

# CancerPPD: a database of anticancer peptides and proteins

Atul Tyagi, Abhishek Tuknait, Priya Anand, Sudheer Gupta, Minakshi Sharma, Deepika Mathur, Anshika Joshi, Sandeep Singh, Ankur Gautam and Gajendra P.S. Raghava\*

Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh 160036, Punjab, India

Received August 5, 2014; Revised September 15, 2014; Accepted September 16, 2014

## ABSTRACT

CancerPPD (<http://crdd.osdd.net/raghava/cancerppd/>) is a repository of experimentally verified anticancer peptides (ACPs) and anticancer proteins. Data were manually collected from published research articles, patents and from other databases. The current release of CancerPPD consists of 3491 ACP and 121 anticancer protein entries. Each entry provides comprehensive information related to a peptide like its source of origin, nature of the peptide, anticancer activity, N- and C-terminal modifications, conformation, etc. Additionally, CancerPPD provides the information of around 249 types of cancer cell lines and 16 different assays used for testing the ACPs. In addition to natural peptides, CancerPPD contains peptides having non-natural, chemically modified residues and D-amino acids. Besides this primary information, CancerPPD stores predicted tertiary structures as well as peptide sequences in SMILES format. Tertiary structures of peptides were predicted using the state-of-art method, PEPstr and secondary structural states were assigned using DSSP. In order to assist users, a number of web-based tools have been integrated, these include keyword search, data browsing, sequence and structural similarity search. We believe that CancerPPD will be very useful in designing peptide-based anticancer therapeutics.

## INTRODUCTION

Cancer is one of the most devastating diseases accounting for millions of deaths worldwide every year (1). Conventional chemotherapy remains the principle mode of cancer treatment but it is not effective due to the adverse effects on normal cells and frequent development of multidrug-resistance by cancer cells (2). This grim situation under-

scores the urgent need to develop novel therapeutic means to tackle this deadly disease. In this context, small peptides known as anticancer peptides (ACPs) have shown tremendous potential as many ACPs exhibit cancer-selective toxicity and thus may avoid the shortcomings of the conventional chemotherapy (3). With several advantages, including high specificity, low intrinsic toxicity, high tissue penetration and ease of modifications, peptides have become the preferred choice as therapeutics compared to small molecules and antibody (4,5). Over the last decade, peptide-based therapeutics have revolutionized the pharmaceutical market (4). The number of approved peptide-based drugs has been increasing from the last few decades (4), which reflects the potential of peptides as therapeutics.

Cancer is one of the diseases where peptide therapeutics have been extensively explored over the years (6–10). ACPs are small peptides (less than 50 amino acids), and are often cationic in nature (consisting of basic and hydrophobic residues) (11). Most of the ACPs are derived from antimicrobial peptides (AMPs) and thus AMPs and most ACPs have similar characteristics (11,12). Similar to bacterial cells, surface of cancer cells is also negatively charged and due to this, many AMPs show broad-spectrum toxicity against both bacteria and cancer cells. The electrostatic interaction of ACPs with negatively charged components of plasma membrane of cancer cells is believed to play a crucial role in the cancer-selective toxicity of ACPs (13). Based on the cytotoxicity profiles, ACPs can be grouped in two main categories (14). First category consists of ACPs, which are toxic to bacterial and cancer cells but not toxic to normal cells. Second category includes ACPs, which are toxic to all, bacterial, cancer and normal cells (14). Though ACPs have been studied extensively, the mechanism of ACPs is not fully understood so far. Many studies have concluded that most ACPs, like AMPs, exhibit membrane-lytic mode of action (12–14). However, few ACPs have shown other modes of action also like induction of apoptosis via disruption of mitochondrial membrane. It has been reported that the selective toxicity of many ACPs toward cancer cells is attributed to the differences in the lipid content and other components

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

of biological membranes between normal and cancer cells (11,12).

