

RESEARCH ARTICLE

THPdb: Database of FDA-approved peptide and protein therapeutics

Salman Sadullah Usmani¹, Gursimran Bedi¹, Jesse S. Samuel², Sandeep Singh, Sourav Kalra, Pawan Kumar, Anjuman Arora Ahuja, Meenu Sharma, Ankur Gautam*, Gajendra P. S. Raghava^{1*}

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

✉ These authors contributed equally to this work.

✉ Current address: Centre for Computational Biology, Indraprastha Institute of Information Technology (IIIT) Okhla Phase III, New Delhi, India

* raghava@imtech.res.in, raghava@iiitd.ac.in (GPSR); ankurgautam@imtech.res.in (AG)



OPEN ACCESS

Citation: Usmani SS, Bedi G, Samuel JS, Singh S, Kalra S, Kumar P, et al. (2017) THPdb: Database of FDA-approved peptide and protein therapeutics. *PLoS ONE* 12(7): e0181748. <https://doi.org/10.1371/journal.pone.0181748>

Editor: Qing-Xiang Amy Sang, Florida State University, UNITED STATES

Received: December 7, 2016

Accepted: July 6, 2017

Published: July 31, 2017

Copyright: © 2017 Usmani et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: We have upload this data set to a stable, public data repository "figshare" which could be accessed at following URL: https://figshare.com/articles/THPdb_Database_of_FDA_Approved_Peptide_and_Protein_Therapeutics/5198005.

Funding: Authors are thankful to funding agencies, Council of Scientific and Industrial Research (CSIR) (project Open Source Drug discovery and GENESIS BSC0121), and Department of Biotechnology (project BTISNET), Govt. of India for financial support and fellowships.

Abstract

THPdb (<http://crdd.osdd.net/raghava/thpdb/>) is a manually curated repository of Food and Drug Administration (FDA) approved therapeutic peptides and proteins. The information in THPdb has been compiled from 985 research publications, 70 patents and other resources like DrugBank. The current version of the database holds a total of 852 entries, providing comprehensive information on 239 US-FDA approved therapeutic peptides and proteins and their 380 drug variants. The information on each peptide and protein includes their sequences, chemical properties, composition, disease area, mode of activity, physical appearance, category or pharmacological class, pharmacodynamics, route of administration, toxicity, target of activity, etc. In addition, we have annotated the structure of most of the protein and peptides. A number of user-friendly tools have been integrated to facilitate easy browsing and data analysis. To assist scientific community, a web interface and mobile App have also been developed.

Introduction

With the significant advancements in biologics and biopharmaceutical over the years, peptides and proteins have emerged with a host of new applications in the diagnostic as well as the therapeutic sector [1, 2]. As per the current calculations, the market for peptide and protein drugs is estimated around 10% of the entire pharmaceutical market and will make up an even larger proportion of the market in the future [3, 4]. Since early 1980s, a total 239 therapeutic proteins and peptides are approved for clinical use by US-FDA [5].

Since the introduction of first recombinant protein therapeutic, human insulin [6], proteins have emerged as a major new class of therapeutics with nearly 380-marketed pharmaceutical products. Proteins as therapeutics have drawn a lot of scientific attention because they are part of many receptors and channels in the membranes, which facilitate transport of molecules within the body [7]. Despite their involvement in various biochemical reactions, proteins also elicit high specificity with the molecular target [7]. On the other hand, peptides have also been

Competing interests: The authors have declared that no competing interests exist.

used widely as therapeutics in cancer, diabetes, infectious diseases and many other disorders that have long been studied for a solution [8, 9]. An advantage of peptides over other drugs is that they are highly versatile, providing a wide variety of pharmaceutical targets, high specificity and show low toxicity levels [5, 9]. A total 60 peptide-based drugs are already in the market and several other therapeutic peptides are currently being evaluated in different phases of clinical trials [5].

Despite the several advantages, there are a few drawbacks associated with these protein and peptide-based therapeutics. Limitations like low solubility, proteolytic degradation, relatively short circulating half-life, physiochemical instability and immunogenicity restrict their ease of administration and compliance at the user end [1, 4, 10]. They generally have high molecular weights, low lipophilicity and presence of some charged functional groups, which tend to hamper their absorption at various sites of the gastro-intestinal tract (GIT), resulting in their low bioavailability [4]. Several strategies like glycosylation, PEGylation, manipulation of amino acid sequence, conjugation to serum albumin have emerged in order to improve the pharmacokinetic and pharmacodynamics properties as well as to decrease the immunogenicity and proteolytic cleavage [1, 5]. With the advent of new technologies, incorporation of non-natural amino acids and pseudo peptide bonds also provide wider chemical diversity with diversified potentials and making it economically more viable [4, 11].

