Briefings in Bioinformatics, 00(00), 2020, 1–7

doi: 10.1093/bib/bbaa064
Problem Solving Protocol

MPTherm: database for membrane protein thermodynamics for understanding folding and stability

A. Kulandaisamy, R. Sakthivel and M. Michael Gromiha

Corresponding author: M. Michael Gromiha, Indian Institute of Technology Madras, Chennai 600036, Tamil Nadu, India.
Tel.: +914422574138; Fax: +91 44 2257 4102; E-mail: gromiha@iitm.ac.in

Abstract

The functions of membrane proteins (MPs) are attributed to their structure and stability. Factors influencing the stability of MPs differ from globular proteins due to the presence of membrane spanning regions. Thermodynamic data of MPs aid to understand the relationship among their structure, stability and function. Although a wealth of experimental data on thermodynamics of MPs are reported in the literature, there is no database available explicitly for MPs. In this work, we have developed a database for MP thermodynamics, MPTherm, which contains more than 7000 thermodynamic data from about 320 MPs. Each entry contains protein sequence and structural information, membrane topology, experimental conditions, thermodynamic parameters such as melting temperature, free energy, enthalpy etc. and literature information. MPTherm assists users to retrieve the data by using different search and display options. We have also provided the sequence and structure visualization as well as cross-links to UniProt and PDB databases. MPTherm database is freely available at http://www.iitm.ac.in/bioinfo/mptherm/. It is implemented in HTML, PHP, MySQL and JavaScript, and supports the latest versions of major browsers, such as Firefox, Chrome and Opera. MPTherm would serve as an effective resource for understanding the stability of MPs, development of prediction tools and identifying drug targets for diseases associated with MPs.

Key words: thermodynamics; stability; transmembrane; database; mutation

Introduction

Protein stability is the free energy difference between its folded and unfolded states. The stability of a protein is predominantly affected by mutations, ligand binding, biological and environmental changes such as pH, temperature, ions, buffers and additives [1]. Experimentally, the protein stability is determined by thermal and chemical denaturation methods. In addition, site-directed mutagenesis is an effective and widely used strategy to elucidate protein stability upon mutations [2, 3]. These experimentally determined thermodynamic data are widely scattered in the literature, and are rapidly expanding. Gromiha et al. [4] developed ProTherm, thermodynamic database for proteins and

A. Kulandaisamy received MTech degree in bioinformatics from Bharathidasan University, India. He is currently doing PhD at Indian Institute of Technology (IIT) Madras and his research is focused on developing computational resources and analyzing the effects mutations in membrane proteins.
R. Sakthivel received PhD in Medical Biochemistry from University of Madras, India. He gained postdoctoral trainings at the University of Pittsburgh and University of Alabama, USA. Currently, he is working at IIT Madras, India. His main research is the molecular mechanisms of Epidermal Growth Factor Receptor (EGFR) signaling pathways, designing of novel small molecule inhibitors for Bcl-2, mutations in cancer genes and protein aggregation.
M. Michael Gromiha received his PhD in Physics from Bharathidasan University, India and served as STA fellow, RIKEN Researcher, Research Scientist and Senior Scientist at Computational Biology Research Center, AIST, Japan till 2010. Currently, he is working as a Professor at Indian Institute of Technology (IIT) Madras, India. His main research interests are structural analysis, prediction, folding and stability of globular and membrane proteins, protein interactions and development of bioinformatics databases and tools.
Submitted: 10 February 2020; **Received (in revised form):** 19 March 2020