Despite the huge therapeutic importance of ACPs, to date, no dedicated repository of ACPs and anticancer proteins has been developed as yet. A few AMPs databases like Antimicrobial Peptide Database V2 (15), Collection of Antimicrobial Peptides Database (16) and Database of Anuran Defense Peptides (17) contain information on anticancer activities of some AMPs, however, the information related to ACPs is not comprehensive. Last decade witnessed an exponential growth in the ACP-based research, which led to the discovery of hundreds of novel ACPs and their derivatives. Many of these ACPs have shown promising results in various pre-clinical studies and few ACP-based formulations have already reached clinical trials (3). This important information is scattered in the literature and thus very difficult to access. In order to assist the scientific community engaged in developing anticancer drugs, we have made an attempt to collect and compile all the scattered information related to ACPs and anticancer proteins and developed a repository-CancerPPD. We believe that CancerPPD will be helpful for both bioinformatics and experimental researchers working in the field of ACP-based therapeutics.

## MATERIALS AND METHODS

### Data collection and compilation

In order to develop a comprehensive information resource on ACPs and proteins, an extensive search was carried out to collect information on ACPs and anticancer proteins. For this, first research articles and patents providing information related to ACPs were extracted from various search engines like PubMed, Google scholar and Patent lens. Specific searches were carried out using a combination of keywords like 'ACPs', 'antitumor peptides', 'anti-angiogenic peptides', 'anti-metastatic peptides' and 'host defense peptides'. This exhaustive search yielded around 750 research articles and 30 patents. From these articles and patents, only experimentally verified ACPs and other relevant experimental information were extracted manually. In addition to ACPs, information related to anticancer proteins was also extracted from UniProt and PubMed. For this, text search was carried out in PubMed and UniProt using keywords 'anticancer protein' OR 'antitumor protein'. Finally, 624 ACPs and 121 anticancer proteins along with detailed information like nature of peptides, origin of peptides, sequence, modifications, assay types, cell line tested, etc. were compiled systematically. Since we did not want to lose any information, multiple entries for a single peptide have been made if identical peptide has been tested on different cell lines or it has been evaluated by different assays. Therefore, total entries in CancerPPD are 3491.

### Database architecture and web interface development

After the collection and compilation of all the information, the database was launched using Apache HTTP server on Linux Platform. MySQL an object-relational database management system was used to manage all data in the back-end. It provides commands to retrieve and store the data into the database. HTML, PHP and JAVA scripts

were used to improve the front-end web interface. All common gateway interface and database interfacing scripts were written in the PHP and Perl programming language. The architecture of CancerPPD is shown in Figure 1.

## DATABASE CONTENT

CancerPPD contains two types of information; primary and secondary information. Primary information was manually curated from literature and consists of various fields, including (i) PMID, (ii) peptide sequence, (iii) name of the peptide, (iv) length of the peptide, (v) configuration (linear or cyclic), (vi) chirality (L/D/Mix), (vii) chemical modification, (viii) N-terminal modification, (ix) C-terminal modification, (x) origin of the peptide, (xi) anticancer activity of peptide, (xii) tested cell lines, (xiii) assay types, (xiv) test times, (xv) cancer types and (xvi) target tissues.

Secondary information, like tertiary structures and simplified molecular-input line-entry system (SMILES), is derived from the primary information. In order to make it a comprehensive resource, we have compiled structures and SMILES of all the peptides. To provide structural information, all the peptides were searched and mapped on the Protein Data Bank (PDB) (18) sequences and if an exact match was obtained, we assigned the same structure to that peptide. A total of 32 peptide structures were obtained from PDB. Those peptides, which failed to map on PDB sequences, were predicted using PEPstr algorithm (19), which is the state-of-the-art algorithm for predicting the tertiary structure of peptides. Briefly, PEPstr uses PSIPRED (20) and BetaTurns (21) to predict secondary structure and types of beta-turns, respectively. It then uses the ideal dihedral angle of these intermediate predicted states as restraints to make an initial structure. Following energy minimization and short molecular dynamics of the initial structure using AMBER (22), the final structure is given as the predicted peptide tertiary structure. PEPstr handles peptides with only natural residues ranging in length from 7 to 25 residues. In the present work, we relaxed the length criteria from 5 to 40 residues. Using PEPstr, we were able to predict the structure of 211 peptides. A total of 9 sequences with lengths of >40 residues each, were treated as proteins and their structure was predicted using I-TASSER (23,24) server, which was among the best template-based protein structure prediction servers in CASP10 (25) assessment. To handle three peptides with length less than five residues, we generated the initial structure with linear conformation using phi and psi dihedral angles of 180°. Molecular dynamics was performed on the initial structure using AMBER and the structure having lowest energy in the trajectory was finally considered as the final predicted structure of the peptide.

