields, spanning from medicine to agriculture, food preservation, and antiseptic materials.

We have shown how the generation of SR families is able to group peptides based on user-defined characteristics. ADAPTABLE also has built-in functions to localize conserved residues, thus highlighting motifs (well-definite arrangements of amino acids likely to be responsible for the different activities). The motifs are created by taking into account intrasequence correlations to avoid nonfunctional chimeras. This function is designed to provide optimal scaffolds for drug design. Recent studies suggest that this approach could be very useful for scientists working in this area (Schmitt et al, 2016; Almaaytah et al, 2017). We believe that the architecture of ADAPTABLE, besides offering easy access to all available information on a specific peptide described in many different databases, can also be used by the scientific community to (i) design new peptides using motifs responsible for the specificity towards a specific organism, (ii) predict several properties of a generic sequence, (iii) discover experimentally untested activities

for a given peptide by retrieving information on similar sequences in its SR family, and (iv) generate optimal scaffold for drug design, thanks to the generation of the SR family representing sequence.

ADAPTABLE has been designed as a self-updating platform, thanks to its automated tools that aggregate data from upstream sources. This is a fundamental characteristic because we expect that the number of peptides will continue to increase in the following years. ADAPTABLE will, therefore, continue to incorporate more AMPs and databases. To this end, we have established a simple text input that allows external contributors to translate their data into the ADAPTABLE format, whose structure is shown at <http://gec.u-picardie.fr/adaptable/faq.html#newdb>. We subscribe to a commitment-to-updates that involves being responsible for maintaining the website and updating the resources regularly.

### Additional Files

The ADAPTABLE tutorial file ([http://gec.u-picardie.fr/adaptable/ADAPTABLE\\_tutorial.pdf#view=Fith](http://gec.u-picardie.fr/adaptable/ADAPTABLE_tutorial.pdf#view=Fith)) contains specific case examples guiding the user step-by-step through the calculation and interpretation of results.

## Materials and Methods

### ADAPTABLE AMPs database

ADAPTABLE incorporates automated tools to periodically download, process, and merge data to keep synched with data sources: ADAM (Lee et al, 2015), ANTISTAPHYBASE (Zouhir et al, 2017), APD (Wang & Wang, 2004; Wang et al, 2009; Wang et al, 2016), AVPdb (Qureshi et al, 2014), BaAMPs (Di Luca et al, 2015), BACTIBASE (Hammami et al, 2007, 2010), CAMPR3 (Waghu et al, 2016), CancerPPD (Tyagi et al, 2015), ConoServer (Kaas et al, 2008, 2012), CPPsite (Gautam et al, 2012; Agrawal et al, 2016), DADP (Novković et al, 2012), DBAASP (Gogoladze et al, 2014; Pirtskhalava et al, 2016), Defensins (Seebah et al, 2007), DRAMP (Fan et al, 2016; Kang et al, 2019), Hemolytik (Gautam et al, 2014), HIPdb (Qureshi et al, 2013), InverPep (Gómez et al, 2017), LAMP (Zhao et al, 2013), MilkAMP (Théolier et al, 2014), ParaPep (Mehta et al, 2014), Peptaibol (Whitmore & Wallace, 2004), PhytAMP (Hammami et al, 2009), SATPdb (Singh et al, 2016), UniProt (The UniProt Consortium, 2018), YADAMP (Piotto et al, 2012), PubChem (Kim et al, 2019), The Microbe Directory (Shaaban et al, 2018), and the Protein Data Bank (Berman et al, 2000).

Data are merged in a single database of more than 40,000 unambiguous sequence entries. In the future, even more databases and resources will be implemented to enrich the information available for each sequence. Third-party contributors can develop their own tools to generate a final output adhering to our standardized simple text format (see “FAQ and Tutorial” section of the Supplemental Data 1).

**Figure 2. Screenshots of “Summary” (top) and “Full output” (bottom) for each SR family.**

The Summary (top) provides the minimal percentage of members displaying each property and, for each member, the links to their ADAPTABLE single entry, and to external databases. The full output (bottom) shows the properties in detail for each member and the peptide best representing each family. The SR family members can be visualized sorted by similarity to the father and colored by amino acid frequency, residue type, polarity, predicted secondary structure, or DSSP prediction (see inserts). Multiple pie charts give a visual representation of each distinct property.