Over the last decade, a plethora of databases encompassing information on protein and peptides with different therapeutic functionalities [12–19] have been developed, reflecting the increased interest for proteins and peptides as therapeutics among the scientific community. A substantial interest has also been shown towards peptide-based subunit vaccine and immunotherapeutic [20]. The information on US-FDA approved protein and peptide therapeutics along with their pharmacokinetic / pharmacodynamic properties, their advantages, chemical modifications, and their limitations are very important. However, this information is not easily accessible and it is scattered in the literature. To the best of authors' knowledge, to date, no single freely available platform exists which is totally dedicated to US-FDA approved protein and peptide therapeutics. Therefore, keeping the above facts in mind, we have developed THPdb, which is a comprehensive resource of all US-FDA approved protein and peptide therapeutics along with their corresponding drug variants available in the market. We anticipate that the information available in the THPdb will certainly be very helpful to the researchers working in the field of peptide and protein-based drug discovery.

Methods and materials

Data collection and compilation

In order to collect and compile the latest information on US-FDA approved protein and peptide therapeutics; first, we searched DrugBank [21] using keyword “biotech drugs” with “Approved” and “Investigational” filters. This search resulted into 369 biotechnology-based drugs on 20th Apr 2017. Among these, only 239 were peptides and proteins, which were subsequently shortlisted for inclusion in the THPdb. Further, in order to provide the complete profile and comprehensive information (regarding the mechanism of action, physical form, route of administration, category of activity, interaction with other drugs, recommended dosages, contraindications, side effects and the explored therapeutic area of these shortlisted peptides and proteins), specific searches were carried out in the PubMed and patent databases using strings ‘therapeutic proteins’, ‘therapeutic peptides’, and using the name of the individual protein and peptide. These searches resulted into a total 985 publications and 70 patents. The relevant information under various heads was collected and compiled manually from these articles.

In order to improve the pharmacokinetic parameters like the volume of distribution (Vd), clearance (CL), or the bioavailability (AUC), different modifications are being made to the existing drug molecules (drug variants). Here, in THPdb, each such modification with their brand names compiled in a separate entry. Therefore, multiple entries have been made if a drug molecule has been modified in different ways or it is being sold under different Brand names. Finally, a total 380 drug variants of these peptides and proteins have been systematically compiled in THPdb.

Database architecture and web interface

THPdb database has been built using a standard platform based on the Linux-Apache-MySQL-PHP (LAMP). Red Hat Linux (version 6.2) as the operating system, MySQL (version 14.12) for managing the data and Apache (version 2.2.17) as the HTTP server was used for developing this database. HTML5, PHP, JAVA scripts have been used for developing the mobile and tablet compatible front ends and MySQL for developing the back end. For the entire database interface and the common gateway, networks PHP and PERL coding have been used. The architecture of THPdb database is given in Fig 1.

Organization of data

Primary information. Information on protein and peptide therapeutics were obtained manually from different resources and stored as primary information. Each entry includes the following information: name of the therapeutic molecule, amino acid sequence or primary structure, molecular weight, chemical formula, half-life, melting point, hydrophobicity, iso-electric point, toxicity, metabolism, pharmacodynamics, mechanism of action, toxicity, indication/disease, metabolism, absorption, Vd, CL, categories, patent no.'s, date of issue and date of expiry of the patent, drug interactions and the targets. In addition, information of company name, brand name, formulation, recommended dosage, side effects, contraindications, and the route of administration have also been collected and stored as the primary data.

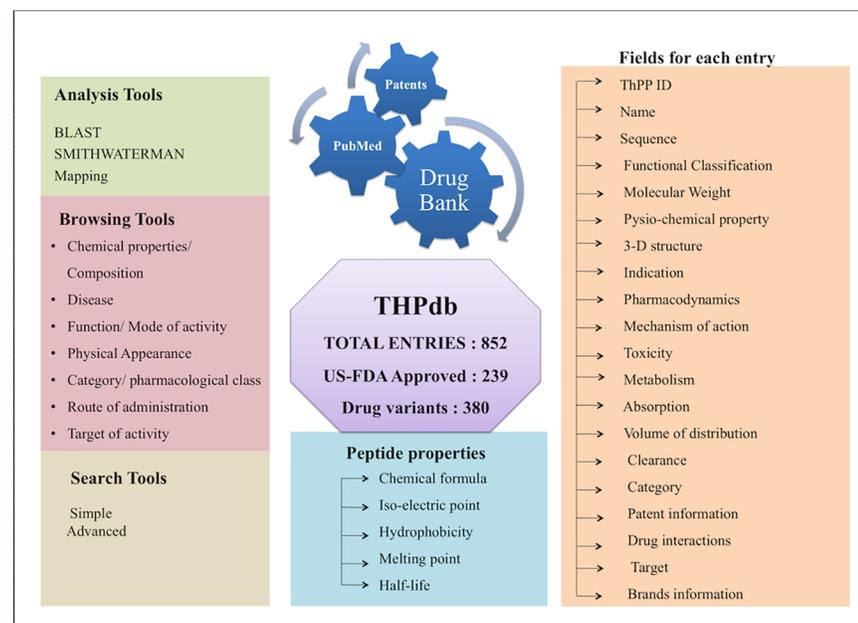


Fig 1. Schematic representation of THPdb architecture.

