RESEARCH LETTER

DBAASP: database of antimicrobial activity and structure of peptides

Giorgi Gogoladze[1], Maia Grigolava[1], Boris Vishnepolsky[1], Mindia Chubinidze[1], Patrice Duroux[2], Marie-Paule Lefranc[2] & Malak Pirtskhalava[1]

[1]Laboratory of Bioinformatics, Ivane Beritashvili Center of Experimental Biomedicine, Tbilisi, Georgia; and [2]IMGT®, the international ImMunoGeneTics Information System®, Institut de Génétique Humaine, Université Montpellier 2, Montpellier, France

Correspondence: Malak Pirtskhalava, Laboratory of Bioinformatics, Ivane Beritashvili Center of Experimental Biomedicine, 14 Gotua Str., Tbilisi 0160, Georgia. Tel.: +995 574 16 23 97; fax: +995 32 237 34 11; e-mail: m.pirtskhalava@lifescience.org.ge

Received 5 February 2014; revised 10 April 2014; accepted 22 May 2014. Final version published online 10 July 2014.

DOI: 10.1111/1574-6968.12489

Editor: Michael Galperin

Keywords
antimicrobial peptide; database; antimicrobial activity; haemolytic/cytotoxic activity.

Abstract

The Database of Antimicrobial Activity and Structure of Peptides (DBAASP) is a manually curated database for those peptides for which antimicrobial activity against particular targets has been evaluated experimentally. The database is a depository of complete information on: the chemical structure of peptides; target species; target object of cell; peptide antimicrobial/haemolytic/cytotoxic activities; and experimental conditions at which activities were estimated. The DBAASP search page allows the user to search peptides according to their structural characteristics, complexity type (monomer, dimer and two-peptide), source, synthesis type (ribosomal, nonribosomal and synthetic) and target species. The database prediction algorithm provides a tool for rational design of new antimicrobial peptides. DBAASP is accessible at http://www.biomedicine.org.ge/dbaasp/.

Introduction

Antimicrobial peptides (AMPs) make up the innate immune defence system against microbial infections. They are found in different kingdoms of life, including humans (Wang, 2008). These peptides are broad-spectrum antibiotic substances that can be used as therapeutic agents for the treatment of diseases caused by bacteria, fungi, viruses, parasites and cancerous cells. AMPs belong to three classes according to their origin. The first class is natural ribosomally synthesized peptides present in all organisms (Brogden, 2005). More than 1000 of such peptides have been identified. The second class is natural nonribosomally synthesized peptides produced in bacteria (Caboche *et al.*, 2010). The third class is non-natural synthetic peptides.

Many AMPs have no regular secondary structure in solution and acquire functional form after interaction with the lipid bilayer. Peptide amphipathicity is required for antimicrobial activity, with hydrophilic and hydrophobic amino acids being aligned on separate sides of the structured peptide. According to their primary and secondary structures, AMPs can be divided into: (1) linear α-helical peptides, (2) peptides with a β-sheet, (3) peptides with a β-hairpin and (4) peptides rich in particular amino acids (Reddy *et al.*, 2004).

Due to the increase of microorganism resistance to conventional antibiotics, new agents are needed for therapeutic use. Artificial AMPs have been considered as potential drugs against infectious diseases. To develop such drugs, knowledge of the relationships between peptide structure and antimicrobial activity is necessary. To understand the physicochemical parameters that are responsible for antimicrobial activity and high therapeutic index, peptide structure–activity relationships need to be studied. Full information on the chemical structure and activity of peptides is needed to carry out such study. Information about the target object of the cell – lipid bilayer, membrane protein, cytoplasmic protein, DNA and RNA – is also required. Information on the detailed chemical structure of peptides, and their antimicrobial/haemolytic/cytotoxic activities, is scattered among different databases and not complete. Prediction tools in the

FEMS MICROBIOLOGY LETTERS

databases are mainly based on machine learning algorithms. Prediction by machine learning requires the submission of both positive and negative samples in the training and testing datasets. There are no databases that allow the user to form an experimentally validated negative (non-AMP) set for machine learning.