Many ACPs also contain chemically modified residues and, therefore, cannot be handled directly by PEPstr. The unavailability of any algorithm (either online or standalone) for handling non-natural or chemically modified residues, prompted us to extend the PEPstr algorithm to handle these modifications for the prediction of peptide structures. Terminal modifications (acetylation/amidation), N to C cyclization, disulfide bridges between cysteine residues as well as the flipping of stereochemistry of a residue from L- to

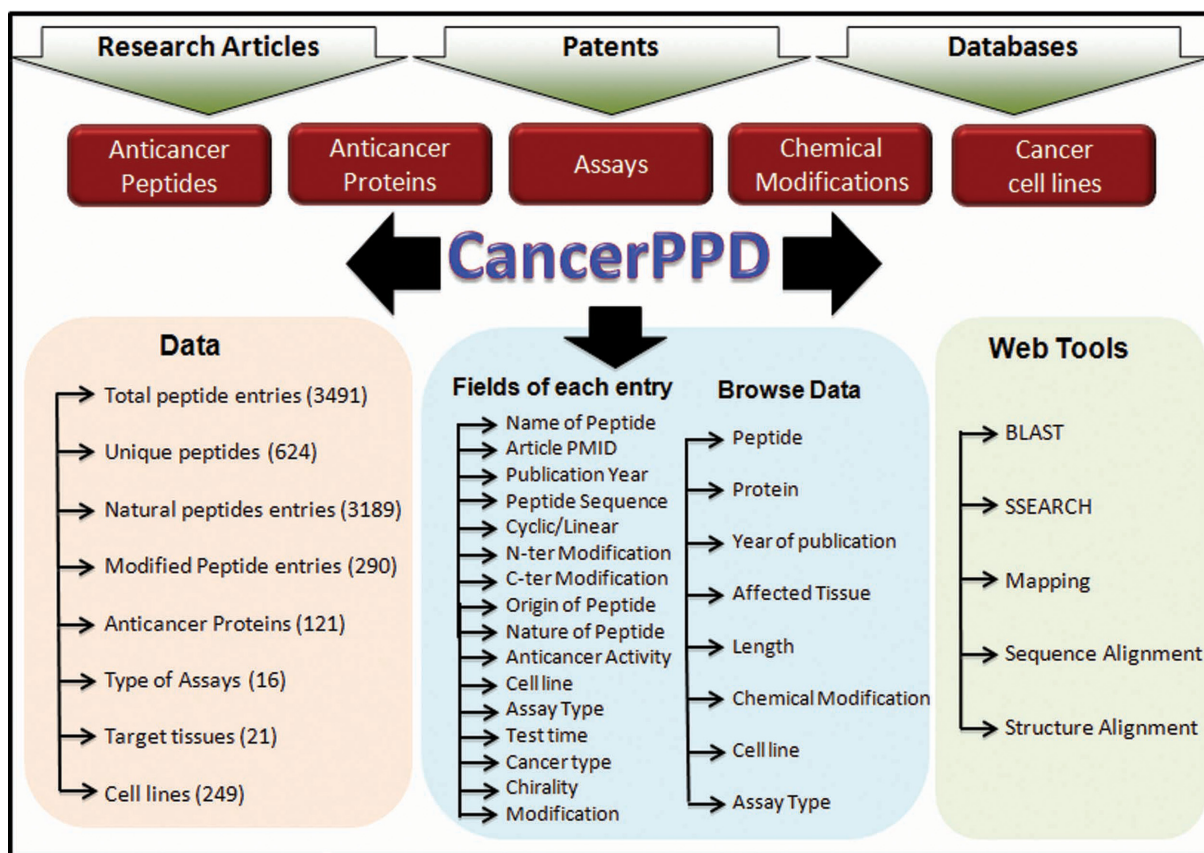


Figure 1. Architecture of CancerPPD.