“Threshold percentage to group families” in the “FAMILY GENERATOR” webpage).

In particular, the largest SR families in a group are originated by fathers that are compatible with the largest number of entries and, therefore, contain only the essential traits. ADAPTABLE sorts SR families by size so that a larger family index might correspond to increasing activity specificities.

For example, consider SR families 3 and 8: they are in the same group and SR family 3 is originated by a father containing the motif XZ-YYY (where X, Y, and Z are amino acid types) that acts upon biomembranes via a carpet mechanism. The elements of a similar, but smaller, SR family (SR family 8) generated by a father containing the motif XZY-YYY might act via the same mechanism but with preference for Gram-positive bacterial membranes. Such a finding would prompt the researcher to study the two fathers to better understand the molecular-level causes of the increased selectivity.

### Generation of the SR family–representing sequence

Specific arrangement of amino acids (motifs) is often correlated with the local structure (Richardson & Richardson, 1988), the interaction with other partners, or the stability of the peptide (Ashenberg et al, 2013). Although the father constitutes a good representative of each SR family, it might contain parts which are not essential traits of the ensemble. For this reason, ADAPTABLE generates a representative peptide for each SR family of sequences to highlight only the essential motifs, those from which the activity could originate. To preserve the information contained in the residue conservation across the SR family, without losing the information on intramolecular interactions within each sequence, ADAPTABLE generates one single template peptide representing the family, starting from each most abundant residue “aa” at position  $p$ . Each sequence is built by using the most frequent amino acid found at distance  $p'$  within the peptides having “aa” at position  $p$  (if the frequency is below 10%, the position  $p'$  is considered variable and replaced by a dash). Finally, the most-representative peptide is chosen: the one displaying the maximum value of the sum of frequencies at each position  $p$ .

### Motifs (position-dependent and position-independent)

The method for generating the representative sequence (that explained above) is a good tool for highlighting motifs because it reflects the mutual distance of amino acid types within a peptide. The representative sequence contains position-dependent motifs because they are drawn on well-defined positions (e.g., -X-YX-----ZZ-X with X, Y, and Z generic amino acid types, and X-YX and ZZ-X as two independent motifs).

Information of even greater significance derives from the position independence of the method; that is, the occurrence probability of amino acid  $j$  at distance  $d$  from amino acid  $i$  is calculated independently using the position of  $i$  in the sequence. The motifs are then called position-independent, and ADAPTABLE reports their analysis in the graphical output (“Generate additional graphical analysis” option set to “y” in the “Family Generator” page).

### Property calculations

ADAPTABLE features a simple algorithm to predict solubility in water based on the number of hydrophobic (I, L, M, F, V, W, and Y) relative to charged (K, R, H, D, and E) residues. Its prediction outputs are soluble (less than 5 amino acids or <50% hydrophobic and >25% charged), poorly soluble (<50% hydrophobic and ≤25% charged), almost insoluble (50–75% hydrophobic), and insoluble (hydrophobic ≥75%).

It also performs a prediction of the secondary structure based on the Chou–Fasman method (Chou & Fasman, 1974, 1978, 1979).

### Additional graphical analysis

Besides the analysis of motifs and of the SR family–representative peptide, other types of graphical outputs can be created by the “Family Generator” tool (“Generate additional graphical analysis” option set to “y”) that reports on the amino acid composition and position in the sequence.

The sequence logo generation based on WebLogo (Schneider & Stephens, 1990; Crooks, 2004) can be performed when the “Run SeqLogo” option is set to “y” in the “Family Generator” page. Examples are given in the Tutorial (Supplemental Data 1).

### Implementation

ADAPTABLE is developed using mainly GAWK (version 4.1 or newer at the time of the writing) (Robbins, 2003) and Bash (4.3 or newer) (Ramey & Fox, 2015). Graphics generation relies on Matplotlib (2.2 or newer) (Hunter, 2007) and Python (3.5 or newer) (Van Rossum & Drake, 2011). Parallelization is achieved using GNU parallel (Tange, 2011) and Python Joblib. The web interface relies on Javascript and CSS3 and adheres to HTML5 web standard, running Apache 2.4 server on a Linux system.

