

TSTMP: TARGET SELECTION FOR HUMAN TRANSMEMBRANE PROTEINS

Julia Varga¹, László Dobson², István Reményi² and Gábor E. Tusnady²

1) Department of Applied Biotechnology and Food Science, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Budapest, Hungary

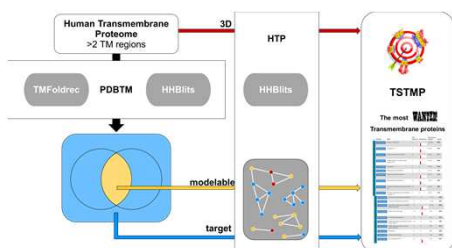
2) „Momentum” Membrane Protein Bioinformatics Research Group, Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, Hungary;

INTRODUCTION

Transmembrane proteins account for 25-30% of the human proteome, however only a fraction of determined protein structures belongs to this class. Currently, there are only **60 known 3D structures** among the polytopic human transmembrane proteins and **about a further 600 could be reliably modeled** using existing structures.

One possible approach to determine targets for protein crystallization is the **selection of those proteins whose structures could be a template of homology modeling** for the highest possible number of proteins. However, such available target lists for transmembrane proteins are not updated regularly. Our aim was to create a **regularly updated database** for the target selection of human transmembrane proteins.

METHODS



An automated pipeline was used for the creation of the database that combines the results of a state-of-the-art transmembrane specific **fold recognition** algorithm² and a **sequence similarity search** algorithm³ enhanced with **topology data**⁴ for searching model structures and **generating protein clusters**.

TargetTrack database was also incorporated into the database to allow the monitoring of the crystallization process of transmembrane proteins.

CATEGORIES IN DATABASE	
3D	PDBTM ⁵ structure for the protein exists
MODELABLE	at least one possible model template for homology modeling exists
TARGET	no 3D structure or model template

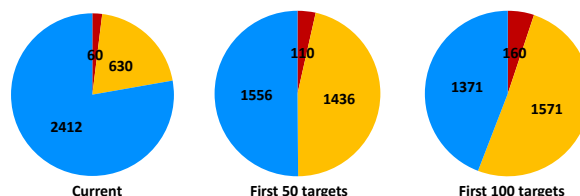
RESULTS

THE MOST WANTED LIST

#	Identifier	Name	HTM segments	#TargetTrack (targets)	Homologues	Cluster
1	Target_00001	Olfactory receptor 8B1	7	1	419 (207)	0005
2	Target_00002	Tetraspanin-4	4	1	21 (23)	0016
3	Target_00003	Sodium channel family 2 member 1	12	2	34 (23)	0117
4	Target_00004	Protein tyrosine phosphatase receptor type 2	2	1	21 (23)	0005
5	Target_00005	Tetraspanin-7	4	2	18 (18)	0016
6	Target_00006	Palmitoyltransferase ZDHHC2	4	2	18 (18)	0005
7	Target_00007	Tetraspanin-12	4	1	24 (14)	0125
8	Target_00008	Monocarboxylate transporter 4	12	1	18 (18)	0039
9	Target_00009	Protein phosphatase 4A family class A member 10	10	1	13 (13)	0035
10	Target_00010	Membrane spanning 4 domain subfamily 4 member 10	4	2	12 (12)	0122

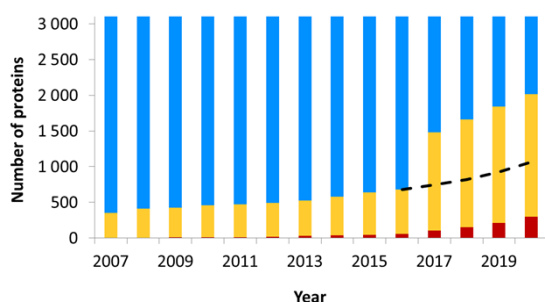
The *Most Wanted list* shows the way the human transmembrane proteome could be modeled with the smallest number of crystallized proteins.

EVIDENCE LEVELS IN THE DATABASE



Although crystallizability is greatly affected by a huge amount of unforeseen or not investigated factors, in an optimistic case the **number of 'modelable' proteins could be doubled** by crystallizing the first 50 targets from the *Most wanted list* considering our ranking.

IMPROVEMENT OF THE STRUCTURAL COVERAGE OF THE PROTEOME



The growth in the number of human 3D protein structures is slow but **exponential**. Extrapolating data of previous years concludes that solving structures of transmembrane proteins selected according to our *Most wanted list* increase the number of modelable proteins more than choosing proteins randomly (black, dashed).

TARGETTRACK STATUSES

Number of homologues	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
selected	99	56	33	25	13	12	11	10	13	10	7	5	16	5	3	1	2	2	2	4	8	3	2	
cloned	15	18	8	3	6	2	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
expressed	25	11	10	5	3	4	7	3	3	1	2	2	4	2										
work stopped	9	6	2	3	2	5	2	3																
other soluble	3	2																						
membrane protein solubilized	4	5																						
purified	12	4	2	5	2	1	2	1	2															
crystallized	2																							
diffraction	2																							

Proteins with high number of target homologues are selected or in process of crystallization **below number** compared to their importance.

CONCLUSIONS

In TSTMP database we examined the **possibility of modeling a transmembrane protein structure** assuming prior knowledge about its relatives, and created a **list of target proteins** from the human transmembrane proteome, as well as a strategy to achieve a higher structural coverage. We provided a possible strategy for reaching higher structural coverage of human transmembrane proteins.



<http://tstmp.enzim.ttk.mta.hu>

(1) Chen, L., Oughtred, R., Berman, H.M. and Westbrook, J. (2004) TargetDB: a target registration database for structural genomics projects. *Bioinformatics*, 20, 2860–2.

(2) Kozma, D. and Tusnady, G.E. (2015) TMFoldRec: a statistical potential-based transmembrane protein fold recognition tool. *BMC Bioinformatics*, 16, 201.

(3) Remmert, M., Biegert, A., Hauser, A. and Söding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM

alignment. *Nat. Methods*, 9, 173–5.

(4) Dobson, L., Reményi, I., Tusnady, G.E., Kozma, D., Simon, J., Tusnady, G., Tusnady, G., Dosztányi, Z., Simon, J., Tusnady, G., et al. (2015) The human transmembrane proteome. *Biol. Direct*, 10, 31.

(5) Kozma, D., Simon, J. and Tusnady, G.E. (2013) PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, 41, D524–D529.