D-form were carried out using inbuilt functions in AMBER11. Special force field libraries (26–28) available for standard molecular dynamics packages like AMBER and GROMACS (29) were used to incorporate other terminal modifications (beta-alanine, hydroxylation) or chemical modifications (ornithine, naphthyl alanine, etc.). In this way, a total of 617 peptide structures were obtained. Due to the lack of force field libraries, the structure of a few peptides having complex modifications was not predicted.

Although the PEPstr algorithm incorporated very short molecular dynamics (25 picoseconds) on the initial structure, we extended the simulation of all structures to 1 nanosecond. After obtaining the predicted structures of the peptides, we assigned the eight types of secondary structure (H = alpha helix; B = beta-bridge; E = extended strand; G = 3/10 helix; I = pi helix; T = turn; S = bend; C = loop) using DSSP software (30). DSSP takes input in PDB file format and assigns the secondary structure based on the hydrogen bonding patterns and geometrical features between amino acids. We also represent the predicted structures in SMILES format. Open Babel software (31) was used to convert the 3D structures in 2D SMILES notation.

## RESULTS

### Implementation

A user-friendly web interface (Figure 2) has been designed to query the database. Various tools have been integrated,

which facilitate data retrieval, search and analysis conveniently. Following is the brief description of various options available in CancerPPD.

### Data retrieval tools

*Simple search.* This option provides basic facility to retrieve data from the database. It allows users to perform keyword search on any field of the database like PMID, origin of peptide, cancer cell lines, nature of peptide. Users can select different fields to be displayed. The keyword should be without spaces.

*Complex query.* This option is for the users who wish to perform complex search to extract the desired information from CancerPPD. It offers a multiple query system, by default it allows to perform four queries at a time, and the user can perform keyword search on any selected field. The server provides facility to use standard logical operators (e.g. =, >, < and LIKE). Users can combine output of different queries using operators like 'AND and OR'. It also has the option to add or remove queries to be executed.

### Peptide sequence search

This option provides facility to search a given peptide sequence against sequences of all peptides available in CancerPPD. It provides two options called exact and substring

# CancerPPD

## Database of Anticancer Peptides & Proteins

Home
Information
Data Submission
Developers
Contact
Assistance
Related Databases

The Mobile Version of CancerPPD

Search

- Basic
- Conditional
- Peptide
- SMILES

Browse

- Protein wise
- Peptides wise
- Tissue wise
- Cell Line wise
- Year-wise
- Assay Type
- Length
- Modifications

Tools

- BLAST
- Smith-Waterman
- Mapping
- Sequence Alignment
- Structure Alignment

Download

- Sequence
- Structure
- Reference
- Acknowledgement

**CancerPPD** is a unique resource of its kind, which provides detailed information related to experimentally verified anticancer peptides (ACPs) and proteins. Data was curated manually from both published articles, patents as well as from other repositories. Since structures play important roles in the anticancer activity, we have predicted tertiary structures of anticancer peptides using state-of-art method PEPstr and secondary structure states are assigned using DSSP. The important feature of cancerPPD is that it also provides information related to various chemical modifications like non-natural, D-amino acids, modified-amino acid like ornithine. The database is cross-linked with various other related resources in order to provide comprehensive information related to ACPs.

