TMFunction: database for functional residues in membrane proteins

M. Michael Gromiha^{1,*}, Yukimitsu Yabuki¹, M. Xavier Suresh¹, A. Mary Thangakani², Makiko Suwa¹ and Kazuhiko Fukui¹

¹Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), AIST Tokyo Waterfront Bio-IT Research Building, 2-42 Aomi, Koto-ku, Tokyo 135-0064 and ²Advanced Technology Inc., Tokyo, Japan

Received August 13, 2008; Revised and Accepted September 22, 2008

ABSTRACT

We have developed the database TMFunction, which is a collection of more than 2900 experimentally observed functional residues in membrane proteins. Each entry includes the numerical values for the parameters IC50 (measure of the effectiveness of a compound in inhibiting biological function), $V_{\rm max}$ (maximal velocity of transport), relative activity of mutants with respect to wild-type protein, binding affinity, dissociation constant, etc., which are important for understanding the sequencestructure-function relationship of membrane proteins. In addition, we have provided information about name and source of the protein, Uniprot and Protein Data Bank codes, mutational and literature information. Furthermore, TMFunction is linked to related databases and other resources. We have set up a web interface with different search and display options so that users have the ability to get the data in several ways. TMFunction is freely available at http://tmbeta-genome.cbrc.ip/TMFunction/.

INTRODUCTION

Membrane proteins perform a diverse variety of functions and are used as main drug targets of pharmaceutical agents. The collection of information on potential amino acid residues for the function of membrane proteins is important for understanding the sequence–structure–function relationship of membrane proteins as well as predicting the functional residues from sequence/structure. Tusnady *et al.* (1) constructed a database for transmembrane proteins, which covers the three-dimensional structures of membrane proteins deposited in Protein Data Bank (PDB) and information on membrane spanning α -helices and β -strands obtained with the TMDET algorithm (2). Saier et al. (3) developed a comprehensive classification system for membrane transport proteins known as the Transport Classification Database (TCDB). Further, functional databases have been developed for G-protein-coupled receptors (GPGRs), human seven transmembrane receptors, Arabidopsis integral membrane proteins and so on (4,5). Edvardsen et al. (6) created a G-protein-coupled receptor mutant database, which is mainly focused on different families of GPCRs. Mutational databases have also been developed for the structure, function and thermodynamics of proteins (7,8). On the other hand, several methods have been proposed for discriminating transmembrane α -helical and β -strand proteins, predicting their membrane-spanning segments, and functional classification of membrane proteins (9–16). In spite of these studies, there is no database available for the broad collection of potential residues, which are important for the functions of different classes of membrane proteins including receptors, transporters, channel proteins, etc. This information can be obtained from experimental studies on membrane proteins that have reported the measured values of several parameters for membrane protein function. The collection of such data is important and necessary for analyzing and predicting potentially important residues for practical applications.

In this work, we have developed a database, TMFunction, which is a collection of more than 2900 experimental data about important amino acid residues in membrane proteins, reported in the literature. It has information about functionally important residues, numerical values for the parameters IC50, $V_{\rm max}$, activity, affinity, etc., along with sequence and structure information for the protein, mutational and literature information. This database will help in understanding the relationship between amino acid sequences/structures and functions of membrane proteins. We have developed a WWW interface to facilitate searching the database and displaying the results with different options.

© 2008 The Author(s)

^{*}To whom correspondence should be addressed. Tel: +81 3 3599 8046; Fax: +81 3 3599 8081; Email: michael-gromiha@aist.go.jp

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/ by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