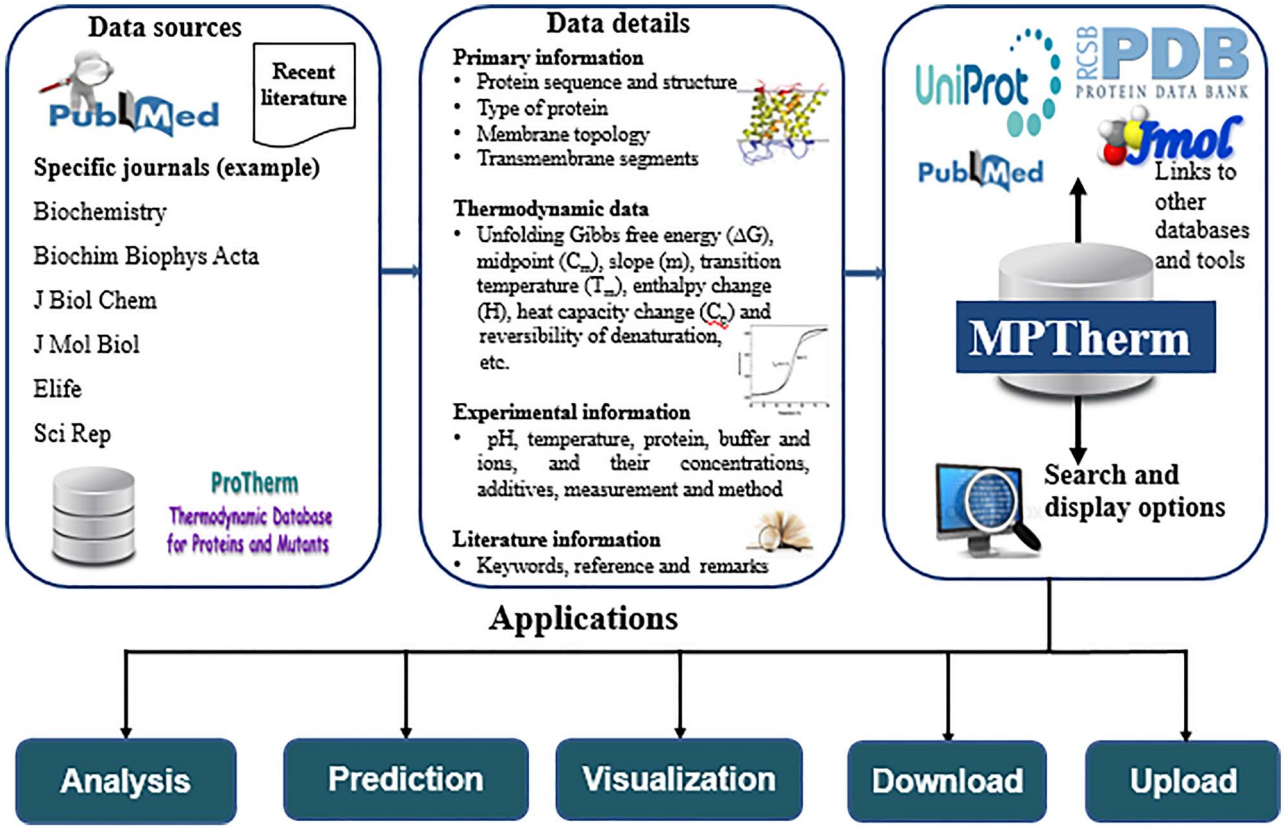

Figure 1. The pipeline diagram for overall workflow of MPTherm database.

mutants, which has more than 25 000 thermodynamic data of proteins and their mutants, and it serves as a potential resource for proteins stability studies [4, 5]. However, it has only limited data on membrane proteins (MPs).

On the other hand, the folding process of MPs differs from water-soluble ones. The folding pathway of these proteins follows a two-step procedure such as insertion into the membrane and the folding of the protein in the membrane lipid bilayer [6–8]. Understanding the folding mechanism of MPs is imperative since they perform versatile cellular and molecular function as well as they play as primary drug targets for different diseases such as cancers, cystic fibrosis, etc. [9,10]. Further, the thermodynamic data on MPs lead to understand the mechanism of disease-causing mutations in different diseases [11].

In an attempt to elucidate the structural and functional relationship of MPs, several databases have been developed using different aspects [12]. For example, PDBTM [13] and MPStruc (http://blanco.biomol.uci.edu/mpstruc/) contain the structural information of MPs. The functionally important residues and the topology information of MPs are accumulated in TMFunction [14] and TOPDB [15], respectively. In addition, a few databases are established to focus on functional aspects: TCDB [16] contains information on transporters, GPCRdb [17] has the complete report of GPCR family or signaling proteins and so on. Recently, we have developed a database MutHTP [18], which has the information on disease-causing and neutral mutations of MPs. However, there is no database available in the literature for thermodynamic data of MPs and their mutants. In this work, we developed MPTherm, a thermodynamic database for MPs with sequence and structure information, membrane topology, experimental conditions, thermodynamic or stability data, literature information and cross-linked with other databases.