Total Peptide Entries : 3491
Total Cell Lines : 249
Total Tissue Types : 21

### Mechanisms of ACP's action

**Key Features**

- (1) **Data Retrieval:** Tools for data fetching like simple, and advanced search were integrated. In addition, to search peptides, peptide search and SMILES search options are also provided.
- (2) **Data Analysis:** BLAST, Smith-waterman, sequence and structure mapping tools are incorporated, which facilitate the similarity-based search.
- (3) **Browsing:** Various browsing fields have been provided, which facilitate the data browsing in a very convenient way.
- (4) **SMILES and Structures:** An important feature of CancerPPD is that it contains information of ACPs in SMILES format. In addition, predicted tertiary structures of all ACPs are available which makes it comprehensive resource.

Figure 2. Schematic representation of CancerPPD web interface.

search. In case of exact search, the server extracts those peptides from the database, which have identical amino acid sequence with the query peptide. In case of substring search, the server extracts those peptide sequences that contain amino acids present in the query peptide.

**SMILES search.** The above options allow users to search the peptide sequences at amino acid level. Some time, researchers wish to understand whether a particular atom or bond or group is more frequent in ACPs. In order to assist users and to understand the property of ACPs at atom/bond level, we have maintained the structures of peptides in SMILES format. Server facilitates users to search their chemicals in SMILES format in the database search.

### Browsing tools

We have developed a browsing facility, which facilitates the retrieval of information in a classified form. A user-friendly interface has been integrated for browsing the ACPs on the major fields that include ACPs, anticancer proteins, type of cell line tested, year of discovery, assay classes and peptide length. Users can browse the ACPs and anticancer proteins using these fields. In addition, users can browse ACPs based on their chemical modifications like N-terminal, C-terminal modifications. In CancerPPD, information of 249 types of cancer cell lines originated from around 21 types of tissues has been compiled. Users can browse ACPs evaluated against a particular type of cancer cell line. Further, to make it a more informative resource, cell lines were linked with Catalogue of Somatic Mutations (COSMIC) (32) and Cancer Cell Line Encyclopedia (CCLE) (33).

### Analysis tools

CancerPPD integrates various web-based tools for performing various analyses like sequence similarity search, multiple sequence alignment and structure alignment. We have integrated Basic Local Alignment Search Tool (BLAST) (34) in CancerPPD that allows users to perform BLAST search against peptides in the database. This allows users to identify ACPs in the database that have high sequence similarity with query peptide sequence. Additionally, we have also integrated another similarity search tool to perform sequence similarity based on Smith–Waterman algorithm (35). The peptide-mapping tool allows users to identify ACPs within their proteins of interest. The server searches for ACPs and maps them on the query protein submitted by the user.

### Data statistics

CancerPPD consists of 3491 peptide entries corresponding to 624 unique peptides. While searching for the ACPs in the literature, it was noted that most ACPs have been tested on various cancer cell lines showing different  $IC_{50}$  values. Therefore, to secure all this information, multiple entries of a single ACP have been made if the identical peptide has been found to be tested on different cell lines (e.g. HeLa, A549, MCF-7, etc.) or has been tested by different assays (e.g. MTT, WST-1, LDH, etc.). The information of cancer