### Availability of Supporting Source Code and Requirements

- Project name: ADAPTABLE.
- Project home page: <http://gec.u-picardie.fr/adaptable/>.
  - <https://gitlab.com/pachoramos/ADAPTABLE>.
- Operating system(s): Platform independent (client runs in the web browser).
- Programming language: AWK, Bash, Python, HTML5, and Javascript.
- Other requirements: Web browser adhering to HTML5 standards (i.e., Chrome, Firefox, Safari, and Edge).
- License: EUPL-1.2.

## Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lisa.201900512>.

## Acknowledgements

We would like to thank Professor Manuel Dauchez, University of Reims Champagne-Ardenne, Matrice Extra-cellulaire et Dynamique Cellulaire (MEDYC),



Unité mixte de recherche 7369, Centre national de la recherche scientifique for useful discussion, and the Matrics platform at the University "Picardie Jules Verne" for providing computing resources. We would also like to thank Graham Bentham for English language editing. Funding: This work was partly supported by the Université de Picardie Jules Verne, S2R 2018 - Action 1: "Incitations au dépôt de projets de recherche"; Francisco Ramos-Martín's PhD scholarship was co-funded by Conseil régional des Hauts-de-France and by European Fund for Economic and Regional Development (FEDER). The authors are grateful to the Université de Picardie Jules Verne for its financial support for publication through its S2R action.

### Author Contributions

F Ramos-Martín: conceptualization, data curation, software, validation, visualization, methodology, and writing—original draft.

T Annaval: data curation, software, methodology, and writing—review and editing.

S Buchoux: software, visualization, methodology, and writing—review and editing.

C Sarazin: conceptualization and writing—review and editing.

N D'Amelio: conceptualization, data curation, software, formal analysis, supervision, funding acquisition, validation, project administration, and writing—original draft, review, and editing.

### Conflict of Interest Statement

The authors declare that they have no conflict of interest.