Materials and methods

Data collection and curation

Thermodynamic data of MPs are retrieved mainly from the literature (90%), supplemented with available data in ProTherm database (10%). We used the key words, 'MP thermal denaturation studies, chemical denaturation studies on MPs, MP stability upon mutations, unfolding studies of MPs' etc., to retrieve the papers related with thermodynamics of MPs. In addition, articles published in specific journals are manually checked to collect the missing data. The information on protein sequence and structure, experimental conditions, denaturation method, measurement, thermodynamic data, literature and location of the data are directly taken from the research article. Further, the sequence, structure and function of a typical MP are obtained from the literature as well as from UniProt [19] database. The thermodynamic data for MPs available in ProTherm at https://www.iitm.ac.in/bioinfo/ProTherm/ are retrieved by mapping UniProt and PDB [20] codes. The complete work flow and features of MPTherm database are shown in Figure 1.

Contents of the database

Each entry in this database has a unique identification or reference number as well as separate web page, which contains the following information (Table 1).

Table 1. Description of data items in MPTherm with an example entry showing the experimental thermodynamic data of human Chloride intracellular channel protein along with other information

Description	Example
Entry ID	6
Protein	
Gene name	CLIC1
Protein name	Chloride intracellular channel protein 1
Organism	*Homo sapiens*
UniProt ID	O00299
Sequence length	241
PDB ID	1K0M
Mutation (UniProt)	E85L
Mutation type	Single
Salient features of membrane	
Protein type	α-helical protein
Topology	Extra-cellular or Outside
Number of TM segments	1
Experimental conditions	
pH	5.5
Buffer name	Sodium phosphate
Buffer concentration (mM)	50
Ion	NaN3
Ion concentration (mM)	0.02 percentage
Additives	1 mM Dithiothreitol (DTT)
Protein concentration (μM)	2
Measure	CD
Method	Thermal
Experimental thermodynamic data	
T_m (°C)	44.1
ΔT_m(°C)	−7.3
Reversibility	No
Literature	
PubMed ID	25209805
Authors	Cross, M., Fernandes, M., Dirr, H. and Fanucchi, S.
Title	Glutamate 85 and glutamate 228 contribute to the pH-response of the soluble form of chloride intracellular channel 1
Year, volume, pages and Journal	2015, 398(1–2), 83–93 and Mol. Cell. Biochem.
Data location	Table 2, p. 88
Key words	CLIC; pH-sensor; ion channel; stability; metamorphic protein; structural transition

Underline indicates hyperlinks to other databases.

Protein sequence and structure

The sequence-based information contains gene, protein, organism, UniProt identification number, protein sequence and their length (retrieved from UniProt). In addition, mutation information such as wild-type residue, mutant position in UniProt and mutant residue along with the type of mutation (single, double and multiple) are included. For the structure, we provide the protein data bank (PDB) code, chain name, number of unique chains and structural information on mutant residues. UniProt to PDB residue level mapping was carried out with the help of SIFTS database [21].

Salient features of MPs

We have collected the features with respect to the membrane spanning segments, which consist of the type of the MP (α-helical or β-barrel), number of secondary structural elements across the membrane in a protein and topology. We utilized PDBTM database for known structures and CCTOP [22], TOPCONS [23] and PRED-TMBB2 [24] servers for MPs of unknown structures.

Experimental conditions

Experimental conditions include temperature, pH, name of the buffer, ions and their concentrations, additives, protein concentration, measurement, denaturation method and remarks. The remarks information describes specific comments about the protein and experimental conditions.