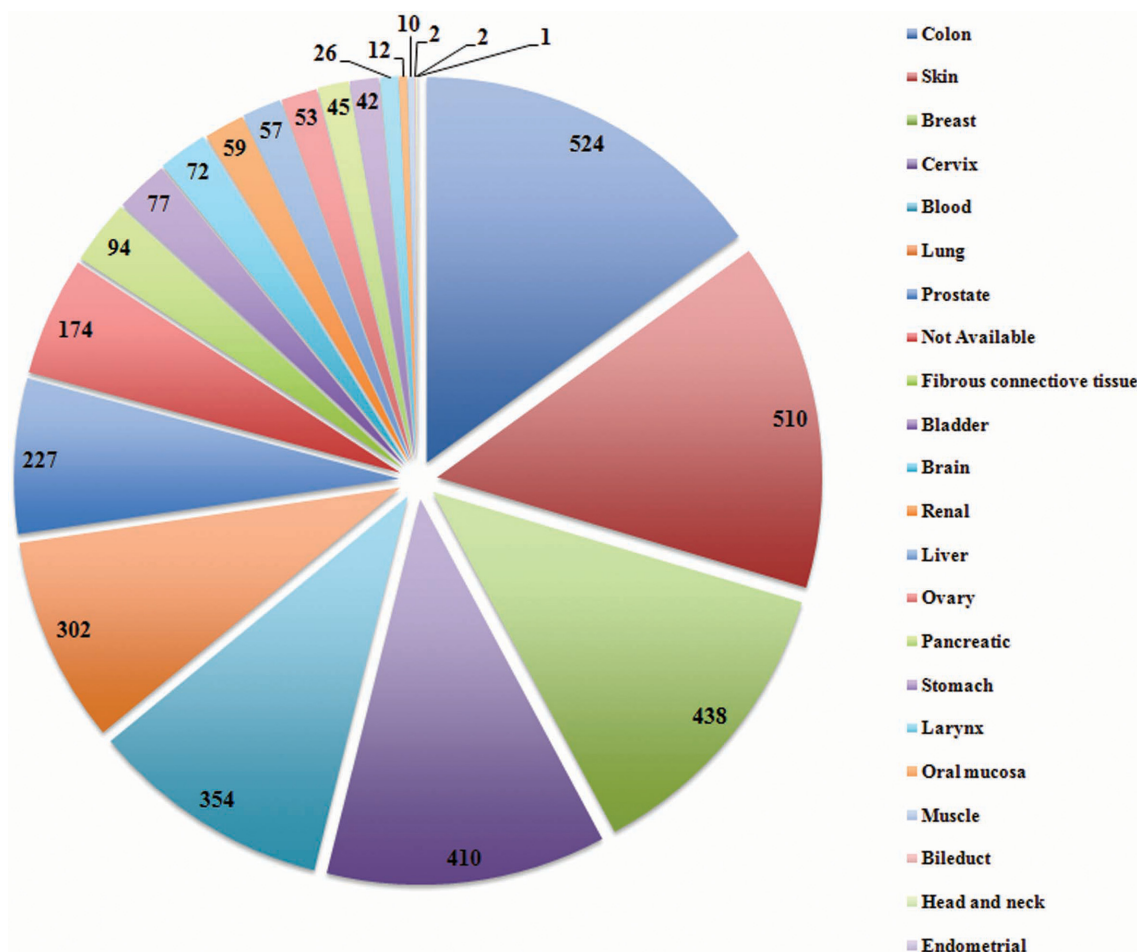
cell lines, which have been used as model system to evaluate the anticancer activity of ACPs is very important and thus information of total 249 cancer cell lines corresponding to 21 tissue types (Figure 3) was compiled and linked with various databases like CCLE and COSMIC. Since the stability of the peptide is one of the major concerns while developing therapeutic peptides, many ACPs and their analogs with different chemical modifications/ non-natural amino acids have been designed and evaluated for their anticancer activity. This information has also been compiled and total 290 peptide entries have been made, which have different chemical modifications. Apart from this, entries of peptide having L-amino acids (3274), D-amino acids (26) and both L- and D-amino acids (178) have been compiled. The most common assay to determine the anticancer activity of ACPs is the cell viability assay, which measure the viability of cells using various substrates like MTT. However, various other assays have also been reported in the literature for determining the anticancer activity of these peptides. The current version of CancerPPD holds information of 16 different types of assay used to test the anticancer potency of peptides.

### DISCUSSION

The therapeutic peptide market emerged almost 40 years back in 1970s. Since then peptides have not been very popular as drug candidates but it has only been in the recent past that the pharmaceutical industries have shown an interest in therapeutic peptides and made a heavy investment in the peptide-based drugs (3,36). This renewal of interest in therapeutic peptides could be due to the various limitations of conventional drugs, including frequent development of drug resistance, non-specificity, poor delivery, etc.