<https://doi.org/10.1371/journal.pone.0181748.g001>

Derived information. Apart from the primary information, information related to function/ activity of the peptides and proteins have also been compiled. Based on the role or mode of activity, therapeutic peptides and proteins are classified into four groups [7], Group I consist of the proteins with enzymatic or regulatory activity, Group II of those with specific targeting activity, Group III are protein vaccines, and Group IV consists of diagnostic agents. The proteins with enzymatic or regulatory activity are further sub-divided into three categories: (Ia) those replacing a deficient or abnormal protein, (Ib) those augmenting an existing pathway, and (Ic) those providing a novel function or activity. Therapeutics with specific targeting activity being sub-divided into two categories: (IIa) peptides interfering with a molecule or organism, and (IIb) peptides delivering other compounds or proteins. Similarly, group III contains protein vaccines, which are sub-classified as (IIIa) vaccines, which protect against a deleterious foreign agent, (IIIb) which treats autoimmune diseases and (IIIc) treats cancer. The last category corresponds to proteins that can be used as diagnostic agents. This classification scheme is explained previously [7]. This information has been stored as secondary data (Fig 2).

To understand the nature of these therapeutic peptides and proteins, it is necessary to understand their physicochemical properties, which has been calculated by in-house PERL scripts that were used to compute amino acid composition and frequencies. All these properties are helpful in analyzing the trends and determining the amino acids dominating in such proteins and peptides. This physiochemical property based information is also stored as secondary data in the database for browsing of peptides.

Many properties exhibited by the therapeutic peptides and proteins are due to their structural characteristics. Thus, it is important to understand the relation between the structures of the biomolecules and its function. We have thus provided the structural annotation of each protein and peptide in our database using the following approaches. The structural information was first extracted from the DrugBank using the Protein Data Bank (PDB) entries [22] (if present) given in the DrugBank followed by downloading it from PDB. In cases, where no structural information was available in either DrugBank or PDB, we used bioinformatics tools

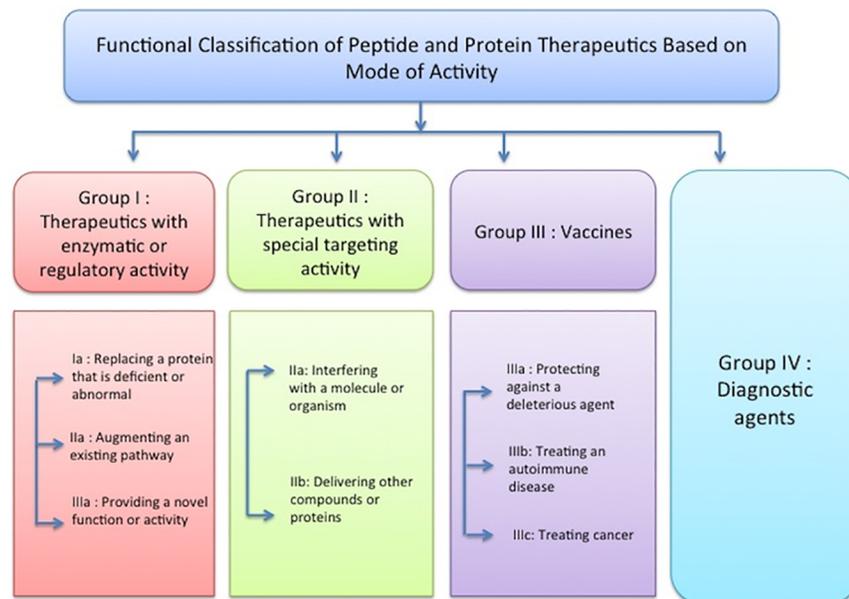


Fig 2. Overview of the function/mode of activity based classification of therapeutic proteins and peptides.

<https://doi.org/10.1371/journal.pone.0181748.g002>

for structure prediction. I-TASSER software [23] was used for predicting structure of proteins having more than 30 amino acids. In case, length of a peptide is 30 or less, then PEPstrMOD [2], which is updated version of PEPstr [24], has been used. If a protein or peptide has any modified residue (e.g. non-natural residues and capping of terminal residues), it was predicted using PEPstrMOD containing special force field libraries to handle modified residues.

Implementation of tools

Data retrieval or search tools. This module of THPdb has been designed to facilitate searching the data using simple and advanced search options. In simple search module, the user can submit the query against any field of the database such as name, sequence, category, classification based on their role/mode of activity, isoelectric point, indication etc. This option results into a customized output according to the search query. In the advanced search module, the user can submit multiple queries simultaneously with Boolean expressions (eg. AND, OR and NOT). Along with these tools and data, we have included a unique feature in our database, which are the downloadable self-explanatory power-point presentations. These presentations are freely available and can be downloaded. The user can use these for academic and research purposes.

