

Comprehensive Functional Identification of Transmembrane Proteins Using the BTP Method

Yasuhito Inoue

gs02409@si.hirosaki-u.ac.jp

Toshio Shimizu

slsimi@si.hirosaki-u.ac.jp

Department of Electronic and Information System Engineering, Faculty of Science and Technology, Hirosaki University, Hirosaki 036-8152, Japan

Keywords: transmembrane protein, transmembrane topology, loop length, binary topology pattern, functional identification

1 Introduction

It has been revealed that TM protein functions correspond well to the TM topologies [the number of TM segments (tms), the loop length and the N-terminus location] [4, 5] and long loop regions contain functional domains. Recently, our group proposed a novel approach for classifying and identifying the functional classification and identification of TM protein functions based on TM topology information (the binary TM topology pattern (BTP) method) [4], where each loop length is expressed as ‘1’ for a long loop or ‘0’ for a short loop, and BTP is described as a string of binary loop length. By applying the BTP method to TM protein sequences encoded in 12 eukaryotic and 87 prokaryotic genomes, in this study, we have identified and classified comprehensively functions of TM proteins of these genomes.

2 Materials and Methods

When the loop length is longer than a defined threshold length (the detailed procedure of defining the threshold length is described previously [5]), a binary loop length of ‘1’ is assigned to the loop. Otherwise, the binary loop length is ‘0’. A wild-card, ‘*’ is given to the loop of which length cannot be defined explicitly as long or short.

We focused on TM proteins with up to 12-tms except 1- and 9-tms. TM protein sequences extracted from SWISS-PROT (Release 38.0) were categorized into 37 TM protein functions according to the descriptions in the DE, CC and KW lines. For example, 4-tms TM proteins were classified into 3 functional groups: “receptor”, “gap junction” and “others”. The BTPs obtained for the three functional groups are “10010”, “0110*” and “0*0***”, respectively, with the threshold lengths of {47, 30, 28, 80, 42} residues. For 6-tms TM protein functional groups, “channel”, “MIP channel”, “transporter” and “others”, BTPs of “1**11*1”, “0**1000”, “0***10*” and “000****” were obtained with the threshold length of {100, 15, 24, 11, 14, 38, 72}, respectively.

In total 454,202 open reading frames (ORFs) of 12 eukaryotic and 87 prokaryotic (72 bacterial and 15 archean) genomes were downloaded from the GenBank, Ensemble, PlasmoDB and DOE-JGI database. We detected TM proteins from the proteomes by SOSUI [1]. After removing the signal peptide regions detected by DetecSig [3], the sequences were subjected to TM topology prediction by ConPred [2]. The BTPs of 37 functional groups were applied to the finally predicted TM protein sequences for functional identification and classification.

3 Results and Discussion

The functional identification results are summarized in Table 1, where the results are listed only for 4-, 6- and 12-tms TM proteins. Among 7,352 TM protein sequences predicted as 4-tms, for example, 189, 521 and 3,999 sequences were discriminated as “receptor”, “gap junction” and “others”, respectively. The functions of remaining 2,643 sequences couldn’t be identified, i.e., “not classified”. As with 6-tms TM proteins, 148 “channel”, 1,346 “MIP channel”, and 3,427 “transporter” and 701 “others” sequences were identified, while 1,589 sequences were remained not classified. It is interesting to know that the BTP method didn’t detect neither 4-tms “receptor” nor 6-tms “channel” genes in prokaryotic genomes, which seem to be specific to only eukaryotic species. The eukaryotic-specific transporter, 12-tms “ABC transporter” gene was not found in prokaryotic genomes by the BTP method, expectedly in this case too.

Table 1: Functional classification/identification of 4-, 6- and 12-tms TM proteins encoded in 12 eukaryotic, 31 proteobacterial, 22 gram-positive bacterial, 19 other bacterial and 15 archaea genomes by the BTP method. Threshold lengths are {47, 30, 28, 80, 42}, {100, 15, 24, 11, 14, 38, 72} and {41, 12, 16, 13, 7, 14, 292, 29, 20, 12, 19, 14, 274} residues, respectively.

Functional groups	BTPs	Eukaryote	Proteobacteria	Firmicutes	Other bacteria	Archaea	Total
4-tms		2,857	1,821	1,140	926	608	7,352
Receptor	10010	177	4	6	2	0	189
Gap junction	0110*	196	127	73	79	46	521
Others	0*0**	1,161	1,120	758	565	395	3,999
<i>Not classified</i>		<i>1,323</i>	<i>570</i>	<i>303</i>	<i>280</i>	<i>167</i>	<i>2,643</i>
6-tms		2,583	1,859	1,235	882	652	7,211
Channel	1**11*1	144	0	2	2	0	148
MIP channel	0**100	280	454	304	194	114	1,346
Transporter	0***10*	1,288	907	563	405	264	3,427
Others	000****	184	172	150	92	103	701
<i>Not classified</i>		<i>687</i>	<i>326</i>	<i>216</i>	<i>189</i>	<i>171</i>	<i>1,589</i>
12-tms		1,127	1,359	516	382	284	3,668
Sodium transporter	*****0*1**1*	149	48	32	31	24	284
Sugar transporter	***0**000***0	431	621	177	137	105	1,471
ABC transporter	1*101*1*****	16	0	0	0	0	16
Others	*****0*0	302	360	202	123	105	1,092
<i>Not classified</i>		<i>229</i>	<i>330</i>	<i>105</i>	<i>91</i>	<i>50</i>	<i>805</i>

References

- [1] Hirokawa, T., Boon-Chieng, S., and Mitaku, S., SOSUI: classification and secondary structure prediction system for membrane proteins, *Bioinformatics*, 14(4):378–379, 1998.
- [2] Ikeda, M., Arai, M., Lao, D.M., and Shimizu, T., Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies, *In Silico Biol.*, 2(1):19–33, 2002.
- [3] Lao, D.M. and Shimizu, T., Methods for detecting the signal peptide in transmembrane and globular proteins, *Genome Informatics*, 12:340–342, 2001.
- [4] Otaki, J.M. and Firstein, S., Length analyses of mammalian G-protein-coupled receptors, *J. Theor. Biol.*, 211(2):77–100, 2001.
- [5] Sugiyama, Y., Poluliakh, N., and Shimizu, T., Identification of transmembrane protein functions by binary topology patterns, *Protein Eng.*, 16(7):479–488, 2003.