

SATPdb: a database of structurally annotated therapeutic peptides

Sandeep Singh, Kumardeep Chaudhary, Sandeep Kumar Dhandra, Sherry Bhalla, Salman Sadullah Usmani, Ankur Gautam, Abhishek Tuknait, Piyush Agrawal, Deepika Mathur and Gajendra P.S. Raghava*

Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, India

Received September 01, 2015; Revised September 30, 2015; Accepted October 13, 2015

ABSTRACT

SATPdb (<http://crdd.osdd.net/raghava/satpdb/>) is a database of structurally annotated therapeutic peptides, curated from 22 public domain peptide databases/datasets including 9 of our own. The current version holds 19192 unique experimentally validated therapeutic peptide sequences having length between 2 and 50 amino acids. It covers peptides having natural, non-natural and modified residues. These peptides were systematically grouped into 10 categories based on their major function or therapeutic property like 1099 anticancer, 10585 antimicrobial, 1642 drug delivery and 1698 antihypertensive peptides. We assigned or annotated structure of these therapeutic peptides using structural databases (Protein Data Bank) and state-of-the-art structure prediction methods like I-TASSER, HHsearch and PEPstrMOD. In addition, SATPdb facilitates users in performing various tasks that include: (i) structure and sequence similarity search, (ii) peptide browsing based on their function and properties, (iii) identification of moonlighting peptides and (iv) searching of peptides having desired structure and therapeutic activities. We hope this database will be useful for researchers working in the field of peptide-based therapeutics.

INTRODUCTION

The past decade has seen an unprecedented growth in peptide-based research (1–2). Over the last decade, the peptide drug market has been growing well, and this could be exemplified by the peptide drug statistics. Currently, around 60–70 peptide drugs have already been approved and out of these, 5 drugs were approved in 2012 alone. Around 200 peptide drugs are in various clinical trials, and around 600 are being evaluated in pre-clinical studies. It is estimated that thousands of potential peptide lead molecules are on

the laboratory bench. These figures clearly indicate the importance of peptide-based therapeutics in the field of drug discovery (3–6). These peptide-based drugs have several advantages over drugs based on small molecules, proteins and antibodies.

Though considerable progress has been made in the field of peptide-based therapeutics, a number of challenges remain to be addressed like poor oral delivery, low bioavailability and low stability of peptides (7). The problem of poor plasma half-life and low bioavailability can be resolved using rational designing of peptide therapeutics by improving their physicochemical properties. Similarly, stability of peptides can be improved using incorporation of D-amino acids, cyclization, alterations in peptide backbone chemistry, terminal and side chain modifications (8). In addition, a significant progress has been made over the years in the field of computational peptidology (9). Several computational methods have been developed for predicting pharmacologically important properties of peptides like toxicity (10), half-life (11), antiangiogenic (12), antimicrobial (13–15) immunogenic peptides (16). These *in silico* prediction tools not only help in designing peptide analogs with improved physicochemical properties but also help in screening peptide libraries for the desired therapeutic property. *In silico* screening of therapeutic peptides (e.g. antimicrobial, cell-penetrating peptides) followed by experimental validation is the most effective approach for discovering novel therapeutic peptides (17).

In the past, a large number of databases have been developed to maintain different kinds of peptides that include antimicrobial (15,18–20), antiviral (21–22), cell-penetrating (23), tumor homing (24), hemolytic (25) peptides. Many identical peptides exist in different databases and have multiple functions (i.e. moonlighting characteristics). These moonlighting peptides have enormous therapeutic potential as single peptide may be used for multiple tasks (26). These peptides can be useful in the repositioning of peptide drugs, which have already passed toxicity and other safety tests and reduce the significant costs incurred by pharmaceutical companies during clinical trials. Ideally, a

*To whom correspondence should be addressed. Tel: +91 172 2690557; Fax: +91 172 2690632; Email: raghava@imtech.res.in

researcher would be interested in having prior information about all the properties of a peptide in order to explore its full therapeutic potential. Presently, identification of different properties of a peptide is a daunting task, as one has to explore different databases available at different websites. In addition, some databases do not maintain the structure of peptides that is required to understand the structure-function relationship as well as for designing better therapeutic peptides. Therefore, we developed SATPdb database, which is a collection of therapeutically important peptides from different peptide databases/datasets. These peptides were curated and classified based on their major function, therapeutic property and sub-function. SATPdb is a unique resource, which allows users to explore the different properties as well as structure of peptides.

MATERIALS AND METHODS

Data collection

We obtained therapeutically important peptides from 20 databases (13,15,18,21–23,25,27–38) using export option provided by databases or using ‘wget’ command. In case of ‘wget’, we downloaded information of peptides in ‘HTML’ format, which were processed using in-house PERL scripts to extract desired information. In addition, we also obtained peptides from two peptide datasets containing antiangiogenic and toxic peptides that were used in the development of prediction methods ‘AntiAngioPred’ (12) and ‘Toxinpred’ (10) respectively. We removed all the peptides having undefined amino acid in their sequence (for example sequence having amino acid X and meaning of X is not defined). We also excluded peptide sequences having more than 50 amino acids. Our database contains peptide sequences having 2–50 amino acids.