Thermodynamic data

The thermodynamic data on MPs are obtained by chemical and/or thermal denaturation methods. Chemical denaturation data include the parameters, unfolding Gibbs free energy change (ΔG^{H2O}) obtained in the absence of denaturant, midpoint of denaturant (C_m), slope of the denaturation curve (m) and reversibility of denaturation. To determine the effect of mutations in protein stability, the change in unfolding Gibbs free energy ($\Delta\Delta G^{H2O}$) is computed using the equation:

$$\Delta\Delta G^{H_2O} = \Delta G^{H_2O}(\text{mutant}) - \Delta G^{H_2O}(\text{wild type}).$$

Further, the unfolding Gibbs free energy change (ΔG), unfolding Gibbs free energy change for the mutants ($\Delta\Delta G$), melting

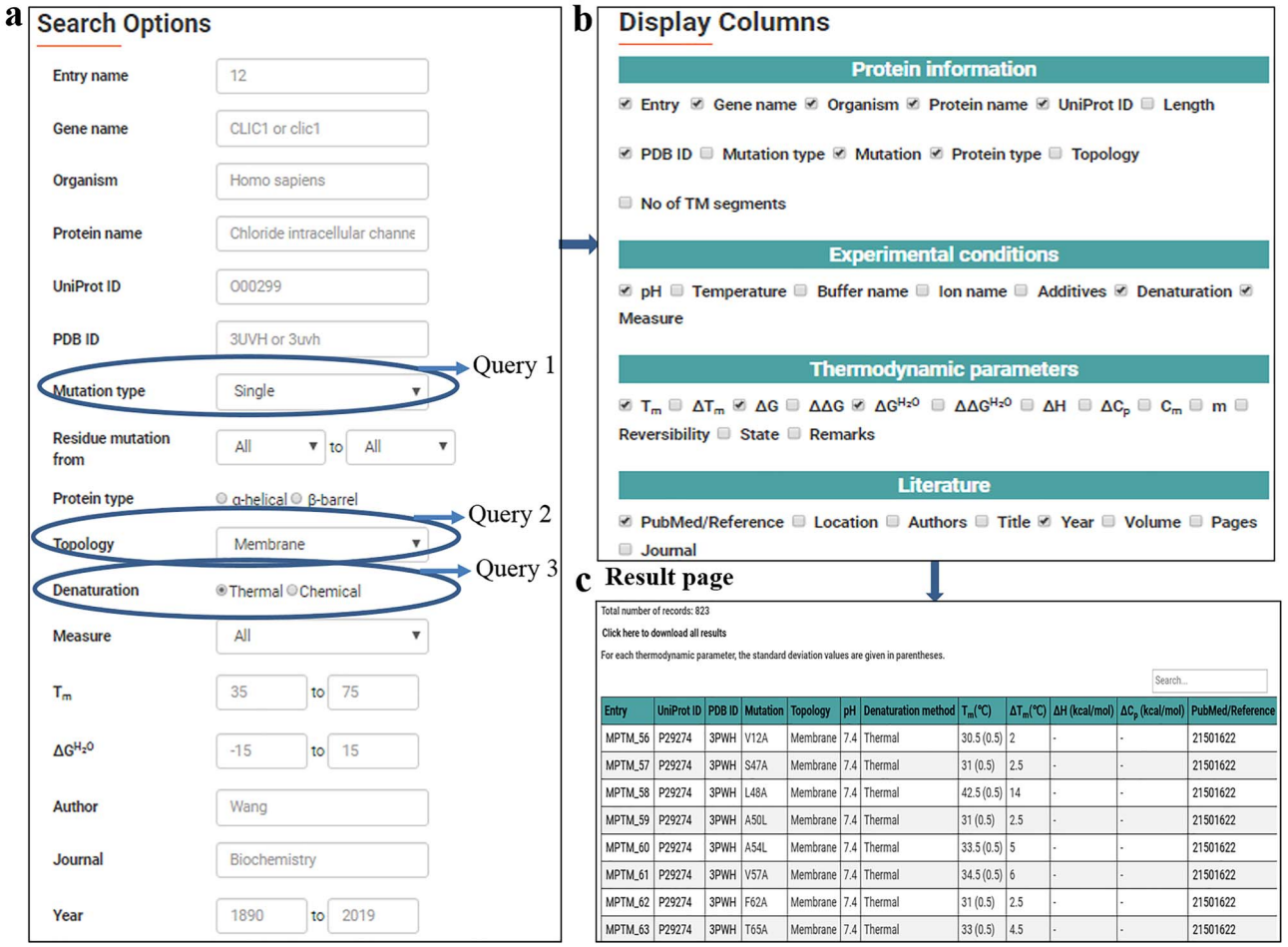

Figure 2. An example of data retrieval from MPTherm database using different search and display options. a) Building a query with multiple search options. b) Different display options provided in the MPTherm database. c) A part of the results obtained from MPTherm.