## References

- Agrawal P, Bhalla S, Usmani SS, Singh S, Chaudhary K, Raghava GPS, Gautam A (2016) CPPsite 2.0: A repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res* 44: D1098–D1103. doi:10.1093/nar/gkv1266
- Aguilera-Mendoza L, Marrero-Ponce Y, Tellez-Ibarra R, Llorente-Quesada MT, Salgado J, Barigye SJ, Liu J (2015) Overlap and diversity in antimicrobial peptide databases: Compiling a non-redundant set of sequences. *Bioinformatics* 31: 2553–2559. doi:10.1093/bioinformatics/btv180
- Aisenbrey C, Marquette A, Bechinger B (2019) The mechanisms of action of cationic antimicrobial peptides refined by novel concepts from biophysical investigations. *Adv Exp Med Biol* 1117: 33–64. doi:10.1007/978-981-13-3588-4\_4
- Albiol Matanic VC, Castilla V (2004) Antiviral activity of antimicrobial cationic peptides against Junin virus and herpes simplex virus. *Int J Antimicrob Agents* 23: 382–389. doi:10.1016/j.ijantimicag.2003.07.022
- Almaaytah A, Ajingi Y, Abualhajjaa A, Tarazi S, Alshar'i N, Al-Balas Q (2017) Peptide consensus sequence determination for the enhancement of the antimicrobial activity and selectivity of antimicrobial peptides. *Infect Drug Resist* 10: 1–17. doi:10.2147/IDR.S118877
- Ashenberg O, Ian Gong L, Bloom JD (2013) Mutational effects on stability are largely conserved during protein evolution. *Proc Natl Acad Sci U S A* 110: 21071–21076. doi:10.1073/pnas.1314781111
- Bechinger B (1996) Towards membrane protein design: pH-sensitive topology of histidine-containing polypeptides. *J Mol Biol* 263: 768–775. doi:10.1006/jmbi.1996.0614
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28: 235–242. doi:10.1093/nar/28.1.235
- Bhadra P, Yan J, Li J, Fong S, Siu SWI (2018) AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. *Sci Rep* 8: 1697. doi:10.1038/s41598-018-19752-w
- Brahmachary M, Krishnan SP, Koh JL, Khan AM, Seah SH, Tan TW, Brusica V, Bajic VB (2004) ANTIMIC: A database of antimicrobial sequences. *Nucleic Acids Res* 32: 586–589. doi:10.1093/nar/gkh032
- Chou PY, Fasman GD (1974) Prediction of protein conformation. *Biochemistry* 13: 222–245. doi:10.1021/bi00699a002
- Chou PY, Fasman GD (1978) Empirical predictions of protein conformation. *Annu Rev Biochem* 47: 251–276. doi:10.1146/annurev.bi.47.070178.001343
- Chou PY, Fasman GD (1979) Prediction of the secondary structure of proteins from their amino acid sequence: Meister/advances. In *Advances in Enzymology and Related Areas of Molecular Biology*, Meister A (ed), pp 45–148. Hoboken, NJ: John Wiley & Sons, Inc. Available at: <http://doi.wiley.com/10.1002/9780470122921.ch2>.
- Crooks GE (2004) WebLogo: A sequence logo generator. *Genome Res* 14: 1188–1190. doi:10.1101/gr.849004
- Dayhoff MO, Barker WC, Hunt LT (1983) Establishing homologies in protein sequences. *Methods Enzymol* 91: 524–545. doi:10.1016/s0076-6879(83)91049-2
- Di Luca M, Maccari G, Maisetta G, Batoni G (2015) BaAMPs: The database of biofilm-active antimicrobial peptides. *Biofouling* 31: 193–199. doi:10.1080/08927014.2015.1021340
- Fan L, Sun J, Zhou M, Zhou J, Lao X, Zheng H, Xu H (2016) DRAMP: A comprehensive data repository of antimicrobial peptides. *Sci Rep* 6: 24482. doi:10.1038/srep24482
- Field D, Des F, Cotter PD, Colin H, Ross RP (2015) Bioengineering lantibiotics for therapeutic success. *Front Microbiol* 6: 1363. doi:10.3389/fmicb.2015.01363
- Fjell CD, Hancock REW, Cherkasov A (2007) AMPper: A database and an automated discovery tool for antimicrobial peptides. *Bioinformatics* 23: 1148–1155. doi:10.1093/bioinformatics/btm068
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. *The Proteomics Protocols Handbook*, pp 571–607. Available at: doi:10.1385/1-59259-890-0:571.
- Gautam A, Chaudhary K, Singh S, Joshi A, Anand P, Tuknait A, Mathur D, Varshney GC, Raghava GPS (2014) Hemolytik: A database of experimentally determined hemolytic and non-hemolytic peptides. *Nucleic Acids Res* 42: D444–D449. doi:10.1093/nar/gkt1008
- Gautam A, Singh H, Tyagi A, Chaudhary K, Kumar R, Kapoor P, Raghava GPS (2012) CPPsite: A curated database of cell penetrating peptides. *Database* 2012: bas015. doi:10.1093/database/bas015
- Giegerich R, Wheeler D (1996) Pairwise sequence alignment. In *BioComputing Hypertext Coursebook*, pp 1–6. Available at: <https://pdfs.semanticscholar.org/5d74/9c057d37f8ffc6a46ec52e347b6b0598f4f4.pdf>.
- Gogoladze G, Grigolava M, Vishnepolsky B, Chubinidze M, Duroux P, Lefranc M-P, Pirtskhalava M (2014) DBAASP: Database of antimicrobial activity and structure of peptides. *FEMS Microbiol Lett* 357: 63–68. doi:10.1111/1574-6968.12489
- Gómez EA, Giraldo P, Orduz S (2017) InverPep: A database of invertebrate antimicrobial peptides. *J Glob Antimicrob Resist* 8: 13–17. doi:10.1016/j.jgar.2016.10.003
- Goyal RK, Mattoo AK (2016) Plant antimicrobial peptides. In *Host Defense Peptides and Their Potential as Therapeutic Agents*, 111–136. Available at: doi:10.1007/978-3-319-32949-9\_5.
- Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS Open Source Drug Discovery Consortium, (2013) In silico approach for predicting toxicity of peptides and proteins. *PLoS One* 8: e73957. doi:10.1371/journal.pone.0073957