Browsing. A user-friendly browsing interface has been developed to facilitate easy retrieval of the information from THPdb. We have stored physicochemical properties of each therapeutic peptide and protein such as the hydrophobicity, isoelectric point, melting point, molecular weight and half-life. Users can browse peptide and protein entries for the desired physicochemical properties.

To access the data in THPdb, various cross-linked browsable tables have been provided. The users can browse on seven major fields: (i) disease area of activity, (ii) number of amino acids in the protein or peptide, (iii) classification based on their function/mode of activity, (iv) physical appearance state (v) route of administration (vi) category/pharmacological class and (vii) target based classification.

Sequence alignment. In order to facilitate sequence similarity based search, we have integrated BLAST and Smith-Waterman algorithm. Users can submit their protein or peptide in FASTA format with default or desired parameters of BLAST [25]. The server performs BLAST search for the user's query sequence against the amino acid sequences or the primary structures of all the peptides and proteins in the database. Similarly, Smith-Waterman algorithm [26] has also been integrated. In addition to above similarity search facility, the server also allows sequence mapping based on identical residues. Sub-search and super-search have been integrated as two mapping options. By using sub-search module, user can map query peptide against all the peptides and proteins entered in THPdb, whereas super-search allows mapping of the protein sequence against THPdb and identification of segments that are identical to therapeutic peptides.

Mobile App

To assist the mobile users, we have developed a mobile App. THPdb (mobile app) is a comprehensive database application of US-FDA approved peptide and protein therapeutics developed to facilitate its smooth usage on mobile platforms. The current mobile version of THPdb compatible with android mobiles exploits python(v2.7.11) and kivy(v1.9.2). The App lists various disease types and a drop down menu of potential therapeutic peptides followed by detailed information of drugs for individual protein and peptide. The important information for individual entity has been sorted into individual panels called "peptide-card" which can be easily downloaded. Each peptide-card features/hosts information about the peptide/ protein

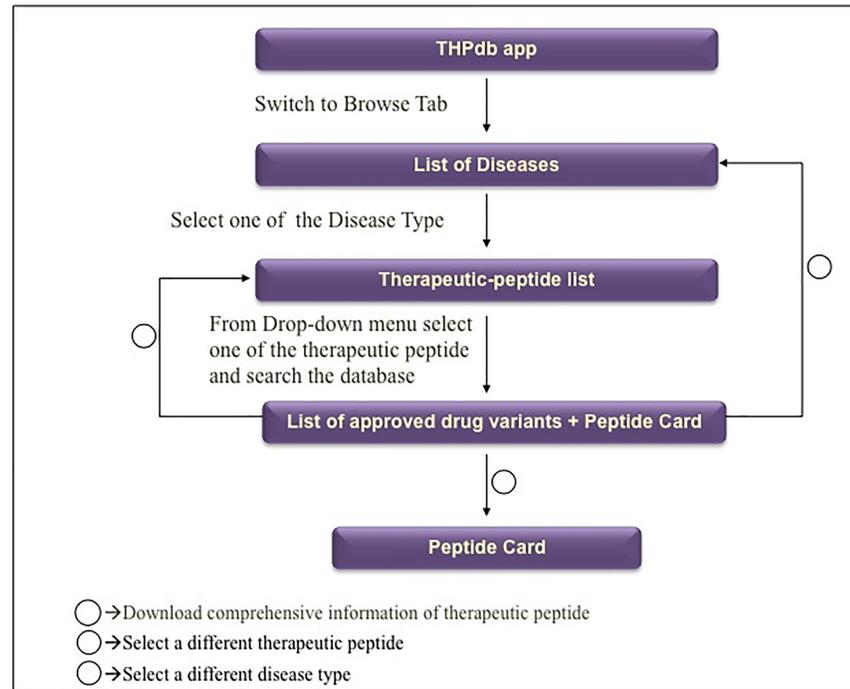


Fig 3. The flow diagram explains the work and design of THPdb mobile application.

<https://doi.org/10.1371/journal.pone.0181748.g003>

therapeutics, their description, sequence, indication, pharmacodynamics, mechanism of action, ADME information, toxicological information, metabolism, absorption, half life, clearance rate, patent information, drug-drug interactions, potential targets, physicochemical properties to name a few. The ‘THPdb’ mobile application is provided in standard “.apk” file format which can be downloaded from <http://crdd.osdd.net/raghava/thpdb/mobile.php>

App working outline. THPdb app is divided into various levels to provide an interactive outline of the database. As shown in Fig 3, database exploration begins with the selection of a disease type from the “Browse” tab. In the next level, a peptide of interest is searched among the list of US-FDA approved peptide and protein therapeutics for the selected diseases to obtain a list of potentially approved drug(s). The hit shares a handful of information like drug brands, their manufacturers, and route of administration of corresponding therapeutic peptide /protein. More comprehensive information of important fields can be retrieved by downloading the “Peptide Card” that is available at the bottom of the each peptide page.