Curation and compilation of peptides

We curated the function and therapeutic properties of each peptide from their source database. We tried our best to extract only those peptides whose therapeutic activities have been experimentally validated. Most of the databases report single function or property of a peptide like CPPsite, which is a collection of peptides having cell-penetrating property and AHTPDB reports only antihypertensive nature of peptides. There are other databases such as APD2, CAMP, CancerPPD, Hemolytik, ParaPep and PhytAMP that contain information on different activities of a peptide. For example, APD2 reports diverse therapeutic properties (antioxidant, anticancer, antiparasitic and antiviral) of peptides apart from antimicrobial activity. We manually curated these peptides along with their functions.

Based on the functional information, all peptides were grouped into 10 major functional categories; each functional category had at least 100 peptides. These functional categories are: (i) antimicrobial, (ii) anticancer, (iii) antiviral, (iv) antibacterial, (v) antifungal, (vi) antiparasitic, (vii) antihypertensive, (viii) cell-cell communication, (ix) drug delivery vehicle and (x) toxic. Though we have made separate functional category for antibacterial, antifungal, antiparasitic and antiviral peptides, these peptides are also

covered under antimicrobial. There are many peptides available in several databases assigned function as antimicrobial (not assigned specifically as antibacterial/antifungal, etc.). All such types of peptides are compiled under antimicrobial functional category. There are few peptides, which possess some other functions also like antioxidant, immunomodulatory, insecticidal. Such peptides have been compiled under section ‘Additional Information’. There are few peptides particularly from conserver for which functional information has not been provided. Since we don’t want to lose any information on therapeutic peptides, such peptides whose functions are not available in the source databases are covered under the section ‘Miscellaneous’ category.

Each functional category has been further classified into sub-categories based on therapeutic activity of peptides. For instance, peptides belonging to anticancer functional category were further divided into antitumor and antiangiogenic sub-categories. Similarly, antiparasitic peptides were further categorized into antiplasmodium, antitypanosomic and antileishmania.

Structural annotation of peptides

A systematic approach was used to perform structural annotation of peptides with following steps. First, all the peptide sequences in SATPdb database were searched for an identical sequence in Protein Data Bank (PDB) (39). In case, an identical sequence was available in PDB then we assigned structure of peptide as given in PDB. If the identical sequence was not available in PDB then we used different structure prediction techniques for predicting the structure of peptides depending on length of peptides. We used a simple approach for predicting tertiary structure of peptides having 2–4 amino acids. First, an initial structure of the peptide with linear conformation was generated by assigning phi and psi torsion angle of each peptide residue as 180°. This initial structure was then subjected to energy minimization followed by molecular dynamics simulation to get the final predicted structure using PEPstrMOD. In order to predict the structure of peptides having 5–30 amino acids, we used web server PEPstrMOD (in parallel communication). PEPstrMOD is an updated version of PEPstr (40) which is a state-of-art method used to predict tertiary structure of natural peptides with length ranging from 7 to 25 amino acids. In addition, PEPstrMOD has been used for predicting the structure of peptides having non-natural or modified residues. These modifications include terminal modifications (e.g. Acetylation, Amidation), D-amino acids, post-translational modifications (e.g. phosphorylation, hydroxylation). PEPstrMOD integrates special force field libraries (FFNCAA (41), FFPTM (42) and SwissSideChain (43–44)) to handle modified amino acids. The peptides with length ranging from 31 to 50 amino acids having high homology with known structures in PDB (HHsearch probability value $\geq 70\%$) were predicted using homolog modeling. First, best templates were identified using HHblits (45) and HHsearch (46) then MODELLER software (47) was used to generate the tertiary structure of a peptide using best templates. The probability value is more reliable criteria for selecting templates rather than sequence identity or E-value (46). Finally I-TASSER Suite (48) was

used for solving the structure of remaining peptides. We also assigned the eight types of secondary structure states for the peptides in SATPdb using DSSP software (49).

Database architecture and web interface

We have developed SATPdb database using Apache HTTP server (version 2.2.17) integrated with PHP (version 5.2.14) and MySQL (version 14.12) on server machine with Red Hat Enterprise Linux (version 6.2) as operating system. PHP and JavaScript (version 1.7) were used to develop the front-end of the database while MySQL was used to process the data at the back-end. The graphical representation of an overall architecture of SATPdb is displayed in Figure 1. We used CSS and HTML5 to create a responsive template, in order to make our site compatible for mobile, tablet and desktop.

Integration of web tools

In order to assist the users in searching, analyzing and retrieving the data from SATPdb, we develop user-friendly interfaces. Following is a brief description of menus and sub-menus available in database SATPdb.