	瀢		Func	ctional	me t Databa: Thursd	co II se of Mi lay, July 31,	AFUN embrane 2008	e Protei	n	J		X	
					5	SEARCH							
	Home	PI	ease fill or cl	hoose ne	cessary en	tries belo	v and set d	isplay opt	ions				
	Search	Proteir	All CB2 Cannab Carnitin e/org	inoid Recep panic cation	ptor transporter; (OCTN2			I	Uni Prot I	D AII ACI AG	M3_RAT (P084 TRA_RAT (P29	183) 1095)
	Statistics	Source	All Aspergillus r	o acid trans	porter channe		T TMHol		l		ICF.	IR_HUMAN (P	13569)
	Links	300100	Bacteriophag Bos taurus (I	ge M13 Bovine) on	-	-		All	FLICIN	5			
	Help	Funct	Clectrogen Export	icity			Parameter	Accessit Activity Affinity	bility	1			
Ĩ	Contact	Mutatio	an 10		Single	- Multiple	□ wild ty	rpe					
		Author Year	rom [10		1		Searc	5	Clear			
1	Display Option	Author Year		10			UniProt	Searc	PD	Gear		Туре	
1	Display Option	Author Year (? PP	ROTEIN		OURCE		UniProt Data	ID C	PDI Fui	B code		Type Experime	int
	Display Option	Author Year 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	ROTEIN scation uthor		OURCE aramete MID	er V	UniProt Data Journal	ID 🔽	PDI Fui	B code		Type Experime	int
SFAP	Display Option	Author Year 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	ROTEIN bocation uthor allts				UniProt Data Journal	ID F	PDI Fui	B code Inction		Type Experime	int
S F A P P S U	Display Option	Author Year 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	ROTEIN acation uthor		OURCE aramete MID		UniProt Data Journal	ID F	PDI Fui	B code Inction		Type Experime	nt
S F F F F F F F F F F F F F F F F F F F	Display Option	Author Year ; 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	ROTEIN bocation uthor allts		OURCE aramete MID	er V	UniProt Data Journal	ID F	PDI Fui	B code Inction		Type Experime	nt
S F F P P S U U	Display Option	Author Year 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	ROTEIN pocation uthor 11ts		OURCE aramete MID		UniProt Data Journal	ID I	PD) Fui	B code Inction	Intioe	Type Experime	PubMas
S F F P F S U U	Display Option	Author Year : 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	COTEIN COTEIN Cotion Athor Athor Athor MRP1 HUM MRP1 HUM MRP1 2335270	TO T	Mutation 51233A	Relative res	UniProt Data Journal Journal	ID Searc	PD) Fui	B code Inction	anction	Type Experime	PubMe ID 192544
S FF N P FF S UU	Display Option	Author Year ; Year ; LC C Au Cest Is	COTEIN Cation Juthor Ilts	TO P	Mutation \$1233A \$1235A	Pr V Pr V	UniProt Data Journal Journal stance factor	ID Searc ID V V V V V V V V V V V V V V V V V V V	PD Fui	B code Inction	metion	Type Experime	PubMin 192544

Figure 1. An example of searching conditions, display options and results of TMFunction: (a) main menu for the search options in TMFunction. The items function (drug) and single mutants are selected for search as indicated by arrows; (b) display options in TMFunction. We have selected entry, protein, Uniprot ID, mutation, parameter, data, function and PMID to show in the output; (c) part of the results obtained from TMFunction.

Relative re

Relative re

S1233A Relative resistance factor to VP-16 13.3

13.7

17.3

4.4

4.2

29.3

ce factor to

Drug resistance

Drug resistance

Drug resistance

Drug resistance

Drug te

Drug resistance(partial)

etpartial

11925441

11925441

11925441

S1237A

Q1239A

T1241A

¥1243F

N1245A

MRP1 HUMAN (P33527)

MRP1 HUMAN (P33527)

MRP1 HUMAN (P33527)

in 1:

190 MRP1 191 Multide MRP1

192 MRP1 193 Multide MRP1

198 Multidrug resistance protein 1; MRP1

192

194

CONTENTS OF THE DATABASE

Each entry in the database includes the following information:

Sequence and structure information: name and source of the protein, Uniprot (17) and PDB (18) codes, type of the integral membrane protein (α -helical or β -strand), mutational details (single, double or wild-type; mutation has been identified with mutated residue, residue number and mutant residue; e.g. A102V) and location of mutants.

Functional information: factors that are mainly affected by the mutation of amino acid residues in membrane proteins, such as relative activity of mutants, affinity for binding, channel, drug, glycosylation, membrane insertion, cellular signaling, membrane translocation, transport, etc.

Functional data: numerical values for affinity (%), B_{max} , IC50, drug sensitivity, K_{d} , K_{m} , V_{max} , uptake (%), etc. *Experimental methods and conditions*: measurement,

Experimental methods and conditions: measurement, method and the ligand used for the control data.

Literature information: keywords, reference, authors and remarks.

DATABASE STATISTICS

The first release of TMFunction contains 2907 entries from 83 different proteins, which perform 29 diverse functions. It has data for 2092 single mutants, 580 multiple mutants and 273 wild-types. The majority of the data concerns transmembrane helical proteins (2760) followed by β -barrel membrane proteins (147). The data are obtained from more than 100 scientific articles published in 20 different journals.