temperature (T_m), change in melting temperature upon mutation (ΔT_m), enthalpy change (ΔH), heat capacity change (ΔC_p) and reversibility of denaturation are included for the thermal denaturation. The positive and negative signs of the numerical values of the parameters $\Delta\Delta G$, $\Delta\Delta G^{H2O}$ and ΔT_m indicate the stabilizing and destabilizing mutations, respectively. The unit and notation of all thermodynamic parameters are mentioned at https://www.iitm.ac.in/bioinfo/mptherm/faqs.php.

Literature information

It contains PubMed id, names of the authors, title of the paper, journal name, volume and page numbers, keywords and location of the data.

Sequence and structure visualization

The full length of the protein sequence is retrieved from UniProt database and the wild-type residue of a mutation is highlighted in red. Similarly, we obtained the protein structure from PDB and provide options to visualize the wild-type or mutant residue through Jmol viewer [25].

Link with other databases

Each entry of MPTherm database is cross-linked with UniProt and PDB databases. Also, references for all the data are linked with PubMed literature database through their respective identifiers.

Results and discussion

Search and display functionality

The thermodynamic database for MPs and their mutants provides a user-friendly web interface with the following search and display options.

(i) Retrieve data based on MPTherm database identification number.
(ii) Users can obtain information on their protein of interest by entering the name of the gene, protein and/or organism, UniProt and/or PDB code.
(iii) To retrieve the information on specific mutations, users have the option to select the type of the mutation such as single, double, multiple or wild-type as well as any specific mutation (e.g. Ala to Val).
(iv) Specifying the MP type (α or β) and membrane topology such as membrane, cytoplasm and extra-cellular as well as the number of membrane spanning segments.
(v) Extracting the data based on denaturation methods (thermal, denaturant), measurements such as CD, DSC, activity, etc.