Data retrieval web tools. We have incorporated six web-based tools to assist users in searching peptides based on their sequence, secondary and tertiary structure. These tools are briefly described as follows. (i) Search: it allows to perform an extensive search against major fields in SATPdb. (ii) Segment search: it allows to search a query peptide sequence and identify identical peptides or peptides with a common segment. (iii) Peptide mapping: this tool searches and maps the peptides in SATPdb on a query protein sequence and is useful for identification of therapeutically important segments in a query protein. (iv) Sequence similarity: it allows to perform similarity search against SATPdb using BLAST (50–51), (v) Secondary structure: this tool identifies peptides having desired secondary structure and (vi) Tertiary structure: it allows users to perform similarity search based on tertiary structure.

Searching moonlighting peptides. Many peptides in SATPdb belong to two or more functional categories and, therefore, possess moonlighting properties. Therefore to assist in searching peptides, which may have moonlighting characteristics, we have provided a ‘Desired Function’ facility to search for peptides with a desired function. For example, users can search for peptides, which may be used as a drug delivery vehicle as well as possess anticancer property but should not be toxic. In addition, this menu allows users to identify peptides having multiple functions and peptides having exclusive function. Figure 2 represents a Venn diagram (plotted using ‘colorfulVennPlot’ package (<https://cran.r-project.org/web/packages/colorfulVennPlot/>)) of peptides common among four functional categories.

Browsing in SATPdb. This menu allows users to extract information in a classified form where a user can obtain peptides having specific property. User can browse peptide en-

tries based on function/sub-function, additional information (activity or function which is not categorized in main functional category), physicochemical properties, peptide modifications, secondary structure states and frequency of peptides in multiple functional categories. Therefore, if a user is interested in extracting all the peptides which can be used as a drug delivery vehicle or peptides which are anticancer in nature, user can use ‘Browse Function/Sub-function’ to get the desired results. Similarly, if a user is interested in peptides with more than 75% positive charged residues, a user can use ‘Browse Properties’ section to get the desired entries.

RESULTS

Peptide functions

Based on curated information on functions, therapeutic peptides in SATPdb have been grouped in 10 major functional categories. The maximum numbers of peptides are in the category of antimicrobial peptides (35%), followed by antiviral (11.5%), antihypertensive (5.6%), drug delivery peptides (5.5%) and anticancer (3.6%) peptides (Table 1). A significant number of peptides (14%) belong to category of toxic peptides. These toxic peptides are either cytotoxic (55%) or hemolytic (45%) in nature. Antiparasitic peptides are further categorized into antiplasmodium (68%), antileishmania (15%) and antitrypanosomic (17%). Similarly, drug delivery peptides are classified into tumor homing (40%), cell-penetrating (51%) and blood-brain barrier peptides (9%). One percent of peptides fall under ‘Additional Information’ section that contain diverse class of peptides like cysteine protease inhibitor, insecticidal, antioxidant, chemotactic, immunomodulatory, etc.

Peptide statistics

The current release of SATPdb consists of 19192 unique peptides collected from 22 different published peptide databases/datasets. All peptides in SATPdb have length between 2 and 50 amino acids but the maximum numbers of peptides (7396) have length between 11 and 20 amino acids. It is interesting that despite the fact these peptides belongs to different resources, most of the peptides have multiple functions, for example 7512 peptides have 2–3 functions, 581 peptides have 4–6 functions while 10 peptides have more than six functions.

Structure statistics

SATPdb stores information of various types of peptides, including linear (18616), cyclic (538), peptides having D-amino acids (437) and peptides with various chemical modifications (1407). A total 644 peptide structures were collected from PDB and stored as such in SATPdb while 13444 peptide structures were predicted using PEPstrMOD including 12770 peptides having natural amino acids and 674 peptides having modified residues. 2589 peptides having more than 30 natural amino acids were predicted using homology-based approach while 607 peptides were predicted using I-TASSER suite. The structures of few peptides having complex chemical modifications were not predicted. Finally, SATPdb maintains a total of 17284 peptide