- Hammami R, Ben Hamida J, Vergoten G, Fliss I (2009) PhytAMP: A database dedicated to antimicrobial plant peptides. *Nucleic Acids Res* 37: D963–D968. doi:10.1093/nar/gkn655
- Hammami R, Zouhir A, Ben Hamida J, Fliss I (2007) BACTIBASE: A new web-accessible database for bacteriocin characterization. *BMC Microbiol* 7: 89. doi:10.1186/1471-2180-7-89
- Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I (2010) BACTIBASE second release: A database and tool platform for bacteriocin characterization. *BMC Microbiol* 10: 22. doi:10.1186/1471-2180-10-22
- Hoskin DW, Ayyalusamy R (2008) Studies on anticancer activities of antimicrobial peptides. *Biochim Biophys Acta* 1778: 357–375. doi:10.1016/j.bbame.2007.11.008
- Huang Y, Huang J, Chen Y (2010) Alpha-helical cationic antimicrobial peptides: Relationships of structure and function. *Protein Cell* 1: 143–152. doi:10.1007/s13238-010-0004-3
- Hunter JD (2007) Matplotlib: A 2D graphics environment. *Comput Sci Eng* 9: 90–95. doi:10.1109/mcse.2007.55
- Joo NE, Ritchie K, Kamarajan P, Miao D, Kapila YL (2012) Nisin, an opoptogenic bacteriocin and food preservative, attenuates HNSCC tumorigenesis via CHAC1. *Cancer Med* 1: 295–305. doi:10.1002/cam4.35
- Juretić D, Vukičević D, Petrov D, Novković M, Bojović V, Lučić B, Ilić N, Tossi A (2011) Knowledge-based computational methods for identifying or designing novel, non-homologous antimicrobial peptides. *Eur Biophys J* 40: 371–385. doi:10.1007/s00249-011-0674-7
- Kaas Q, Westermann J-C, Halai R, Wang CKL, Craik DJ (2008) ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* 24: 445–446. doi:10.1093/bioinformatics/btm596
- Kaas Q, Yu R, Jin A-H, Dutertre S, Craik DJ (2012) ConoServer: Updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res* 40: D325–D330. doi:10.1093/nar/gkr886
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22: 2577–2637. doi:10.1002/bip.360221211
- Kang X, Dong F, Shi C, Liu S, Sun J, Chen J, Li H, Xu H, Lao X, Zheng H (2019) DRAMP 2.0, an updated data repository of antimicrobial peptides. *Sci Data* 6: 148. doi:10.1038/s41597-019-0154-y
- Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al (2019) PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res* 47: D1102–D1109. doi:10.1093/nar/gky1033
- Lacerda AF, Vasconcelos EA, Pelegrini PB, Grossi de Sa MF (2014) Antifungal defensins and their role in plant defense. *Front Microbiol* 5: 116. doi:10.3389/fmicb.2014.00116
- Lee H-T, Lee C-C, Yang J-R, Lai JZC, Chang KY (2015) A large-scale structural classification of antimicrobial peptides. *Biomed Res Int* 2015: 475062. doi:10.1155/2015/475062
- Mehta D, Anand P, Kumar V, Joshi A, Mathur D, Singh S, Tuknait A, Chaudhary K, Gautam SK, Gautam A, et al (2014) ParaPep: A web resource for experimentally validated antiparasitic peptide sequences and their structures. *Database* 2014: bau051. doi:10.1093/database/bau051
- Myers EW (1991) *An Overview of Sequence Comparison Algorithms in Molecular Biology*. Tucson, AZ: Department of Computer Science, University of Arizona. Available at: <http://myerslab.mpi-cbg.de/wp-content/uploads/2014/06/compbio.survey.pdf>.
- Ng XY, Rosdi BA, Shahrudin S (2015) Prediction of antimicrobial peptides based on sequence alignment and support vector machine-pairwise algorithm utilizing LZ-complexity. *Biomed Res Int* 2015: 212715. doi:10.1155/2015/212715
- Novković M, Simunić J, Bojović V, Tossi A, Juretić D (2012) DADP: The database of anuran defense peptides. *Bioinformatics* 28: 1406–1407. doi:10.1093/bioinformatics/bts141