Results

THPdb data statistics

The current version of THPdb holds a total of 852 entries providing information on 239 US-FDA approved peptide and protein drugs and their 380 drug variants. Corresponding drug variants of therapeutic proteins and peptides were classified on the basis of a classification presented by Leader et al. Of the 380 drug variants, 229 involved in regulatory and enzymatic activity followed by 78 belonging to the therapeutics with special targeting activity. A total of 58 drug variants belong to the vaccines category and 15 to the diagnostic agents.

In addition, these drug variants have also been grouped on the basis of disease in which they are being used for therapy. A total of 89 drug variants show activity in case of metabolic

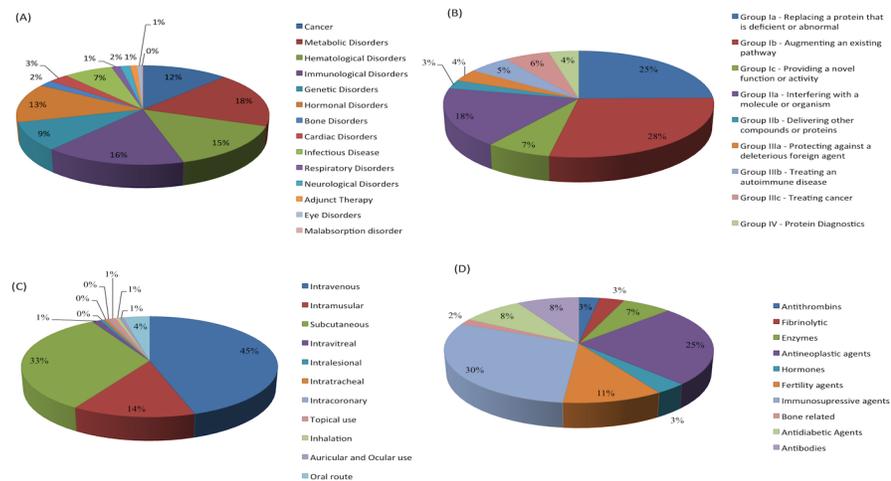


Fig 4. A schematic representation of distribution of therapeutic peptides and proteins based on disease, function/ mode of activity, route of administration, and pharmacological class.

<https://doi.org/10.1371/journal.pone.0181748.g004>

disorders, 80 have activity in the immunological disease area, 74 for hematological diseases, 61 in the cancer therapy, 63 in hormonal disorders, 46 variants useful for genetic disorders, 35 in infectious disease, 14 in cardiovascular disorders, 10 have the potential to cure bone disorders, 07 used in neurological disorder, 06 for respiratory disorder, 05 variants are given as adjunct, 03 in eye disorder, and 01 variant has been used in malabsorption disorder.

The routes by which these variants are being delivered have also been compiled systematically. Total 158 drugs by intravenous infusions, 116 drug variants are delivered by subcutaneous injections, 49 drugs by intramuscular route, 13 drugs via the oral route, 4 by intra-vitreal, 2 by intralesional, 2 by intratracheal, 1 by intracoronary, and 3 drug variants are being used externally as ointments. All these data statistics have been summarized in Fig 4.

In order to assist users, we have provided power-point presentations containing detailed information of each US-FDA approved protein and peptide therapeutic including their sequence or primary structure, molecular weight, chemical formula, description, indication, pharmacodynamics, mechanism of action and its interaction with other drugs. These presentations can be downloaded from the THPdb website.

Comparison with existing resources

Despite the fact that peptides and proteins are gaining tremendous scientific attention as therapeutics nowadays, no single repository exists till date which can provide comprehensive information on all US-FDA approved protein and peptide therapeutics. However, two databases, ChEMBL [27] and DrugBank provide the information on US FDA approved protein and peptide therapeutics to a certain extent but they don't cover all the US-FDA approved protein and peptide therapeutics and their drug variants. As of 1st May 2017, ChEMBL provides information of only 191 protein and peptide therapeutics and information of 48 protein and peptide therapeutics is missing. In addition, among these 191 protein and peptide therapeutics, ChEMBL, provides structural information of only 148 protein and peptide therapeutics. Similarly, in case of drug variants, ChEMBL provides information of 144 variants only. DrugBank store some information about protein and peptide therapeutics, but it doesn't give detailed description of corresponding drug variants of these peptide and protein therapeutics. In addition, DrugBank provides PDB ID of only 107 protein and peptide therapeutics, whereas

THPdb provide structural information of about 156 protein and peptide therapeutics. In order to overcome the limitations of the above resources, we have incorporated detailed information of all 239 protein and peptide therapeutics and their 380 variants. Besides this, comprehensive information (their sequence, physico-chemical properties, structure, pharmacokinetics, pharmacodynamics properties, etc) of each protein and peptide therapeutic and their drug variant was provided in the form of power-point presentations. User can freely download the power-point presentations. Thus, THPdb provides the latest and comprehensive information and is a complementary resource to the existing databases.