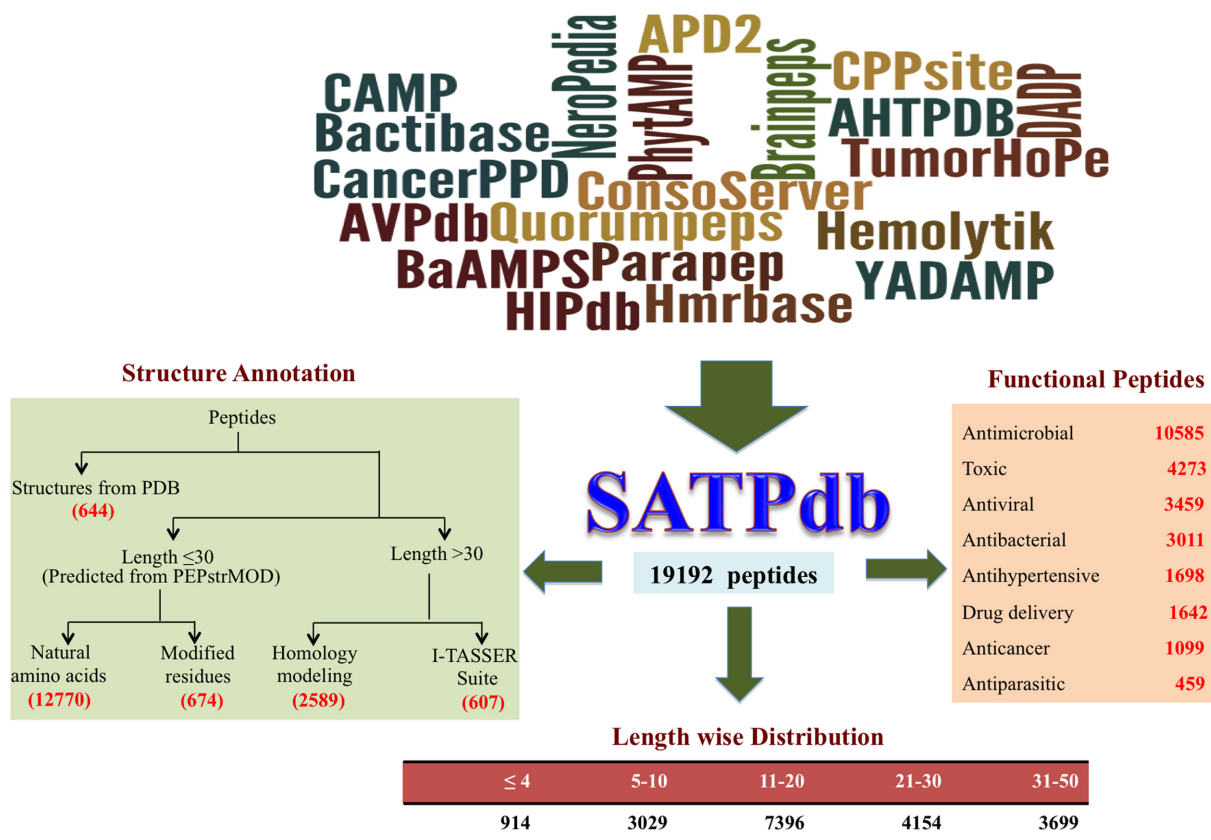


Figure 1. Schematic representation of architecture of SATPdb.

Table 1. List of functional and sub-functional categories of peptides in SATPdb

Functional category	Peptides ^a	Sub-functional category	Databases/Datasets
Antibacterial	3011	Anti-gram(+ve), Anti-gram(-ve)	APD2, BACTIBASE, CAMP, CancerPPD, Hemolytik, ParaPep, PhytAMP
Antifungal	1435	Antiyeast	APD2, CAMP, CancerPPD, Hemolytik, ParaPep, PhytAMP
Antiviral	3459	NA	APD2, AVPdb, CAMP, CancerPPD, Hemolytik, HIPdb, ParaPep, PhytAMP
Antiparasitic	459	Antiplasmodium, Antileishmania, Antitrypanosomic	APD2, CAMP, CancerPPD, Hemolytik, ParaPep
Antimicrobial	10585	NA	APD2, AVPdb, BaAMPs, BACTIBASE, CAMP, CancerPPD, DADP, Hemolytik, HIPdb, ParaPep, PhytAMP, YADAMP
Antihypertensive	1698	NA	AHTPDB, CancerPPD
Anticancer	1099	Antitumor, Antiangiogenic	AntiAngioPred, APD2, CAMP, CancerPPD, Hemolytik, ParaPep
Cell-cell communication	712	Hormones, Quorum sensing	Hmrbase, NeuroPedia, Quorumpeps
Drug delivery vehicle	1642	Cell-penetrating, Tumor homing, passing blood-brain barrier	Brainpeps, CPPsite, Hemolytik, ParaPep, TumorHoPe
Toxic	4273	Hemolytic, Cytotoxic	APD2, Hemolytik, ParaPep, ToxinPred
Miscellaneous	1837	NA	ConoServer

^aNumber of peptides.

tertiary structures. All the annotated peptide structures, as well as, peptide sequences can be downloaded from ‘Download’ web page available at SATPdb.

Utility of SATPdb

SATPdb is a powerful resource, which can provide structural annotation of most of the therapeutic peptides published so far. Since structure of a peptide plays an important role in its function (52), a user can exploit structural information of peptides provided in SATPdb for performing

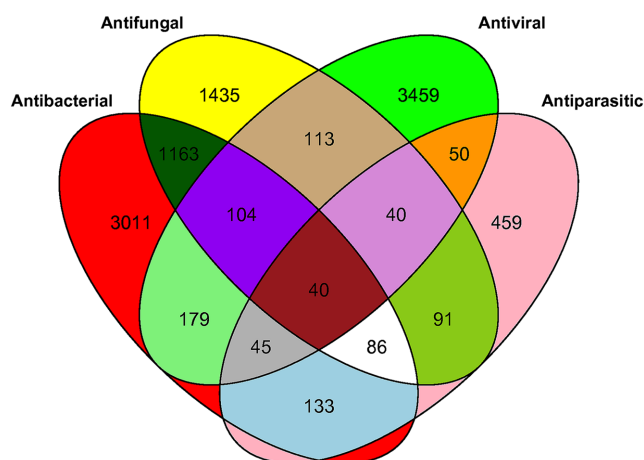


Figure 2. Venn diagram demonstrating peptides with multiple functions (antibacterial, antifungal, antiviral and antiparasitic).