- Oren Z, Shai Y (1998) Mode of action of linear amphipathic  $\alpha$ -helical antimicrobial peptides. *Biopolymers* 47: 451–463. doi:10.1002/(SICI)1097-0282(1998)47:6<451::AID-BIP4>3.0.CO;2-F
- Oyinloye BE, Adenowo AF, Kappo AP (2015) Reactive oxygen species, apoptosis, antimicrobial peptides and human inflammatory diseases. *Pharmaceuticals* 8: 151–175. doi:10.3390/ph8020151
- Piotto SP, Sessa L, Concilio S, Iannelli P (2012) YADAMP: Yet another database of antimicrobial peptides. *Int J Antimicrob Agents* 39: 346–351. doi:10.1016/j.ijantimicag.2011.12.003
- Pirtskhalava M, Gabrielian A, Cruz P, Griggs HL, Squires RB, Hurt DE, Grigolava M, Chubinidze M, Gogoladze G, Vishnepolsky B, et al (2016) DBAASP v.2: An enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res* 44: 6503. doi:10.1093/nar/gkw243
- Qureshi A, Thakur N, Kumar M (2013) HIPdb: A database of experimentally validated HIV inhibiting peptides. *PLoS One* 8: e54908. doi:10.1371/journal.pone.0054908
- Qureshi A, Thakur N, Tandon H, Kumar M (2014) AVPdb: A database of experimentally validated antiviral peptides targeting medically important viruses. *Nucleic Acids Res* 42: D1147–D1153. doi:10.1093/nar/gkt1191
- Ramey C, Fox B (2015) *Bash 4.3 Reference Manual*. Samurai Media Limited. Available at: [https://books.google.com/books/about/Bash\\_4\\_3\\_Reference\\_Manual.html?hl=&id=Ddn1jgEACAAJ](https://books.google.com/books/about/Bash_4_3_Reference_Manual.html?hl=&id=Ddn1jgEACAAJ).
- Richardson JS, Richardson DC (1988) Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240: 1648–1652. doi:10.1126/science.3381086
- Robbins A (2003) *GAWK: Effective AWK Programming: A User's Guide for GNU Awk*. Available at: <https://books.google.com/books/about/GAWK.html?hl=&id=ADJjtWAAACAAJ>.
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5: 725–738. doi:10.1038/nprot.2010.5
- Schmitt P, Rosa RD, Destoumieux-Garzón D (2016) An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. *Biochim Biophys Acta* 1858: 958–970. doi:10.1016/j.bbame.2015.10.011
- Schneider TD, Stephens RM (1990) Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* 18: 6097–6100. doi:10.1093/nar/18.20.6097
- Seebah S, Suresh A, Zhuo S, Choong YH, Chua H, Chuon D, Beuerman R, Verma C (2007) Defensins knowledgebase: A manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic Acids Res* 35: D265–D268. doi:10.1093/nar/gkl866
- Shaaban H, Westfall DA, Mohammad R, Danko D, Bezdán D, Afshinnekoo E, Segata N, Mason CE (2018) The Microbe Directory: An annotated, searchable inventory of microbes' characteristics. *Gates Open Res* 2: 3. doi:10.12688/gatesopenres.12772.1
- Singh S, Chaudhary K, Dhanda SK, Bhalla S, Usmani SS, Gautam A, Tuknait A, Agrawal P, Mathur D, Raghava GPS (2016) SATPdb: A database of structurally annotated therapeutic peptides. *Nucleic Acids Res* 44: D1119–D1126. doi:10.1093/nar/gkv1114
- Tamamura H, Xu Y, Hattori T, Zhang X, Arakaki R, Kanbara K, Omagari A, Otaka A, Ibuka T, Yamamoto N, et al (1998) A low-molecular-weight inhibitor against the chemokine receptor CXCR4: A strong anti-HIV peptide T140. *Biochem Biophys Res Commun* 253: 877–882. doi:10.1006/bbrc.1998.9871
- Tange O (2011) GNU parallel: The command-line power tool. *login: The USENIX Magazine* 36: 42–47. doi:10.5281/zenodo.16303
- Tassanakajon A, Somboonwiwat K, Amparyup P (2015) Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Dev Comp Immunol* 48: 324–341. doi:10.1016/j.dci.2014.05.020

