Classification of Membrane Proteins by Types of Transmembrane Helices Using SOSUI System

Shigeki Mitaku [1] **Takatsugu Hirokawa** [12] **Mitsuo Ono** [1]

mitaku@cc.tuat.ac.jp hirokawa@cc.tuat.ac.jp mitsuo@cc.tuat.ac.jp

[1] Tokyo University of Agriculture and Technology, Department of Biotechnology, Tokyo 184-5855, Japan

[2] Present address: Ryoka System Inc., Computational Science and Technology Division

1 Introduction

There are many amino acid sequences in proteomes which are not homologous to any other sequences. Therefore, methods to classify proteins independent of the sequence homology are strongly required for computational analysis of proteomes. Proteins may be divided into two categories: soluble and membrane proteins. Since membrane proteins are characterized by the existence of long hydrophobic transmembrane helices, the classification of amino acid sequences into two types of proteins, soluble and membrane proteins, is possible with considerably high accuracy. The classification of all amino acid sequences in several proteomes was reported recently, leading to the conclusion that the fraction of membrane proteins is about 30% [1]. However, those methods provided only the information about the number of transmembrane helices. We have previously proposed a new method (SOSUI) to classify amino acid sequences by two types of transmembrane helices, primary and secondary transmembrane segments. In this work, we have analyzed four kinds of dataset including two set of all amino acid sequences from proteomes, using the SOSUI system [2] which not only predicts transmembrane helix regions but also classifies them by the strength of interaction with lipid membranes.

2 Methods

We have prepared four kinds of dataset of amino acid sequences: (1) ORFs of *Mycoplasma genitalium*, (2) ORFs of *Methanococcus jannaschii*, (3) membrane proteins of known topology, and (4) membrane proteins whose 3D-structure is known. The number of membrane proteins were 105, 359, 80 and 17, respectively.

All amino acid sequences were analyzed by a SOSUI system by which the number of transmembrane helices and the fraction of primary and secondary transmembrane segments were obtained. The performance of the SOSUI system was reported in ref. [2]. The SOSUI system is available at URL: http://www.tuat.ac.jp/~mitaku/.

3 Results and Discussion

The distribution of the number of transmembrane helices showed two peaks. The largest peak was found at the single spanning membrane proteins and a broad peak was observed at the helix number of six or seven. This characteristics of the number distribution was common to three dataset except for membrane proteins with known 3D-structure. Because histograms of the number of transmembrane helices for two biological species is similar to that for the dataset from membrane proteins of known topology, it is expected that membrane proteins with seven transmembrane helices, many of which are receptor proteins are abundant in biological systems.

Statistical analysis showed that two thirds of transmembrane segments were primary transmembrane helices, which will be able to penetrate into membrane by themselves, while the remaining transmembrane helices were secondary ones, which are considered to be stabilized by the interaction with other transmembrane segments. However, fifty percent of transmembrane helices were secondary ones in the dataset of known 3D-structures. Although the statistical significance of the increased fraction of secondary transmembrane helices is not clear yet because of the paucity of data, this type of transmembrane helices may stabilize the tertiary structure of membrane proteins.

Functionally important residues usually have polar groups. Therefore, functionally important helices should have polar residues, which are the characteristics of secondary transmembrane helices. The number of secondary transmembrane helices in most membrane proteins was less than three. The classification of membrane proteins by the fraction of secondary transmembrane helices may provide information about the functional features of the proteins.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, "Genome Science", from the Ministry of Education, Science, Sports and Culture of Japan.

References

[1] Frishman, D. and Mewes, H.W., Protein structural classes in five complete genomes, *Nature structural biology*, 4:626–628, 1997.

[2] Hirokawa, T., Seah, B.-C., and Mitaku, S., SOSUI: classification and secondary structure prediction system for membrane proteins, *Bioinformatics Application Note*, 14, 1988.