various studies like molecular simulations, docking, binding pockets and peptide interaction with receptors. In addition, there are many peptides, which were initially designed to have a particular therapeutic property but later on characterized to show some other functionalities as well. For example, TAT peptide was initially discovered as cell-penetrating peptide (53) but later studies have revealed that it also possess antimicrobial activity (54). Now few studies have shown that it has the capability to cross the blood brain barrier (55). Such information is very important but scattered in the literature/databases and thus it is very difficult for a user to gather all such information for a single peptide. SATPdb is an attempt to fill this gap and it brings together most of the therapeutic peptides on a single platform. Using SATPdb, a user can perform meta-analysis and can search peptides having multiple functionalities. In order to demonstrate an application of SATPdb in real life, we performed following two case studies.

Case Study 1: Antimicrobial peptides having drug delivery property with minimum side effects.

We used ‘Desired Function’ sub-menu of ‘Moonlighting’ menu of SATPdb for identification of peptides with desired function. We searched ‘antimicrobial AND drug delivery vehicle NOT toxic’ as a query in SATPdb. This search extracted 11 peptide sequences, which possess above desired properties/functions. Among these 11 peptides, 3 of them also possess a third function among which 1 peptide is antihypertensive, and the other 2 are involved in cell-cell communication (Table 2).

Case 2. Peptides having antihypertensive and anticancer activity with minimum side effects.

We extracted four peptides having antihypertensive and anticancer properties but at the same time they do not have any toxicity. These peptides were extracted using ‘Desired Function’ sub-menu of ‘Moonlighting’ menu of SATPdb. Moreover, all these peptides also possess antioxidant property and three of them possess immunomodulatory effects (Table 2). These peptides may be useful as their activity have been tested experimentally in previous studies.

DISCUSSION

The increasing frequencies of reports of drug resistance are becoming a serious predicament to global healthcare (56). There is a pressing need to discover novel and more effective therapeutic agents. To tackle this grim situation, therapeutic peptides have gained significant attention of the researchers as safe and effective alternatives with their high efficacies and low toxicity, high cell-penetration and ease of synthesis (3,5). Though many peptide databases exist, which cover different biological properties like anticancer, antimicrobial, antiparasitic, cell-penetrating, hemolytic, antihypertensive, etc., there is a need for a unified platform integrating majority of the peptides from various databases. SATPdb is developed in an attempt to make a meta-database comprising a collection of peptide sequences from 22 different peptide databases/datasets. All entries in SATPdb are cross-linked with individual databases to provide an option for easy switching to the original source of the information present in individual databases. Most of the peptide entries are structurally annotated and tertiary structures of peptides are provided for downloading. Moreover, many moonlighting peptides with desired functions can be effectively searched using SATPdb.

Briefly, users can take advantage of SATPdb in following ways: users can (i) search a peptide of interest in 22 peptide databases/datasets at one go and therefore save time, (ii) browse peptides of SATPdb with similar physicochemical properties or secondary structure content, (iii) extract moonlighting peptides with desired functions and (iv) extract structural information of most of the peptides including peptides with non-natural residues which can be used for further structure-to-function analysis and docking studies. In summary, SATPdb is a useful resource and we hope that it will expedite the peptide-based research.

UPDATE OF SATPdb

We will update SATPdb at regular intervals and in the updated version we will include newly developed peptide databases as well as the new release of the databases presently incorporated in SATPdb.

LIMITATIONS

Although many peptide databases are covered in SATPdb, yet other databases (like epitope-based, PepBank, etc.) are not included in SATPdb. Epitope-based databases like IEDB holds more than 1 47 000 peptide epitopes. As SATPdb also stores and maintains annotated structures of peptides, performing the structural annotation of a large number of peptides was difficult to handle in the current version of SATPdb. We understand that these are important databases and should be included in the SATPdb. In the next update we will also include other important databases like PepBank (57), IEDB (58), MHCBN (59). Moreover, structures of 1908 peptides ($1581 \leq 30$ residues and $327 > 30$ residues) were not annotated due lack of appropriate force field to handle complex modifications in these peptides. In future, with the availability of new force fields to handle residues with complex modifications, the structural annotation of these peptides will be feasible

Table 2. List of peptides with brief description of their functional properties obtained from SATPdb using two case studies