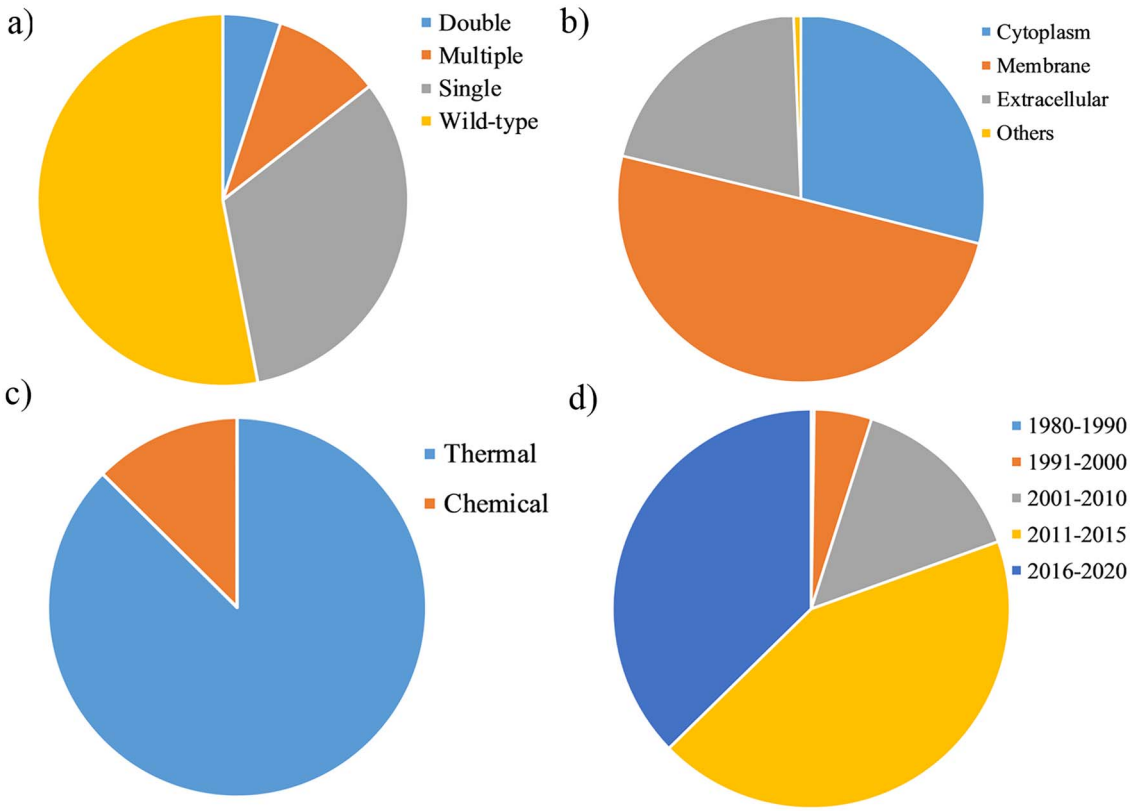

Figure 3. Pie charts show the statistics of MPTherm database in different aspects. a) distribution of wild-type and mutant data, b) distribution of single point mutations with respect to membrane topology, c) distribution of data based on different denaturation methods and d) distribution of data based on year published.

(vi) Limiting data for certain range of particular thermodynamic parameters such as ΔG, $\Delta\Delta G$, ΔG^{H2O}, $\Delta\Delta G^{H2O}$, T_m and ΔT_m.
(vii) Obtaining the data using author names, journal name and year of publication.
(viii) Users can also select the terms to display in the output. It is divided into four sections, such as protein information, experimental conditions, thermodynamic parameters and literature. In addition, users have the flexibility to specify the number of entries to be displayed per page.
(ix) Options to download the data
(x) Each entry is linked with relevant databases (e.g. structure with PDB and sequence with UniProt) for complete information.

We have provided the detailed information about the search and display options in the tutorial page of the database. Sequence and structural visualization are available in the external page of each entry (Table 1).

Example of data retrieval from the database

MPTherm allows users to extract the data from the 'search' web page. An example to extract the thermodynamic data from MPTherm is illustrated in Figure 2a. Users can search with single or combination of multiple conditions. In this example, we showed the combination of multiple conditions such as mutation type as 'single', topology as 'membrane' and the denaturation method as 'thermal'. In the second step, users have the options to select the terms to be displayed in the output. The selected display options in the example are shown in Figure 2b. Once users click the 'submit' button the results are displayed for the query search. In the 'search_result' page, we have also provided the total number of records, results in a table format and an option to download the results (Figure 2c). Users can also search or filter the results using their own keywords.

Database statistics

The current version of MPTherm database contains 7189 thermodynamic data, which includes both wild-type and mutants from 323 proteins. This database contains 53%, 33%, 5% and 9% of wild-type, single, double and multiple mutations, respectively (Figure 3a). Also, a few deletion mutations are available. Further, we have constructed the amino acid substitution matrices for the whole dataset as well as a set of nonredundant point mutations (sequence identity < 30%) and observed that Trp to Phe residue changes are predominant (Supplementary Tables S1 and S2). The substitution matrices are also available at https://www.iitm.ac.in/bioinfo/mptherm/stats.php.

Based on membrane topology, about 50% of single point mutations are located in the transmembrane (TM) region and 29% in the cytoplasmic region (Figure 3b). In addition, the analysis of the distribution of the wild-type and single point mutations revealed that single and seven TM segments containing proteins have more number of data compared to other TM segment proteins (Figure 4). In MPTherm, almost 90% of thermodynamic data belong to α-helical proteins.