The main goal of this work was to create a database which overcomes these lacks. We have implemented the 'Database of Antimicrobial Activity and Structure of Peptides' (DBAASP), which contains information on AMPs of different origins (ribosomal, nonribosomal and synthetic) and complexity (monomers, dimers and two-peptides). DBAASP is manually curated and is a depository of information on those peptides for which antimicrobial activity against particular targets has been evaluated experimentally. The database provides:

(1) Full information on the chemical structure of peptides: complete information regarding post-translational and N/C termini modification of amino acids.

(2) Information about peptide antimicrobial activities and experimental conditions at which activities were estimated.

(3) Information about peptide haemolytic/cytotoxic activities.

(4) Information about the target object of the cell.

DBAASP provides the development of machine-learning-based AMP prediction tools for particular target species that will operate by experimentally validated positive (AMP) and negative (non-AMP) sets of peptides.

Database content

Structure of the database

DBAASP is hosted on a Linux server using JBOSS 7 application server (http://www.jboss.org). All entries are stored in a MYSQL 5.5 (http://www.mysql.com/) database. The application is written in JAVA 7 (http://www.oracle.com), using PRIMEFACES, an open source component library for JavaServer Faces.

Data collection

Information about AMPs was collected from PubMed (http://www.ncbi.nlm.nih.gov/pubmed) using keywords: antimicrobial, antibacterial, antifungal, antiviral, antitumour, anticancer and antiparasitic peptides.

Web interface

The pages that give access to the database sections are as follows: (1) 'Home' is the introductory page, (2) 'Search' provides a link to the page for general search of the peptides, (3) 'Ranking Search' gives access to the search page, which allows users to find peptides by target species and activity measure and gives result in the form of a ranking list, (4) 'Prediction' provides a link to the prediction tool, (5) 'About' gives access to the page describing the goal behind the design of the database, and (6) 'Help' provides a description of the search abilities and abbreviations used.

Information stored in DSAASP

Characterization of antimicrobial peptides in DSAASP

Knowledge on AMPs is presented according to the unique ontology for immunogenetics and immunoinformatics, IMGT-ONTOLOGY, using its concepts of Identification, Classification and Description (Giudicelli & Lefranc, 2012).

Identification: peptide and target summary

Identification holds: (1) name of peptide; (2) synthesis type (ribosomal, nonribosomal, synthetic); (3) complexity (monomer, dimer, two-peptide); (4) target group (Gram-positive, Gram-negative, virus, parasite, insect, cancer, fungus, mammalian cell and mycoplasma); and (5) target object of cell (membrane protein, cytoplasmic protein, lipid bilayer and DNA/RNA).

Classification: peptide source

Classification involves: (1) kingdom of source species (We use Cavalier-Smith's six-kingdom system (Cavalier-Smith, 2004)), and (2) source species.

Description: information about sequence

Description includes: (1) sequence (amino acids with an L-isomer or without a stereoisomer are denoted by uppercase letters, D-stereoisomers are represented by lowercase letters); (2) length of peptide; and (3) unusual and post-translationally modified amino acids (unusual or post-translationally modified amino acids are denoted by an X or x).

Antimicrobial activity against the target species

Antimicrobial activity against the target species comprises information on: target species, activity measure, activity value and unit. A brief description of experimental conditions is also provided.

Haemolytic and cytotoxic activity

For medical purposes, an AMP should have high antimicrobial and low or no haemolytic/cytotoxic activity. For

activity tests, erythrocytes from different species are used most frequently. 'Haemolytic and cytotoxic activity' information comprises: target cell, activity measure for lysis, peptide concentration and unit.