Peptide sequences	Major functions	Sub-functions	Additional information
Case Study 1 ('antimicrobial AND drug delivery vehicle NOT toxic')			
RPKQQQFFGLM	antibacterial, antimicrobial, antifungal, drug delivery vehicle	anti-gram(+ve), anti-gram(-ve), blood brain barrier	NA
PGP	antiviral, antihypertensive, antimicrobial, drug delivery vehicle	blood brain barrier	NA
SEEPPISLDLTFHLLREVLEM	antiparasitic, cell-cell communication, antimicrobial, drug delivery vehicle	antitrypanosomic, hormones, blood brain barrier	Neuropeptide
ARAEQLAQQAHSNRKLMELI MEHFPGP	antiviral, antimicrobial, drug delivery vehicle	blood brain barrier	NA
SVSVGMKPSRP	antiviral, antimicrobial, drug delivery vehicle	tumor homing peptide	NA
YPSKPDNPGEDAPAEDLARY YSALRHYNLITRQRY	antimicrobial, drug delivery vehicle	blood brain barrier	NA
RQIRIWFQNRMRWR	antimicrobial, drug delivery vehicle	cell-penetrating peptide	NA
RQIKIWFQNRMRKWK	antiviral, antimicrobial, drug delivery vehicle	cell-penetrating peptide	NA
DAEFRHDSGYEVHHQKLVFF	antibacterial, antiviral, antimicrobial, antifungal, drug delivery vehicle	anti-gram(+ve), anti-gram(-ve), blood brain barrier	NA
AEDVGSNKGAIIGLMVGGVV	antibacterial, antiparasitic, cell-cell communication, antimicrobial, antifungal, drug delivery vehicle	antitrypanosomic, blood brain barrier	Neuropeptide
HSDAVFTDNYTRLRKQMAVK KYLNSILN	antimicrobial, drug delivery vehicle	cell-penetrating peptide	NA
GLFRALLRLLRSLWRLLRA			
Case Study 2 ('antihypertensive AND anticancer NOT toxic')			
RYLGYL	anticancer, antiviral, antihypertensive, antimicrobial	NA	immunomodulatory, antioxidant
PPPEE	anticancer, antiviral, antihypertensive, antimicrobial	NA	immunomodulatory, antioxidant
YPPFG	anticancer, antiviral, antihypertensive, antimicrobial	NA	immunomodulatory, antioxidant
FKCRRWQWRMKK	anticancer, antiviral, antihypertensive, antimicrobial	NA	antioxidant

AVAILABILITY

SATPdb can be accessed freely at <http://crdd.osdd.net/raghava/satpdb/>.

ACKNOWLEDGEMENTS

Authors' are thankful to Council of Scientific and Industrial Research (CSIR), Indian Council of Medical Research (ICMR), Department of Science and Technology (DST) and Department of Biotechnology (DBT), Government of India for financial support and fellowships.

Authors Contribution: S.S., K.C., P.A. and S.S.U. collected and compiled the data. S.S., K.C. and A.T. performed structural annotation. S.S., K.C., S.B., S.S.U., P.A., S.K.D. and A.T. developed the web interface. S.S., A.G., P.A., D.M., K.C. and G.P.S.R. prepared the manuscript. G.P.S.R. conceived the idea and coordinated the project.

FUNDING

Funding for open access charge: Council of scientific and Industrial Research (CSIR) [fellowships and project: Open

Source Drug Discovery and GENESIS BSC0121]; Department of Biotechnology [project: BTISNET]; Government of India.

Conflict of interest statement. None declared.

REFERENCES

- Albericio, F. and Kruger, H.G. (2012) Therapeutic peptides. *Future Med. Chem.*, **4**, 1527–1531.
- Otvos, L. Jr (2008) Peptide-based drug design: here and now. *Methods Mol. Biol.*, **494**, 1–8.
- Craik, D.J., Fairlie, D.P., Liras, S. and Price, D. (2013) The future of peptide-based drugs. *Chem. Biol. Drug Des.*, **81**, 136–147.
- Fosgerau, K. and Hoffmann, T. (2015) Peptide therapeutics: current status and future directions. *Drug Discov. Today*, **20**, 122–128.
- Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatskiy, M. (2010) Synthetic therapeutic peptides: science and market. *Drug Discov. Today*, **15**, 40–56.
- Kaspar, A.A. and Reichert, J.M. (2013) Future directions for peptide therapeutics development. *Drug Discov. Today*, **18**, 807–817.
- Otvos, L. Jr and Wade, J.D. (2014) Current challenges in peptide-based drug discovery. *Front. Chem.*, **2**, 62, 1–4, PMID 25152873.
- Gentilucci, L., De Marco, R. and Cerisoli, L. (2010) Chemical modifications designed to improve peptide stability: incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Curr. Pharm. Des.*, **16**, 3185–3203.