On the other hand, 86% of data derived from thermal denaturation methods and most of the available data were measured through fluorescence experiments (Figure 3c). The collected data in MPTherm database are obtained from 457 research

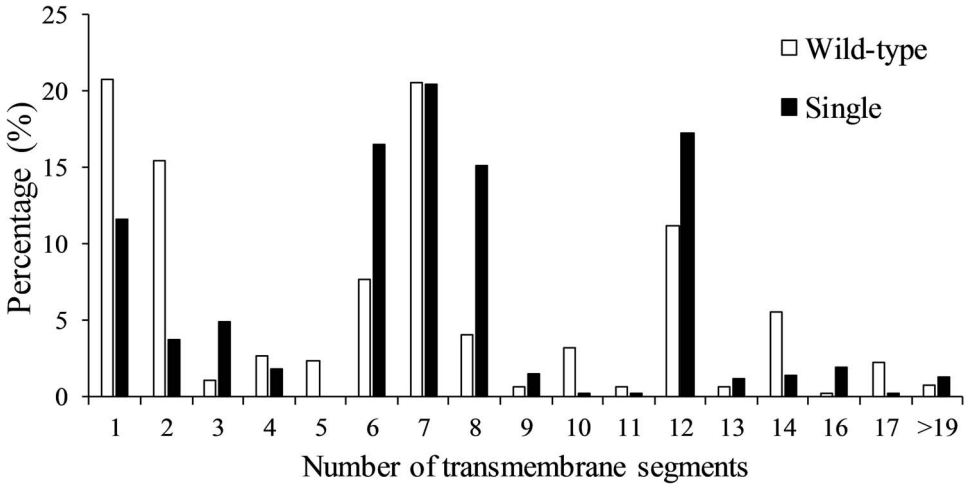

Figure 4. Distribution of wild-type and single point mutation data in the TM segments of the proteins.

papers, and most of them are published in the year between 2000 and 2020 (Figure 3d).

Data upload and download

Users can freely upload their new experimentally determined thermodynamic data of the MPs into MPTherm database. To upload the data, users have to provide the following information in a web form: protein name, UniProt accession number, PubMed or Digital Object Identifier (DOI) number, name of the experiment, experimental conditions and thermodynamic parameters. We have provided an example of the data format in the tutorial section. On the other hand, users can also download the entire MPTherm database by submitting their request to the developed in a web form.

Applications

MPTherm has several potential applications and some of them are: (i) to understand the factors that are governing the folding and the stability of MPs, (ii) to elucidate the relationship among structure, stability and function of MPs, (iii) to understand the mechanism of disease-causing mutations in terms of protein stability, (iv) to characterize the effect of mutations in protein stability and to design the stable mutants in the protein engineering, which has several industrial applications and (v) to develop computational algorithms to identify the stabilizing mutations in MPs.

Availability and update

The MPTherm database is developed using HTML, PHP, MySQL and JavaScript programming languages and it supports the latest version of major browsers such as Firefox, Chrome and Opera. The web interface is available at http://www.iitm.ac.in/bioinfo/mptherm/. The database will be maintained and updated regularly. The updated information will be reflected on the home page of the database. Any constructive comments and suggestions are welcome and should be sent to gromiha@iitm.ac.in

Conclusions

In this work, we have collected the experimental thermodynamic data of MPs and their mutants from the literature and ProTherm database. Based on the collected data, we have developed a specific user-friendly web interface for MPs with a wide number of search and display options. The current version of this database contains ~7000 thermodynamic or stability data from more than 320 proteins. For each entry, we provide protein sequence and structure (if available) information, mutation and their type, salient features with respect to the membrane spanning segments, experimental conditions, denaturation method, measurement, thermodynamic parameters and literature information. In addition, users can also visualize the mutation at the primary and tertiary levels. We suggest that MPTherm serves as a useful resource for understanding the factors, which governs MP stability and to design stable mutants.

Key Points

- We have developed MPTherm database, specific for membrane protein stability. The first version of this database contains ~7000 thermodynamic data of membrane proteins and these data are obtained from more than 450 research papers.
- The information about protein sequence and structure, mutation, specific membrane protein features, experimental conditions and methods, thermodynamic parameters and literature details are provided for each entry.
- MPTherm includes different search and display options, visualization of mutation in the sequence and structure levels, cross-linked with UniProt and PDB databases.
- This database helps in understanding the relationship among thermodynamics, structure and function of membrane proteins.

Supplementary data

Supplementary data are available online at https://academic.oup.com/bib.

Acknowledgements

We acknowledge Prof. Dmitrij Frishman, Technical University of Munich for his constructive comments. We thank the Department of Biotechnology and Indian Institute of Technology Madras for computational facilities.