Discussion

Small peptides and proteins were never considered as the ideal therapeutic molecules, however, over the last few decades, a revival of interest in peptide and protein therapeutics has been seen among the pharmaceutical and biotech industries [1]. This can be evidenced by the increasing rate of peptides and proteins entry per year in the clinical trials [9]. In 2012 alone, six peptides were approved as drugs by US-FDA [11]. Despite the immense popularity of proteins and peptides as therapeutics, the information of all already US-FDA approved protein and peptide therapeutics is far from reach. Therefore, in the present study, an attempt has been made to provide a comprehensive insight on the US-FDA approved peptide and protein therapeutics including all the information on their physical, chemical and activity related characteristics. In THPdb, all this information has been compiled in a systematic manner, which will help the researchers to look into a variety of aspects of a particular protein or peptide simultaneously.

Development of a drug is a complex and time-consuming process. Even though peptides/proteins are highly active in their native form, they need to be modified in several ways in order to become a drug molecule. These series of changes make the lead molecule as a better therapeutic candidate and the prior knowledge of such changes or modifications will be very useful in developing future drugs. The peptide researchers may be interested to know the half-life, chemical modifications, immunogenicity, solubility, side effects, toxicity, pharmacodynamics, different formulations, dosage forms, and the different brand names under which these drugs are currently sold in the market. In this context, THPdb is very helpful in the development of novel therapeutic entities, particularly during the initial phases of drug discovery by understanding how the drawbacks have been overcome in US-FDA approved peptide and protein therapeutics. For example, the various modifications made to improve the therapeutic profiles of US-FDA approved proteins will give researchers many clues to improve the therapeutic profile of their lead molecules.

Utility of database

One of the major advantages of THPdb is that users can get the complete information on US-FDA approved protein and peptide therapeutics, which was not previously available on a single platform. THPdb provides its users with numerous services, like (i) users can check whether their peptides or proteins of interest has already been reported or not, and how different or similar their protein or peptide is with the existing approved therapeutics, (ii) what kind of modifications has been done to increase therapeutic competency of the therapeutics like half-life, rate of clearance, etc., and (iii) users can exploit the structural information of these peptides provided in THPdb for molecular dynamics or docking studies. In addition, a mobile application is also developed with the aim of making it easy for the researchers.

Besides this, THPdb also provides self-explanatory resource in power-point format, which will be very helpful for academicians, and researchers and we hope that it will accomplish our main aim behind this database *i.e* “academic enhancement” and “knowledge enrichment”.

A Query Submission Form

Type **Insulin**

Please paste/insert/type your query to be searched: (insulin)

Select Fields to be SEARCHED:

- Name [Leparudin]
- Chemical Formula [CO₂]
- Category [Hormone]
- Company [Eli Lilly]
- Functional Classification [Iib]
- Indication [cystic fibrosis]
- Sequence [PHWSVLLR]
- Isoelectric Point [4.58]
- Patent number [US7276477]
- Physical Appearance [Lyophilized]
- Brand [Enbrel]
- Route [Intravenous]

Please select field you wish to DISPLAY: (Max seven field be can selected once)

- Therapeutic peptide Name
- Chemical Formula
- Half life
- Pharmacodynamics
- Absorption
- Patent Number
- Target
- Prescribed for
- Route of administration
- Sequence
- Toxicity
- Clearance
- Date of Expiry
- Company
- Formulation
- Contraindication
- Molecular Weight
- Melting Point
- Indication
- Metabolism

B Results

ID	ThPPID	Peptide	Sequence
1205	Th1027	Insulin Regular	A-chain:GIVEQCCTSICSLYQLENYCN B-chain:FVNQHLCGSHLV Full view
1206	Th1027	Insulin Regular	A-chain:GIVEQCCTSICSLYQLENYCN, B-chain:FVNQHLCGSHLV Full view
1207	Th1027	Insulin Regular	A-chain:GIVEQCCTSICSLYQLENYCN, B-chain:FVNQHLCGSHLV Full view
1208	Th1027	Insulin Regular	A-chain:GIVEQCCTSICSLYQLENYCN, B-chain:FVNQHLCGSHLV Full view
1209	Th1027	Insulin Regular	A-chain:GIVEQCCTSICSLYQLENYCN, B-chain:FVNQHLCGSHLV Full view
1275	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1276	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1277	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1278	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1279	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1280	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1281	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view
1282	Th1040	Insulin Lispro	A chain:GIVEQCCTSICSLYQLENYCN;B chain:FVNQHLCGSHLV Full view

Total 52 entries consisting of 17 different Insulin drug variants

C 1634 details

Details of each entry

D

Get the structure

3-D Structure **Th1140 A chainTh1140 B chain**

Fig 5. Representative screen shots of THPdb demonstrating the flow of information after submission of a query in simple search page.