Organization of information in DBAASP peptide card

Full information on each peptide is presented in the peptide card. An example of the peptide card is given in Data S1. There is 'Peptide and Target Summary' information at the top followed by 'Peptide Source' information with a link to NCBI. 'Information about Sequence' is presented in several tables, including: (1) the table 'Monomer' describes sequence termini modification and length of the peptide; (2) the table 'Intrachain Bond' holds information about the position of amino acids involved in intrachain bond and bond type; and (3) the table 'Unusual or Modified Amino Acid' gives the position and type of unusual amino acids and amino acids before modification. Peptide activity information is presented in two tables: 'Activity Against Target Species' and 'Haemolytic and Cytotoxic Activity'. Additional information about the experiment or target is given in the 'Note'. There are references to the articles at the bottom of the card.

Database statistics

Statistical data relating to the database on 10 April 2014 are presented in the following tables: (1) the number of peptides by complexity (Table 1); and (2) the number of monomers by synthesis type, bond type and amino acid modification (Table 2). Additional information on monomer length distribution is given in Fig. 1.

Utilities

Search

Peptides can be searched via the 'Search' page. The page is divided into four sections for easier query construction (Data S2). The result table provides a list of peptides which correspond to the search criteria of the 'Search' page. For each peptide are displayed the ID, Name, N Terminus, Sequence and C Terminus (Data S3). 'View' provides a link to the individual 'DBAASP peptide card' (see above).

Table 1. Number of peptides by complexity

Peptide complexity type	No. of peptides
Monomers	3917
Dimers	27
Two-peptides	110
Total number of peptides	4054

Table 2. Number of monomers by synthesis type, bond type and amino acid modification

Monomer type	No. of monomers
Ribosomal	1382
Nonribosomal	3
Synthetic	2532
With disulfide bond (DSB)	746
With N–C termini peptide bond (NCB)	194
With thioether bond (TIE)	5
With sidechain–mainchain bond (SMB)	2
Without intrachain bond	3021
With modified N terminus	253
With modified C terminus	1722
With modified N and C termini	226
Without N and C termini modification	2168
Peptides with modified side chain	271
Monomers with D amino acid	246

Ranking search

The 'Ranking Search' page allows users to find peptides by target species and activity measure in combination with other search options. The result table provides a list of peptides ranked by activity value for given target species and activity measure (Data S4). The database will soon have a ranking search for healthy cells and lysis measure.

Prediction tool

Most of the available AMP prediction methods use a common approach for different classes of AMPs. In contrast, we propose that a strategy of prediction should be based on the fact that there are several kinds of AMPs. In the current version of DBAASP we focus on the prediction of a particular (largest) class of AMP, the linear cationic antimicrobial peptides (LCAP). The method is based on the analysis of the physicochemical characteristics responsible for the ability of a peptide to interact with an anionic membrane. Characteristics such as hydrophobicity, amphiphaticity, location of the peptide in relation to the membrane, charge and propensity to have a disordered structure were studied. A detailed description of the algorithm is given by Vishnepolsky & Pirtskhalava (2014).

Comparison with other databases

Other AMP databases can be divided into specialized and general. The former include: PhytAMP, dedicated to plant AMPs (Hammami *et al.*, 2009); Bactibase, a database for bacteriocin peptides (Hammami *et al.*, 2007); PenBase, which provides information about penaeidins (Gueguen

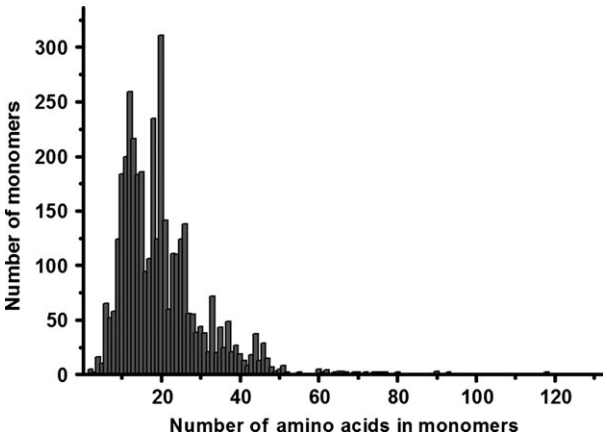

Fig. 1. Distribution of the number of monomers according to monomer length.

et al., 2006); RAPD, a database of recombinantly produced peptides (Li & Chen, 2008); Defensins knowledgebase, which contains information about defensins (Seebah *et al.*, 2007); Peptaibols, a database dedicated to peptaibols (Whitmore & Wallace, 2004); Norine, a database of nonribosomal peptides (Caboche *et al.*, 2008); DADP, a database of anuran peptides (Novkovic *et al.*, 2012); and DAMPD, which contains ribosomal peptides (Sundararajan *et al.*, 2012).