The present manuscript describes a repository of ACPs named CancerPPD, which is an important and a much-needed resource. ACPs belong to an important class of therapeutic peptides, which have received a significant attention over the years. The ability of many ACPs to selectively kill cancer cells without affecting normal cells makes them an attractive alternative candidate for cancer therapy (3,13). The last decade has seen a progressive growth in peptide-based research, particularly in therapeutic peptides, which is exemplified by hundreds of research articles and by the development of the various databases of therapeutic peptides, including CPPsite (37), TumorHope (38), Hemolytik (39), ParaPep (40), Brainpeps (41), Quorumpeps (42), etc. The development of CancerPPD will be an important addition to this elite group of knowledge resources.

Over the last decade, huge data on ACPs has been generated and, therefore, a systematic cataloging of this data will be important to understand the properties of ACPs and to delineate the features responsible for anticancer activity of these peptides. This analysis will further be helpful to design and predict better ACPs. Therefore, CancerPPD has been built with an aim to provide comprehensive information related to ACPs. Apart from ACP sequences, CancerPPD contains predicted tertiary structures of ACPs and ACPs in SMILES format. Users can make the best use of CancerPPD in the following ways: (i) users can search for the best ACPs, (ii) users can also search whether their peptide of interest is already exists in the CancerPPD or not, (iii) users



**Figure 3.** Representation of ACPs (entries) evaluated against various cancers cell lines originated from different tissues.

can analyze the extent to which their peptides are similar to the existing ACPs, (iv) users can map their query sequences as well as structures of their peptides of interest with structure of any ACP available in CancerPPD, (iv) structures available in CancerPPD will be used for docking and various membrane simulations studies, (v) CancerPPD offers a huge and latest data set of ACPs, which can be used for development of various prediction methods for ACPs and (vi) the SMILES of ACPs will be used to develop QSAR models. We hope that CancerPPD will be a useful resource for researchers working in the area of cancer therapeutics.

### LIMITATIONS AND UPDATE OF CancerPPD

In CancerPPD, we have stored the predicted tertiary structures of most of the ACPs along with the peptides with modified amino acids. However, one of the limitations is that the tertiary structures of few peptides have not been predicted due to the complex chemical modifications like lanthionine bridges. The force field libraries for such kind of residue modifications are presently not available. In the future, we will try to predict the structure of such peptides.

One critical task in the field of database development is to timely update it with newly generated information. Therefore, in this database, we have integrated an updating facil-

ity. Scientific community may submit ACPs to CancerPPD using online submission form. Our team will update the database regularly from online submission as well as from newly published literature.

### AVAILABILITY

CancerPPD is freely available at <http://crdd.osdd.net/raghava/cancerppd/>. The mobile version of this database can be accessed at <http://crdd.osdd.net/raghava/cancerppd/mobile/cancerppd/>.

### FUNDING

Council of Scientific and Industrial Research [project Open Source Drug discovery and GENESIS BSC0121]; Department of Biotechnology [BTISNET], Government of India for financial support. Funding for open access charge: Council of Scientific and Industrial Research.

*Conflict of interest statement.* None declared.