<https://doi.org/10.1371/journal.pone.0181748.g005>

Case study: How to retrieve information about “Insulin” from THPdb. Here we have shown step-by-step about what kind of information a user can fetch from THPdb. We have use Insulin as therapeutics in a case study to explain this. Below is the description:

If user wants to know about the different type of insulin drugs (drug variants) approved by US-FDA, user has to type Insulin in the space provided in the simple search option (Fig 5A) or alternatively user can use advance search tool as well. On simple search page, a number of fields are provided on which user can fetch the information. At a time, user can select a

maximum of 7 fields to be displayed. After selecting fields, user has to submit the query and search will end up with all insulin-based drugs and its variants available in the market under different brand names. For example, here, a total 52 entries comprising 17 insulin drugs and its variants are displayed along with information on selected fields (Fig 5B). User can also click on individual entry ID in order to obtain the detailed information for a particular insulin drug as shown in Fig 5C. By doing so, user can get the all possible information of that particular drug like sequence, physico-chemical properties, pharmacokinetics and pharmacodynamics, half-life, side-effects, route of administration, dosage form, brand name, etc. In addition, user can also get the structural information by clicking on 3-D structure option provided in the detailed information page Fig 5D.

Limitations and database update

While compiling the data under various categories for each peptide and protein, we faced some challenges, as the sequence of 57 of the approved peptides and proteins could not be obtained from any of the available sources. These are mainly the monoclonal antibodies and a few peptide antibiotics. This information will be updated in the database as soon as the new information is available in the literature.

Acknowledgments

Authors are thankful to funding agencies, Council of Scientific and Industrial Research (CSIR) (project Open Source Drug discovery and GENESIS BSC0121), and Department of Biotechnology (project BTISNET), Govt. of India for financial support and fellowships.

Author Contributions

Conceptualization: GPSR.

Data curation: SSU GB JSS SK AAA.

Formal analysis: SSU SS PK MS.

Investigation: SSU GB JSS SK AAA.

Methodology: SSU SS.

Project administration: GPSR.

Resources: GPSR.

Software: SSU SS PK MS.

Supervision: GPSR.

Validation: GPSR.

Visualization: SSU AG.

Writing – original draft: SSU AG.

Writing – review & editing: SSU AG GPSR.

References

1. Antosova Z, Mackova M, Kral V, Macek T. Therapeutic application of peptides and proteins: parenteral forever? *Trends Biotechnol.* 2009; 27(11):628–35. Epub 2009/09/22. <https://doi.org/10.1016/j.tibtech.2009.07.009> PMID: 19766335.