9. Zhou,P., Wang,C., Ren,Y., Yang,C. and Tian,F. (2013) Computational peptidology: a new and promising approach to therapeutic peptide design. *Curr. Med. Chem.*, **20**, 1985–1996.
10. Gupta,S., Kapoor,P., Chaudhary,K., Gautam,A., Kumar,R. and Raghava,G.P. (2013) In silico approach for predicting toxicity of peptides and proteins. *PLoS One*, **8**, e73957.
11. Sharma,A., Singla,D., Rashid,M. and Raghava,G.P. (2014) Designing of peptides with desired half-life in intestine-like environment. *BMC Bioinformatics*, **15**, 282, 1–8, PMID 25141912.
12. Ettayapuram Ramaprasad,A.S., Singh,S., Gajendra,P.S.R. and Venkatesan,S. (2015) AntiAngioPred: A Server for Prediction of Anti-Angiogenic Peptides. *PLoS One*, **10**, e0136990.
13. Wang,G., Li,X. and Wang,Z. (2009) APD2: the updated antimicrobial peptide database and its application in peptide design. *Nucleic Acids Res.*, **37**, D933–D937.
14. 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.
15. Waghu,F.H., Gopi,L., Barai,R.S., Ramteke,P., Nizami,B. and Idicula-Thomas,S. (2014) CAMP: Collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res.*, **42**, D1154–D1158.
16. Dhandu,S.K., Gupta,S., Vir,P. and Raghava,G.P. (2013) Prediction of IL4 inducing peptides. *Clin. Dev. Immunol.*, **2013**, 263952, 1–9, PMID 24489573.
17. Gautam,A., Sharma,M., Vir,P., Chaudhary,K., Kapoor,P., Kumar,R., Nath,S.K. and Raghava,G.P. (2015) Identification and characterization of novel protein-derived arginine-rich cell-penetrating peptides. *Eur. J. Pharm. Biopharm.*, **89**, 93–106.
18. Piotto,S.P., Sessa,L., Concilio,S. and Iannelli,P. (2012) YADAMP: yet another database of antimicrobial peptides. *Int. J. Antimicrob. Agents*, **39**, 346–351.
19. Théolier,J., Fliss,I., Jean,J. and Hammami,R. (2014) MilkAMP: a comprehensive database of antimicrobial peptides of dairy origin. *Dairy Sci. Technol.*, **94**, 181–193.
20. Gogoladze,G., Grigolava,M., Vishnepolsky,B., Chubinidze,M., Duroux,P., Lefranc,M.P. and Pirtskhalava,M. (2014) DBAASP: database of antimicrobial activity and structure of peptides. *FEMS Microbiol. Lett.*, **357**, 63–68.
21. Qureshi,A., Thakur,N., Tandon,H. and Kumar,M. (2014) AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. *Nucleic Acids Res.*, **42**, D1147–D1153.
22. Qureshi,A., Thakur,N. and Kumar,M. (2013) HIPdb: a database of experimentally validated HIV inhibiting peptides. *PLoS One*, **8**, e54908.
23. Gautam,A., Singh,H., Tyagi,A., Chaudhary,K., Kumar,R., Kapoor,P. and Raghava,G.P. (2012) CPPsite: a curated database of cell penetrating peptides. *Database (Oxford)*, **2012**, bas015, 1–7, PMID 22403286.
24. Kapoor,P., Singh,H., Gautam,A., Chaudhary,K., Kumar,R. and Raghava,G.P. (2012) TumorHoPe: a database of tumor homing peptides. *PLoS One*, **7**, e35187.
25. Gautam,A., Chaudhary,K., Singh,S., Joshi,A., Anand,P., Tuknait,A., Mathur,D., Varshney,G.C. and Raghava,G.P. (2014) Hemolytik: a database of experimentally determined hemolytic and non-hemolytic peptides. *Nucleic Acids Res.*, **42**, D444–D449.
26. Rodriguez Plaza,J.G., Villalon Rojas,A., Herrera,S., Garza-Ramos,G., Torres Larios,A., Amero,C., Zarraga Granados,G., Gutierrez Aguilar,M., Lara Ortiz,M.T., Polanco Gonzalez,C. *et al.* (2012) Moonlighting peptides with emerging function. *PLoS One*, **7**, e40125.
27. Kumar,R., Chaudhary,K., Sharma,M., Nagpal,G., Chauhan,J.S., Singh,S., Gautam,A. and Raghava,G.P.S. (2015) AHTPDB: a comprehensive platform for analysis and presentation of antihypertensive peptides. *Nucleic Acids Res.*, **43**, D956–D962.
28. Di Luca,M., Maccari,G., Masetta,G. and Batoni,G. (2015) BaAMPs: the database of biofilm-active antimicrobial peptides. *Biofouling*, **31**, 193–199.
29. Hammami,R., Zouhir,A., Le Lay,C., Ben Hamida,J. and Fliss,I. (2010) BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiol.*, **10**, 22, 1–5, PMID 20105292.
30. Van Dorpe,S., Bronselaer,A., Nielandt,J., Stalmans,S., Wynendaele,E., Audenaert,K., Van De Wiele,C., Burvenich,C., Peremans,K., Hsouchou,H. *et al.* (2012) Brainpeps: the blood-brain barrier peptide database. *Brain Struct. Funct.*, **217**, 687–718.
31. Tyagi,A., Tuknait,A., Anand,P., Gupta,S., Sharma,M., Mathur,D., Joshi,A., Singh,S., Gautam,A. and Raghava,G.P.S. (2015) CancerPPD: a database of anticancer peptides and proteins. *Nucleic Acids Res.*, **43**, D837–D843.
32. Kaas,Q., Yu,R., Jin,A.H., Dutertrre,S. and Craik,D.J. (2012) ConoServer: updated knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.*, **40**, D325–D330.
33. Novkovic,M., Simunic,J., Bojovic,V., Tossi,A. and Juretic,D. (2012) DADP: the database of anuran defense peptides. *Bioinformatics*, **28**, 1406–1407.
34. Rashid,M., Singla,D., Sharma,A., Kumar,M. and Raghava,G.P. (2009) Hmrbase: a database of hormones and their receptors. *BMC Genomics*, **10**, 307, 1–10, PMID 19589147.
35. Kim,Y., Bark,S., Hook,V. and Bandeira,N. (2011) NeuroPedia: neuropeptide database and spectral library. *Bioinformatics*, **27**, 2772–2773.
36. Mehta,D., Anand,P., Kumar,V., Joshi,A., Mathur,D., Singh,S., Tuknait,A., Chaudhary,K., Gautam,S.K., Gautam,A. *et al.* (2014) ParaPep: a web resource for experimentally validated antiparasitic peptide sequences and their structures. *Database (Oxford)*, **2014**, bau051, 1–7, PMID 24923818.
37. Hammami,R., Ben Hamida,J., Vergoten,G. and Fliss,I. (2009) PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic Acids Res.*, **37**, D963–968.
38. Wynendaele,E., Bronselaer,A., Nielandt,J., D’Hondt,M., Stalmans,S., Bracke,N., Verbeke,F., Van De Wiele,C., De Tre,G. and De Spiegeleer,B. (2013) Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. *Nucleic Acids Res.*, **41**, D655–D659.
39. Rose,P.W., Prlic,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
40. Kaur,H., Garg,A. and Raghava,G.P. (2007) PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept. Lett.*, **14**, 626–631.
41. Khoury,G.A., Smadbeck,J., Tamamis,P., Vandris,A.C., Kieslich,C.A. and Floudas,C.A. (2014) Forcefield.NCAA: ab initio charge parameters to aid in the discovery and design of therapeutic proteins and peptides with unnatural amino acids and their application to complement inhibitors of the compstatin family. *ACS Synthetic Biol.*, **3**, 855–869.
42. Khoury,G.A., Thompson,J.P., Smadbeck,J., Kieslich,C.A. and Floudas,C.A. (2013) Forcefield.PTM: charge and AMBER forcefield parameters for frequently occurring post-translational modifications. *J. Chem. Theory Comput.*, **9**, 5653–5674.
43. Gfeller,D., Michielin,O. and Zoete,V. (2013) SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Res.*, **41**, D327–D332.
44. Gfeller,D., Michielin,O. and Zoete,V. (2012) Expanding molecular modeling and design tools to non-natural sidechains. *J. Comput. Chem.*, **33**, 1525–1535.
45. Remmert,M., Biegert,A., Hauser,A. and Soding,J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, **9**, 173–175.
46. Soding,J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
47. Eswar,N., Webb,B., Marti-Renom,M.A., Madhusudhan,M.S., Eramian,D., Shen,M.Y., Pieper,U. and Sali,A. (2007) Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.*, **50**, 2.9.1–2.9.31, Chapter 2, Unit 2.9, PMID 18429317.
48. Yang,J., Yan,R., Roy,A., Xu,D., Poisson,J. and Zhang,Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12**, 7–8.
49. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
50. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

51. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
52. Eiriksdottir,E., Konate,K., Langel,U., Divita,G. and Deshayes,S. (2010) Secondary structure of cell-penetrating peptides controls membrane interaction and insertion. *Biochim. Biophys. Acta*, **1798**, 1119–1128.
53. Tyagi,M., Rusnati,M., Presta,M. and Giacca,M. (2001) Internalization of HIV-1 tat requires cell surface heparan sulfate proteoglycans. *J. Biol. Chem.*, **276**, 3254–3261.
54. Jung,H.J., Jeong,K.S. and Lee,D.G. (2008) Effective antibacterial action of tat (47–58) by increased uptake into bacterial cells in the presence of trypsin. *J. Microbiol. Biotechnol.*, **18**, 990–996.
55. Zou,L.L., Ma,J.L., Wang,T., Yang,T.B. and Liu,C.B. (2013) Cell-penetrating Peptide-mediated therapeutic molecule delivery into the central nervous system. *Curr. Neuropharmacol.*, **11**, 197–208.
56. Holohan,C., Van Schaeybroeck,S., Longley,D.B. and Johnston,P.G. (2013) Cancer drug resistance: an evolving paradigm. *Nat. Rev. Cancer*, **13**, 714–726.
57. Shtatland,T., Guettler,D., Kossodo,M., Pivovarov,M. and Weissleder,R. (2007) PepBank—a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinformatics*, **8**, 280, 1–10, PMID 17678535.
58. Vita,R., Overton,J.A., Greenbaum,J.A., Ponomarenko,J., Clark,J.D., Cantrell,J.R., Wheeler,D.K., Gabbard,J.L., Hix,D., Sette,A. *et al.* (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.*, **43**, D405–D412.
59. Bhasin,M., Singh,H. and Raghava,G.P. (2003) MHCBN: a comprehensive database of MHC binding and non-binding peptides. *Bioinformatics*, **19**, 665–666.