SATPdb: a database of structurally annotated therapeutic peptides

Sandeep Singh

Available at: https://works.bepress.com/sandeep_singh/1/
SATPdb: a database of structurally annotated therapeutic peptides


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

1To whom correspondence should be addressed. Tel: +91 172 2690557; Fax: +91 172 2690632; Email: raghava@imtech.res.in

© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
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 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 conoserver 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, antirypansomic 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 (HH-search probability value ≥ 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 predicting the structure of peptides with length ranging from 51 to 100 amino acids. I-TASSER Suite integrates a state-of-art method used to predict the structure of natural peptides with length range from 7 to 25 amino acids.
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 entries 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, antioxidiant, 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 sources, 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 peptides 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

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

*aNumber of peptides.

Structure Annotation

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...
properties and functions. Among these 11 peptides, 3 of them
tracted 11 peptide sequences, which possess above desired
vehicle NOT toxic’ as a query in SATPdb. This search ex-
function. We searched ‘antimicrobial AND drug delivery
menu of SATPdb for identification of peptides with desired
ery property with minimum side effects.

Moreover, all these peptides also possess antioxidant prop-
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 synthe-
sis (3, 5). Though many peptide databases exist, which cover
different biological properties like anticancer, antimicro-
bial, antiparasitic, cell-penetrating, hemolytic, antihyper-
tensive, etc., there is a need for a unified platform integrat-
ing majority of the peptides from various databases. SATP-
db is developed in an attempt to make a meta-database
comprising a collection of peptide sequences from 22 differ-
ent 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 pep-
tides are provided for downloading. Moreover, many moon-
lighting 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 stud-
ies. 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 pep-
tide databases as well as the new release of the databases
presently incorporated in SATPdb.

LIMITATIONS
Although many peptide databases are covered in SAT-
db, yet other databases (like epitope-based, PepBank, etc.)
are not included in SATPdb. Epitope-based databases like
IEDB holds more than 147 000 peptide epitopes. As SAT-
db also stores and maintains annotated structures of pep-
tides, performing the structural annotation of a large num-
er 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 ≤ 30 residues and 327
> 30 residues) were not annotated due lack of appropriate
force field to handle complex modifications in these pep-
tides. In future, with the availability of new force fields
to handle residues with complex modifications, the structural
annotation of these peptides will be feasible.

Figure 2. Venn diagram demonstrating peptides with multiple functions
(antibacterial, antifungal, antiviral and antiparasitic).
## Table 2. List of peptides with brief description of their functional properties obtained from SATPdb using two case studies

<table>
<thead>
<tr>
<th>Peptide sequences</th>
<th>Major functions</th>
<th>Sub-functions</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case Study 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPKPQQFFGLM</td>
<td>antibacterial, antimicrobial, antifungal, drug delivery vehicle</td>
<td>anti-gram(+ve), anti-gram(-ve), blood brain barrier</td>
<td>NA</td>
</tr>
<tr>
<td>PGP</td>
<td>antiviral, antihypertensive, antimicrobial, drug delivery vehicle, antiparasitic, cell-cell</td>
<td>antityrpanosomic, hormones, blood brain barrier</td>
<td>Neuropeptide</td>
</tr>
<tr>
<td>SEEPPISDLTFHLLREVLEM</td>
<td>communication, antimicrobial, drug delivery vehicle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARAEQLAQQAHSNKRKLEII</td>
<td>antiviral, antimicrobial, drug delivery vehicle</td>
<td>blood brain barrier</td>
<td>NA</td>
</tr>
<tr>
<td>MEHFGPG</td>
<td>antiviral, antimicrobial, drug delivery vehicle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSVSGMPSPRP</td>
<td>antiviral, antimicrobial, drug delivery vehicle, antimicrobial, drug delivery vehicle,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YPSKPDPNGEDAPAEDLARY</td>
<td>YSALRHYINLITRQRY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RQIRIFQNRRMRWRR</td>
<td>antimicrobial, drug delivery vehicle</td>
<td>cell-penetrating peptide</td>
<td>NA</td>
</tr>
<tr>
<td>RQKIFQNRRMKWKK</td>
<td>antiviral, antimicrobial, drug delivery vehicle</td>
<td>cell-penetrating peptide</td>
<td>NA</td>
</tr>
<tr>
<td>DAEFRHDSGYEVQHKVLFF</td>
<td>antimicrobial, drug delivery vehicle</td>
<td>anti-gram(+ve), anti-gram(-ve), blood brain barrier</td>
<td>NA</td>
</tr>
<tr>
<td>AEDVGSNKAIGHLMVGVV</td>
<td>antibacterial, antiparasisitc, cell-cell</td>
<td>antityrpanosomic, blood brain barrier</td>
<td>Neuropeptide</td>
</tr>
<tr>
<td>HSDAVFTDNYTRLKQMAVK</td>
<td>communication, antimicrobial, antifungal, drug delivery vehicle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KYLNSILN</td>
<td>antimicrobial, drug delivery vehicle</td>
<td>cell-penetrating peptide</td>
<td>NA</td>
</tr>
<tr>
<td>GLFRALLRLLSRLWLLRA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case Study 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RYLGYL</td>
<td>antancer, antiviral, antihypertensive, antimicrobial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPPEE</td>
<td>antancer, antiviral, antihypertensive, antimicrobial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YPFPG</td>
<td>antancer, antiviral, antihypertensive, antimicrobial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FKCRWQWWMK</td>
<td>antancer, antiviral, antihypertensive, antimicrobial</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### AVAILABILITY

SATPdb can be accessed freely at [http://crdd.osdd.net/raghava/satpdb/](http://crdd.osdd.net/raghava/satpdb/).

### ACKNOWLEDGEMENTS


### 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