FEATURES OF TMFUNCTION

TMFunction includes several features in the search and display options as shown in Figure 1, and as briefly explained subsequently:

- (i) Retrieving data for a particular protein and/or source.
- (ii) Specifying the type of the mutant as single, multiple and/or wild-type.
- (iii) Selecting the function of the protein (transport) as well as the numerical value of the functional parameter (e.g. IC50).
- (iv) Mentioning the type of the protein.
- (v) Extracting data by authors, publication year and keywords.
- (vi) Downloading entire data.

Detailed tutorials describing the usage of TMFunction are available at the home page. For example, the data obtained for the function 'drug' and single mutants is shown in Figure 1a. The terms: entry, protein, Uniprot ID, mutation, parameter, data, function, experiments and Pubmed ID have been selected for displaying the results (Figure 1b). Figure 1c shows the final results obtained with the search conditions and display options.

LINKS TO OTHER DATABASES

Each entry in TMFunction is linked to Uniprot ID (http:// www.uniprot.org/) and PDB code (http://www.rcsb.org) to obtain the sequence and structure information directly. The references for all data are directly connected to the PUBMED literature database of NCBI (http:// www.ncbi.nlm.nih.gov/pubmed/). Further, we have provided links to several related databases and web servers including sequences and structures, functions and genomes, transmembrane helix and strand predictions (http://tmbeta-genome.cbrc.jp/TMFunction/ DBlinkspage.html).

AVAILABILITY AND CITATION OF TMFUNCTION

The database can be freely accessible at http://tmbeta-gen ome.cbrc.jp/TMFunction/. If this database is used as a tool in your published research work, please cite this article including the URL. Suggestions and comments are welcome and should be sent to michaelgromiha@aist.go.jp.

ACKNOWLEDGEMENT

We thank Dr Martin Frith for critical reading of the manuscript.

FUNDING

The support from Computational Biology Research Center (CBRC) and National Institute of Advanced Industrial Science and Technology (AIST) is gratefully acknowledged. Funding for open access charge: Computational Biology Research Center (CBRC).

Conflict of interest statement: None declared.

REFERENCES

- 1. Tusnady,G.E., Dosztanyi,Z. and Simon,I. (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res.*, **33**, D275–D278.
- 2. Tusnady,G.E., Dosztanyi,Z. and Simon,I. (2005) TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. *Bioinformatics*, **21**, 1276–1277.
- 3. Saier, M.H. Jr, Tran, C.V. and Barabote, R.D. (2006) TCDB: the transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res.*, **34**, D181–D186.
- 4. Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F.E. and Vriend, G. (2003) GPCRDB information system for G proteincoupled receptors. *Nucleic Acids Res.*, **31**, 294–297.
- Schwacke, R., Schneider, A., Van Der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W.B., Flugge, U.I. and Kunze, R. (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol.*, **131**, 16–26.
- Edvardsen, O., Reiersen, A.L., Beukers, M.W. and Kristiansen, K. (2002) tGRAP, the G-protein coupled receptors mutant database. *Nucleic Acids Res.*, 30, 361–363.
- Kawabata, T., Ota, M. and Nishikawa, K. (1999) The Protein Mutant Database. Nucleic Acids Res., 27, 355–357.
- Gromiha, M.M., An, J., Kono, H., Oobatake, M., Uedaira, H. and Sarai, A. (1999) ProTherm: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, 27, 286–288.

- Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14, 378–379.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol., 305, 567–580.
- 11. Tusnády,G.E. and Simon,I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics*, **17**, 849–850.
- Lehnert, U., Xia, Y., Royce, T.E., Goh, C.S., Liu, Y., Senes, A., Yu, H., Zhang, Z.L., Engelman, D.M. and Gerstein, M. (2004) Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions. *Q. Rev. Biophys.*, 37, 121–146.
- Gromiha, M.M. and Suwa, M. (2007) Current developments on beta-barrel membrane proteins: sequence and structure analysis, discrimination and prediction. *Curr. Protein Pept. Sci.*, 8, 580–599.

- Gromiha, M.M., Yabuki, Y., Kundu, S., Suharnan, S. and Suwa, M. (2007) TMBETA-GENOME: database for annotated beta-barrel membrane proteins in genomic sequences. *Nucleic Acids Res.*, 35, D314–D316.
- Gromiha, M.M. and Yabuki, Y. (2008) Functional discrimination of membrane proteins using machine learning techniques. *BMC Bioinform.*, 9, 135.
- Li,H., Dai,X. and Zhao,X. (2008) A nearest neighbor approach for automated transporter prediction and categorization from protein sequences. *Bioinformatics*, 24, 1129–1136.
- Wu,C.H., Apweiler,R., Bairoch,A., Natale,D.A., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, 34, D187–D191.
- Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.*, 35, D301–D303.