Transmembrane Topology Pattern and Detection of Transmembrane Protein Functions

Natalia Poluliakh¹
gs98616@si.hirosaki-u.ac.jp

Kohei Saito¹
gs00608@si.hirosaki-u.ac.jp

Toshio Shimizu¹,²
slsimi@si.hirosaki-u.ac.jp

¹ Department of Information Science, Graduate School of Science, Hirosaki University, Hirosaki, 036-8561 Japan
² Department of Electronic and Information System Engineering, Faculty of Science and Technology, Hirosaki University, Hirosaki, 036-8561 Japan

Keywords: transmembrane topology, loop length, topology pattern, binary expression, function identification

1 Introduction

Transmembrane (TM) proteins (TMPs) have important roles in the biological organisms [1, 2]. With the rapid increasing of the number of TMPs, a lot of efforts are made in attempt to classify TMP functions by sequence analysis, and in fact, some databases aimed for it have been constructed (e.g., Pfam [3]). Which characteristics of TMP could provide us important clues to its function? We suppose that the TMP topology information that is readily available in the sequence data is a good prospect for this purpose.

In this study, we assume that a TMP function is defined by its TM topology: the number of transmembrane segments (tms), loop length and N-terminus location. In order to detect TMP functions, we especially focus on its loop length [4, 5] and simplify TM topology by classifying the loop length into two categories (i.e., “short” and “long”) and express it as the topology pattern. Then we examine how TMP functions correspond to topology patterns.

2 Data and Methods

To classify loops into two categories based on their length, we temporarily define 30 residues as the threshold length. If a loop is shorter than the threshold length, it is defined as the “short” loop and a binary code of ‘0’ is assigned to it. Otherwise, it is defined as the “long” loop, and ‘1’ is assigned to it. For example, in the case of 2-tms TMP having a long N-terminal loop, and short connecting and C-terminal loops, the topology pattern is ‘100’.

We focused here on the 4-tms TMPs, 6-tms TMPs and 7-tms TMPs extracted from SWISS-PROT release 38 [6], from which TMPs with undefined topology and fragments were excluded. Finally, we obtained 372 entries of 4-tms TMPs, which mostly include acetylcholine (79 entries), gamma-aminobutyric acid (56), glutamate (28) and glycine (10) receptors, gap junctions (51) and antigens (45). We also obtained 119 entries of 6-tms TMPs, which mainly include transporters (42) and channels (59). For 7-tms TMPs, 735 out of the 740 entries obtained are G-protein coupled receptors (GPCRs).

Then, we investigated how the TMP functions correspond to topology patterns. For this purpose the most favorable case is that one function corresponds to only one topology pattern. In another case that one function corresponds to several topology patterns, we calculate the probability of every loop to be defined as “long”, in order to obtain a topology pattern as the probability model.


3 Results and Discussion

In the case of 4-tms TMPs, only the receptors were examined in their correspondence to topology patterns. Acetylcholine, gamma-aminobutyric-acid and glycine receptors correspond to a topology pattern, (1, 0, 0, 1, 0), while glutamate receptors have a (1, 0, 0, 1, 1) topology pattern. Indeed, these 4 receptor types can be distinguished using the respective topology patterns from other TMPs.

In the 6-tms TMP entries we analyzed, there are 3 groups of channel proteins: cyclic-nucleotide channel (11 entries), potassium channel (12) and aquaporin channel (26). Since each of the channel groups is represented in more than two topology patterns, the topology pattern is derived as the probability model by using the method described above. The obtained topology patterns are as follows: (1.00, 0.00, 0.00, 1.00, 0.00, 0.91, 1.00) for cyclic-nucleotide channel; (1.00, 0.58, 0.00, 0.08, 0.00, 0.91, 1.00), potassium channel; (0.31, 0.04, 0.35, 0.04, 0.00, 0.23, 0.65), aquaporin channel. The distinctive feature of the topology pattern of cyclic-nucleotide channels is in its long N-terminal and 3-4 connecting loops. In the case of the topology pattern of the potassium channels the specificity is in its long N-terminal loop and short 3-4 connecting loop. With regard to the topology pattern of aquaporin channels, every loop except for the C-terminal loop has a tendency to be “short”.

From the results described above, it is possible to distinguish 4-tms receptors and 6-tms channels from other TMPs using our topology pattern detection method proposed in this study.

We also analyzed 735 entries of 7-tms GPCR to obtain its topology pattern. As in the case of 6-tms TMP, several topology patterns correspond to one function, and for this reason, the topology pattern was obtained as a probability model. The topology pattern, (0.75, 0.00, 0.01, 0.01, 0.22, 0.45, 0.00, 0.78) clearly demonstrates the specificity of GPCR, i.e., with long C- and N-terminal loops. These two long loops in the topology pattern could make it possible to distinguish GPCRs from the other 7-tms TMPs (e.g., pump), but these characteristics are insufficient for detection of every particular GPCR type.

It is reasonable to conclude that it is possible to distinguish functions from each other according to the topology patterns, but for more precise detection other parameters must be considered. We are planning to add some other characteristics to the topology pattern, for example, by considering the amino acid hydrophobicity of the loop regions. Thus, we expect to be able to improve the sensibility of the topology pattern to identify TMP functions. Also the threshold length should be optimized to increase the ability of topology patterns for identifying TMP functions.

References