Funding

We acknowledge the support of the Department of Science and Technology, Government of India (INT/RUS/RSF/P-09) to M.M.G. A.K. was supported by the Ministry of Human Resource Development and Initiative for Biological Systems Engineering Travel grant, Indian Institute of Technology Madras, India.

Conflict of interest

The authors declae that there is no conflict of interest.

References

1. Deller MC, Kong L, Rupp B. Protein stability: a crystallographer's perspective. *Acta Crystallogr F Struct Biol Commun* 2016;**72**:72–95.
2. Bryan PN. Site-directed mutagenesis to study protein folding and stability. *Methods Mol Biol* 1995;**40**:271–89.
3. Harris NJ, Booth PJ. Folding and stability of membrane transport proteins in vitro. *Biochim Biophys Acta* 2012;**1818**:1055–66.
4. Gromiha MM, An J, Kono H, *et al*. ProTherm: thermodynamic database for proteins and mutants. *Nucleic Acids Res* 1999;**27**:286–8.
5. Gromiha MM, Anoosha P, Huang LT. Applications of protein thermodynamic database for understanding protein mutant stability and designing stable mutants. *Methods Mol Biol* 2016;**1415**:71–89.
6. White SH, Wimley WC. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* 1999;**28**:319–65.
7. Fleming KG. Energetics of membrane protein folding. *Annu Rev Biophys* 2014;**43**:233–55.
8. Cymer F, Von Heijne G, White SH. Mechanisms of integral membrane protein insertion and folding. *J Mol Biol* 2015;**427**:999–1022.
9. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there? *Nat Rev Drug Discov* 2006;**5**:993–6.
10. Marinko JT, Huang H, Penn WD, *et al*. Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis. *Chem Rev* 2019;**119**:5537–606.
11. Stefl S, Nishi H, Petukh M, *et al*. Molecular mechanisms of disease-causing missense mutations. *J Mol Biol* 2013;**425**:3919–36.
12. Gromiha MM, Ou YY. Bioinformatics approaches for functional annotation of membrane proteins. *Brief Bioinform* 2014;**15**:155–68.
13. Kozma D, Simon I, Tusnady GE. PDBTM: protein data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res* 2013;**41**:D524–9.
14. Gromiha MM, Yabuki Y, Suresh MX, *et al*. TMFunction: database for functional residues in membrane proteins. *Nucleic Acids Res* 2009;**37**:D201–4.
15. Tusnady GE, Kalmar L, Simon I. TOPDB: topology data bank of transmembrane proteins. *Nucleic Acids Res* 2008;**36**:D234–9.
16. Saier MH, Jr, Yen MR, Noto K, *et al*. The transporter classification database: recent advances. *Nucleic Acids Res* 2009;**37**:D274–8.
17. Isberg V, Mordalski S, Munk C, *et al*. GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* 2017;**45**:2936.
18. Kulandaisamy A, Binny Priya S, Sakthivel R, *et al*. MutHTP: mutations in human transmembrane proteins. *Bioinformatics* 2018;**34**:2325–6.
19. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017;**45**:D158–69.
20. Burley SK, Berman HM, Kleywegt GJ, *et al*. Protein data Bank (PDB): the single global macromolecular structure archive. *Methods Mol Biol* 2017;**1607**:627–41.
21. Dana JM, Gutmanas A, Tyagi N, *et al*. SIFTS: updated structure integration with function, taxonomy and sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Res* 2019;**47**:D482–9.
22. Dobson L, Reményi I, Tusnády GE. CCTOP: a consensus constrained TOPology prediction web server. *Nucleic Acids Res* 2015;**43**:W408–12.
23. Tsirigos KD, Peters C, Shu N, *et al*. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res* 2015;**43**:W401–7.
24. Tsirigos KD, Elofsson A, Bagos PG. PRED-TMBB2: improved topology prediction and detection of beta-barrel outer membrane proteins. *Bioinformatics* 2016;**32**:i665–71.
25. Scalfani VF, Williams AJ, Tkachenko V, *et al*. Programmatic conversion of crystal structures into 3D printable files using Jmol. *J Chem* 2016;**8**:66.