2. Singh S, Singh H, Tuknait A, Chaudhary K, Singh B, Kumaran S, et al. PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol Direct*. 2015; 10:73. Epub 2015/12/23. <https://doi.org/10.1186/s13062-015-0103-4> PMID: 26690490; PubMed Central PMCID: PMC4687368.
3. Craik DJ, Fairlie DP, Liras S, Price D. The future of peptide-based drugs. *Chem Biol Drug Des*. 2013; 81(1):136–47. Epub 2012/12/21. <https://doi.org/10.1111/cbdd.12055> PMID: 23253135.
4. Bruno BJ, Miller GD, Lim CS. Basics and recent advances in peptide and protein drug delivery. *Ther Deliv*. 2013; 4(11):1443–67. Epub 2013/11/16. <https://doi.org/10.4155/tde.13.104> PMID: 24228993; PubMed Central PMCID: PMC43956587.
5. Fosgerau K, Hoffmann T. Peptide therapeutics: current status and future directions. *Drug Discov Today*. 2015; 20(1):122–8. Epub 2014/12/03. <https://doi.org/10.1016/j.drudis.2014.10.003> PMID: 25450771.
6. Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, et al. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc Natl Acad Sci U S A*. 1979; 76(1):106–10. Epub 1979/01/01. PMID: 85300; PubMed Central PMCID: PMC4382885.
7. Leader B, Baca QJ, Golan DE. Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov*. 2008; 7(1):21–39. Epub 2007/12/22. <https://doi.org/10.1038/nrd2399> PMID: 18097458.
8. Boohaker RJ, Lee MW, Vishnubhotla P, Perez JM, Khaled AR. The use of therapeutic peptides to target and to kill cancer cells. *Curr Med Chem*. 2012; 19(22):3794–804. Epub 2012/06/26. PMID: 22725698; PubMed Central PMCID: PMC4537071.
9. Vlieghe P, Lisowski V, Martinez J, Khrestchatskiy M. Synthetic therapeutic peptides: science and market. *Drug Discov Today*. 2010; 15(1–2):40–56. Epub 2009/11/03. <https://doi.org/10.1016/j.drudis.2009.10.009> PMID: 19879957.
10. Otvos L Jr., Wade JD. Current challenges in peptide-based drug discovery. *Front Chem*. 2014; 2:62. Epub 2014/08/26. <https://doi.org/10.3389/fchem.2014.00062> PMID: 25152873; PubMed Central PMCID: PMC4126357.
11. Di L. Strategic approaches to optimizing peptide ADME properties. *AAPS J*. 2015; 17(1):134–43. Epub 2014/11/05. <https://doi.org/10.1208/s12248-014-9687-3> PMID: 25366889; PubMed Central PMCID: PMC4287298.
12. Agrawal P, Bhalla S, Usmani SS, Singh S, Chaudhary K, Raghava GP, et al. CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res*. 2016; 44(D1):D1098–103. Epub 2015/11/21. <https://doi.org/10.1093/nar/gkv1266> PMID: 26586798; PubMed Central PMCID: PMC4702894.
13. Singh S, Chaudhary K, Dhanda SK, Bhalla S, Usmani SS, Gautam A, et al. SATPdb: a database of structurally annotated therapeutic peptides. *Nucleic Acids Res*. 2016; 44(D1):D1119–26. Epub 2015/11/04. <https://doi.org/10.1093/nar/gkv1114> PMID: 26527728; PubMed Central PMCID: PMC4702810.
14. Kumar R, Chaudhary K, Sharma M, Nagpal G, Chauhan JS, Singh S, et al. AHTPDB: a comprehensive platform for analysis and presentation of antihypertensive peptides. *Nucleic Acids Res*. 2015; 43(Database issue):D956–62. Epub 2014/11/14. <https://doi.org/10.1093/nar/gku1141> PMID: 25392419; PubMed Central PMCID: PMC4383949.
15. Tyagi A, Tuknait A, Anand P, Gupta S, Sharma M, Mathur D, et al. CancerPPD: a database of anticancer peptides and proteins. *Nucleic Acids Res*. 2015; 43(Database issue):D837–43. Epub 2014/10/02. <https://doi.org/10.1093/nar/gku892> PMID: 25270878; PubMed Central PMCID: PMC4384006.
16. Van Dorpe S, Bronselaer A, Nielandt J, Stalmans S, Wynendaele E, Audenaert K, et al. Brainpeps: the blood-brain barrier peptide database. *Brain Struct Funct*. 2012; 217(3):687–718. Epub 2011/12/30. <https://doi.org/10.1007/s00429-011-0375-0> PMID: 22205159.
17. Wynendaele E, Bronselaer A, Nielandt J, D'Hondt M, Stalmans S, Bracke N, et al. Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. *Nucleic Acids Res*. 2013; 41(Database issue):D655–9. Epub 2012/11/28. <https://doi.org/10.1093/nar/gks1137> PMID: 23180797; PubMed Central PMCID: PMC3531179.
18. Pirtskhalava M, Gabrielian A, Cruz P, Griggs HL, Squires RB, Hurt DE, et al. DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. *Nucleic Acids Res*. 2016; 44(D1):D1104–12. Epub 2015/11/19. <https://doi.org/10.1093/nar/gkv1174> PMID: 26578581; PubMed Central PMCID: PMC4702840.
19. Waghv FH, Barai RS, Gurung P, Idicula-Thomas S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res*. 2016; 44(D1):D1094–7. Epub 2015/10/16. <https://doi.org/10.1093/nar/gkv1051> PMID: 26467475; PubMed Central PMCID: PMC4702787.

20. Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GP. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. *Brief Bioinform.* 2016. Epub 2016/03/27. <https://doi.org/10.1093/bib/bbw025> PMID: 27016393
21. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 2014; 42(Database issue):D1091–7. Epub 2013/11/10. <https://doi.org/10.1093/nar/gkt1068> PMID: 24203711; PubMed Central PMCID: PMC3965102.
22. Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S, et al. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.* 2015; 43(Database issue):D345–56. Epub 2014/11/28. <https://doi.org/10.1093/nar/gku1214> PMID: 25428375; PubMed Central PMCID: PMC3983988.
23. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2015; 12(1):7–8. Epub 2014/12/31. <https://doi.org/10.1038/nmeth.3213> PMID: 25549265; PubMed Central PMCID: PMC398668.
24. Kaur H, Garg A, Raghava GP. PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept Lett.* 2007; 14(7):626–31. Epub 2007/09/28. PMID: 17897087.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215(3):403–10. Epub 1990/10/05. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712.
26. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 1981; 147(1):195–7. Epub 1981/03/25. PMID: 7265238.
27. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 2012; 40(Database issue):D1100–7. Epub 2011/09/29. <https://doi.org/10.1093/nar/gkr777> PMID: 21948594; PubMed Central PMCID: PMC3245175.