- Thakur N, Qureshi A, Kumar M (2012) AVPPred: Collection and prediction of highly effective antiviral peptides. *Nucleic Acids Res* 40: W199–W204. doi:10.1093/nar/gks450
- Théolier J, Fliss I, Jean J, Hammami R (2014) MilkAMP: A comprehensive database of antimicrobial peptides of dairy origin. *Dairy Science & Technology* 94: 181–193. doi:10.1007/s13594-013-0153-2
- Tyagi A, Tuknait A, Anand P, Gupta S, Sharma M, Mathur D, Joshi A, Singh S, Gautam A, Raghava GPS (2015) CancerPPD: A database of anticancer peptides and proteins. *Nucleic Acids Res* 43: D837–D843. doi:10.1093/nar/gku892
- The UniProt Consortium (2018) UniProt: The universal protein knowledgebase. *Nucleic Acids Res* 46: 2699. doi:10.1093/nar/gky092
- Van Rossum G, Drake FL, Jr (2011) *The Python Language Reference Manual Network Theory*. Available at: [https://books.google.com/books/about/The\\_Python\\_Language\\_Reference\\_Manual.html?hl=&id=Ut4BuQAACAAJ](https://books.google.com/books/about/The_Python_Language_Reference_Manual.html?hl=&id=Ut4BuQAACAAJ).
- Vogel G (2017) Meet WHO's dirty dozen: The 12 bacteria for which new drugs are most urgently needed. *Science* doi:10.1126/science.aal0829
- Vogt TC, Bechinger B (1999) The interactions of histidine-containing amphipathic helical peptide antibiotics with lipid bilayers. The effects of charges and pH. *J Biol Chem* 274: 29115–29121. doi:10.1074/jbc.274.41.29115
- Waghu FH, Barai RS, Gurung P, Idicula-Thomas S (2016) CAMPR3: A database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res* 44: D1094–D1097. doi:10.1093/nar/gkv1051
- Wang G (2014) Improved methods for classification, prediction, and design of antimicrobial peptides. *Methods Mol Biol* 1268: 43–66. doi:10.1007/978-1-4939-2285-7\_3
- Wang G, Li X, Wang Z (2009) APD2: The updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res* 37: D933–D937. doi:10.1093/nar/gkn823
- Wang G, Li X, Wang Z (2016) APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res* 44: D1087–D1093. doi:10.1093/nar/gkv1278
- Wang G, Mishra B, Lau K, Lushnikova T, Golla R, Wang X (2015) Antimicrobial peptides in 2014. *Pharmaceuticals* 8: 123–150. doi:10.3390/ph8010123
- Wang L, Dong C, Li X, Han W, Su X (2017) Anticancer potential of bioactive peptides from animal sources. *Oncol Rep* 38: 637–651. doi:10.3892/or.2017.5778
- Wang P, Hu L, Liu G, Jiang N, Chen X, Xu J, Zheng W, Li L, Tan M, Chen Z, et al (2011) Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. *PLoS One* 6: e18476. doi:10.1371/journal.pone.0018476
- Wang Z, Wang G (2004) APD: The antimicrobial peptide database. *Nucleic Acids Res* 32: D590–D592. doi:10.1093/nar/gnh077
- Whitmore L, Wallace BA (2004) The Peptaibol Database: A database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res* 32: D593–D594. doi:10.1093/nar/gkh371
- Xu X, Lai R (2015) The chemistry and biological activities of peptides from amphibian skin secretions. *Chem Rev* 115: 1760–1846. doi:10.1021/cr4006704
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: Protein structure and function prediction. *Nat Methods* 12: 7–8. doi:10.1038/nmeth.3213
- Yang J, Zhang Y (2015) I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res* 43: W174–W181. doi:10.1093/nar/gkv342
- Yang SC, Lin CH, Sung CT, Fang JY (2014) Antibacterial activities of bacteriocins: Application in foods and pharmaceuticals. *Front Microbiol* 5: 241. doi:10.3389/fmicb.2014.00241
- Yan R, Xu D, Yang J, Walker S, Zhang Y (2013) A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. *Sci Rep* 3: 2619. doi:10.1038/srep02619
- Zasloff M (2002) Antimicrobial peptides of multicellular organisms. *Nature* 415: 389–395. doi:10.1038/415389a
- Zhao X, Wu H, Lu H, Li G, Huang Q (2013) Lamp: A database linking antimicrobial peptides. *PLoS One* 8: e66557. doi:10.1371/journal.pone.0066557
- Zouhir A, Hammami R, Fliss I, Hamida JB (2010) A new structure-based classification of gram-positive bacteriocins. *Protein J* 29: 432–439. doi:10.1007/s10930-010-9270-4
- Zouhir A, Taieb M, Lamine MA, Cherif A, Jridi T, Mahjoubi B, Mbarek S, Fliss I, Nefzi A, Sebei K, et al (2017) ANTISTAPHYBASE: Database of antimicrobial peptides (AMPs) and essential oils (EOs) against methicillin-resistant *Staphylococcus aureus* (MRSA) and *Staphylococcus aureus*. *Arch Microbiol* 199: 215–222. doi:10.1007/s00203-016-1293-6



**License:** This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).