Antimicrobial peptide databases of general type are characterized by a high number of entries and broad range of peptide origin. They include: YADAMP, which contains 2525 sequences (Piotto *et al.*, 2012); APD, which has 2338 sequences (Wang *et al.*, 2009); CAMP, which holds 6756 sequences (Waghu *et al.*, 2014); and LAMP, which comprises 5547 sequences (http://biotechlab.fudan.edu.cn/database/lamp/index. php).

DBAASP contains more than 4000 experimentally determined ribosomal, nonribosomal and synthetic peptides. Thus, DBAASP can be included in the group of general AMP databases. Comparisons of general peptide databases are presented in Tables 3 and 4.

Detailed chemical structure information is presented in DBAASP only, although information on target object of the cell is also presented in APD. In DBAASP all entries include: (1) detailed chemical structure of the peptide, (2) antimicrobial activity against the target species and (3) haemolytic/cytotoxic activity; this information is either incomplete or does not exist in the other databases. There are also differences in the search systems. Searching by sequence 'Length' can is also provided by YADAMP. Few search systems allow the user to find peptides by 'N Terminus' modification and 'Synthesis Type'. Only APD and DBAASP provide the possibility of finding peptides by 'C terminus' modification, 'Unusual Amino Acid', 'Intrachain Bond', 'Complexity' and 'Target Object of Cell'. DBAASP allows the user to perform ranking searches by a particular target species and measure of activity. The result of these searches is a ranking list of selected peptides based on activity values. DBAASP

Table 3. Comparison of information available in databases of general type

	Available information					
Database	No. of entries	Origin of peptides	Detailed chemical structure	Target object of cell	Activity against target species	Haemolytic/cytotoxic activity
DBAASP	> 4000	N, S	+	+	+	+
YADAMP	2525	N, S	–	–	NC	–
APD	2353	N, S	NC	+	NC	NC
CAMP	6756	N, S, P	–	–	NC	NC
LAMP	5547	N, S, P	–	–	NC	NC

N, natural peptide; S, synthetic peptide; P, predicted peptide; NC, not complete; '+' and '–' signify the existence or nonexistence the corresponding function respectively.

Table 4. Comparison of general type databases' search system

	Search system options									
Database	Sequence length	N terminus	C terminus	Unusual amino acid	Intrachain bond	Complexity	Synthesis type	Target group	Target object of cell	Prediction tool
DBAASP	+	+	+	+	+	+	+	+	+	+
YADAMP	+	–	–	–	–	–	–	–	–	–
APD	–	–	+	+	+	+	–	+	+	–
CAMP	–	–	–	–	–	–	–	–	–	+
LAMP	–	–	–	–	–	–	–	+	–	–

'+' and '–' signify the existence or nonexistence the corresponding function respectively.

and CAMP allow sequence-based prediction of the existence of antimicrobial activity. The accuracy of DBAASP's prediction on the testing set (91%) is slightly better than that for CAMP (90%) (Vishnepolsky & Pirtskhalava, 2014).

Conclusion

To improve the antimicrobial properties of existing AMPs or design new active ones, data regarding the peptide's chemical structure and antimicrobial activities are needed. DBAASP provides users with detailed information concerning peptide sequence; N and C end modifications; source; bonds; and post-translational modification of amino acids. All records contain information about antimicrobial activity of the peptide. The database is updated regularly and the system is provided with prediction algorithms for AMPs. (The development of prediction algorithm is in progress.) The prediction algorithms and knowledge of a peptide's detailed structure and activity are necessary for the cheaper rational design of new therapeutic peptides.