### REFERENCES

1. Siegel, R., Naishadham, D. and Jemal, A. (2013) Cancer statistics, 2013. *CA Cancer J. Clin.*, **63**, 11–30.

2. 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.
3. Thundimadathil, J. (2012) Cancer treatment using peptides: current therapies and future prospects. *J. Amino Acids*, **2012**, 967347.
4. Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatsky, M. (2010) Synthetic therapeutic peptides: science and market. *Drug Discov. Today*, **15**, 40–56.
5. Otvos, L. Jr. (2008) Peptide-based drug design: here and now. *Methods Mol. Biol.*, **494**, 1–8.
6. Barras, D. and Widmann, C. (2011) Promises of apoptosis-inducing peptides in cancer therapeutics. *Curr. Pharm. Biotechnol.*, **12**, 1153–1165.
7. Li, Z.J. and Cho, C.H. (2012) Peptides as targeting probes against tumor vasculature for diagnosis and drug delivery. *J. Transl. Med.*, **10** (Suppl. 1), S1.
8. Boohaker, R.J., Lee, M.W., Vishnubhotla, P., Perez, J.M. and Khaled, A.R. (2012) The use of therapeutic peptides to target and to kill cancer cells. *Curr. Med. Chem.*, **19**, 3794–3804.
9. Shapira, S., Fokra, A., Arber, N. and Kraus, S. (2014) Peptides for diagnosis and treatment of colorectal cancer. *Curr. Med. Chem.*, **21**, 2410–2416.
10. Gautam, A., Kapoor, P., Chaudhary, K., Kumar, R., Open, Source Drug Discovery C. and Raghava, G.P. (2014) Tumor homing peptides as molecular probes for cancer therapeutics, diagnostics and theranostics. *Curr. Med. Chem.*, **21**, 2367–2391.
11. Gaspar, D., Veiga, A.S. and Castanho, M.A. (2013) From antimicrobial to anticancer peptides. A review. *Front Microbiol.*, **4**, 294.
12. Hoskin, D.W. and Ramamoorthy, A. (2008) Studies on anticancer activities of antimicrobial peptides. *Biochim. Biophys. Acta*, **1778**, 357–375.
13. Mader, J.S. and Hoskin, D.W. (2006) Cationic antimicrobial peptides as novel cytotoxic agents for cancer treatment. *Expert Opin. Investig. Drugs*, **15**, 933–946.
14. Schweizer, F. (2009) Cationic amphiphilic peptides with cancer-selective toxicity. *Eur. J. Pharmacol.*, **625**, 190–194.
15. 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.
16. Wagh, F.H., Gopi, L., Barai, R.S., Ramteke, P., Nizami, B. and Iducila-Thomas, S. (2014) CAMP: collection of sequences and structures of antimicrobial peptides. *Nucleic Acids Res.*, **42**, D1154–D1158.
17. Novkovic, M., Simunic, J., Bojovic, V., Tossi, A. and Juretic, D. (2012) DADP: the database of anuran defense peptides. *Bioinformatics*, **28**, 1406–1407.
18. Rose, P.W., Bi, C., Bluhm, W.F., Christie, C.H., Dimitropoulos, D., Dutta, S., Green, R.K., Goodsell, D.S., Prlic, A., Quesada, M. *et al.* (2013) The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.*, **41**, D475–D482.
19. 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.
20. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
21. Kaur, H. and Raghava, G.P. (2004) A neural network method for prediction of beta-turn types in proteins using evolutionary information. *Bioinformatics*, **20**, 2751–2758.
22. Case, D.A., Cheatham, T.E. 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M. Jr., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
23. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinform.*, **9**, 40.
24. Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.*, **5**, 725–738.
25. Huang, Y.J., Mao, B., Aramini, J.M. and Montelione, G.T. (2014) Assessment of template-based protein structure predictions in CASP10. *Proteins*, **82** (Suppl. 2), 43–56.
26. 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 Synth. Biol.*, doi:10.1021/sb400168u.
27. Khoury, G.A., Thompson, J.P., Smadbeck, J., Kieslich, C.A. and Floudas, C.A. (2013) Forcefield\_PTMT: charge and AMBER forcefield parameters for frequently occurring post-translational modifications. *J. Chem. Theory Comput.*, **9**, 5653–5674.
28. Gfeller, D., Michielin, O. and Zoete, V. (2013) SwissSidechain: a molecular and structural database of non-natural sidechains. *Nucleic Acids Res.*, **41**, D327–D332.
29. Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., van der Spoel, D. *et al.* (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**, 845–854.
30. Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
31. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R. (2011) Open Babel: an open chemical toolbox. *J. Cheminform.*, **3**, 33.
32. Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A. *et al.* (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.*, **39**, D945–D950.
33. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D. *et al.* (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, **483**, 603–607.
34. 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.
35. Pearson, W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.*, **132**, 185–219.
36. 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.
37. 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.
38. 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.
39. 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.
40. 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**, doi:10.1093/database/bau051.
41. 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.
42. 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.