Availability

DBAASP is accessible at http://www.biomedicine.org.ge/dbaasp/.

Acknowledgements

This project is supported by Shota Rustaveli National Science Foundation (Georgia) and Centre National de la Recherche Scientifique (France) Grant No. 09/10. We acknowledge support from ISTC-BTEP Grant No. G-2102 and collaborators from members of the Office of Cyber Infrastructure and Computational Biology (OCICB), National Institute of Allergy and Infectious Diseases (NIAID), Mike Tartakovsky, Alex Rosenthal, Andrei Gabrielian and Darrell Hurt.

References

Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol* **3**: 238–250.

Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P & Kucherov G (2008) NORINE: a database of nonribosomal peptides. *Nucleic Acids Res* **36** (Database issue): D326–D331.

Caboche S, Leclere V, Pupin M, Kucherov G & Jacques P (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *J Bacteriol* **192**: 5143–5150.

Cavalier-Smith T (2004) Only six kingdoms of life. *Proc Biol Sci* **271**: 1251–1262.

Giudicelli V & Lefranc MP (2012) IMGT-ONTOLOGY 2012. *Front Genet* **3**: article79.

Gueguen Y, Garnier J, Robert L *et al.* (2006) PenBase, the shrimp antimicrobial peptide penaeidin database: sequence-based classification and recommended nomenclature. *Dev Comp Immunol* **30**: 283–288.

Hammami R, Zouhir A, Ben Hamida J & Fliss I (2007) BACTIBASE: a new web-accessible database for bacteriocin characterization. *BMC Microbiol* **17**: 89.

Hammami R, Ben Hamida J, Vergoten G & Fliss I (2009) PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Res* **37** (Database issue): D963–D968.

Li Y & Chen Z (2008) RAPD: a database of recombinantly-produced antimicrobial peptides. *FEMS Microbiol Lett* **289**: 126–129.

Novkovic M, Simunic J, Bojovic V, Tossi A & Juretic D (2012) DADP: the database of anuran defense peptides. *Bioinformatics* **28**: 1406–1407.

Piotto SP, Sessa L, Concilio S & Iannelli P (2012) YADAMP: yet another database of antimicrobial peptides. *Int J Antimicrob Agents* **39**: 346–351.

Reddy KV, Yedery RD & Aranha C (2004) Antimicrobial peptides: premises and promises. *Int J Antimicrob Agents* **24**: 536–547.

Seebah S, Suresh A, Zhuo S, Choong YH, Chua H, Chuon D, Beuerman R & Verma C (2007) Defensins knowl-edgebase: a manually curated database and information source focused on the defensins family of antimicrobial peptides. *Nucleic Acids Res* **35** (Database issue): D265–D268.

Sundararajan VS, Gabere MN, Pretorius A, Adam S, Christoffels A, Lehväslaiho M, Archer JAC & Bajic VB (2012) DAMPD: a manually curated antimicrobial peptide database. *Nucleic Acids Res* **40** (Database issue): D1108–D1112.

Vishnepolsky B & Pirtskhalava M (2014) Prediction of linear cationic antimicrobial peptides based on characteristics responsible for their interaction with the membranes. *J Chem Inf Model* **54**: 1512–1523.

Waghu FH, Gopi L, Barai RS, Ramteke P, Nizami B & Idicula-Thomas S (2014) CAMP: collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res* **42** (Database issue): D1154–D1158.

Wang G (2008) Structures of human host defense cathelicidin LL-37 and its smallest antimicrobial peptide KR-12 in lipid micelles. *J Biol Chem* **283**: 32637–32643.

Wang G, Li X & Wang Z (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res* **37** (Database issue): D933–D937.

Whitmore L & Wallace BA (2004) The peptaibol database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res* **32** (Database issue): D593–D594.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. DBAASP peptide card.
Data S2. DBAASP search page.
Data S3. DBAASP search result.
Data S4. DBAASP ranking search and result pages.