A Prediction Method for Transmembrane Segments in Proteins Utilizing Multiple Discrimination Functions

Daisuke Kihara [1] Toshio Shimizu [2] Minoru Kanehisa [1]

kihara@kuicr.kyoto-u.ac.jp slsimi@si.hirosaki-u.ac.jp kanehisa@kuicr.kyoto-u.ac.jp

[1] Institute for Chemical Research, Kyoto University
Gokasho, Uji, Kyoto 611, Japan

[2] Department of Information Science, Faculty of Science, Hirosaki University
3 Bunkyo-cho, Hirosaki 036, Japan

Abstract

A novel method for detecting transmembrane segments in membrane proteins is presented. The method employs discriminant analysis to detect transmembrane regions in a query sequence by taking advantage of the information of different characteristics in different subgroups of membrane proteins classified according to the number of transmembrane segments. The method was applied to sequences of membrane proteins with experimentally determined locations of transmembrane segments.

1 Introduction

Membrane proteins play important roles in biological systems. While there is an increasing amount of sequence data available, only a few three-dimensional structures have been solved by experimental methods. Under these conditions, there is a growing need for computational methods to give insights into the structure of membrane proteins. The starting point of the computational analysis is to predict the location of transmembrane segments in a given sequence. A number of methods have been proposed thus far, and they are typically based on hydropathy analysis [1]. Various improvements were performed in recent studies but their ability of prediction seem to be still not enough. For example, they often miss to predict the last, relatively hydrophilic, segment in proteins that consist of seven transmembrane segments.

Most of the previous methods consider common characteristics of transmembrane segments in all membrane proteins. In contrast, the basic idea of the present study is to utilize the information of different characteristics of transmembrane segments in different groups of membrane proteins. We employ discriminant analysis, which was also utilized in the previous work [2].

2 Method

The amino acid sequences of membrane proteins in TMbase [3] were grouped by the number of times the proteins span the membrane. The segments in each group were again grouped into "subsets" by the order they span the membrane; for example, the seventh segments in the seven-spanning membrane proteins form a subset. Each segment is described by a set of variables, average hydrophobicity, amphiphilicity and polarity. Thus, a subset is characterized by the distribution of these variables. When the subsets are judged by discriminant analysis to be similar, they are fused to form a "set". Discrimination functions are constructed for each set; therefore, each group has a set of discrimination functions that represents the model of the group.

For prediction, a query sequence is compared with each model and the score is calculated. This is analogous to the protein fold recognition approach where the amino acid sequence is compared with each of the three-dimensional profiles of the representative folds (models). Here the discrimination function for each set in a model detects candidates for transmembrane regions independently, so that the combination of these candidates that gives the highest score is adopted for the model. After comparing with all the models, the highest scoring model is considered to be most reliable, but it is also possible to display models whose scores are almost on a par with the highest one.

We have collected from references sequences of membrane proteins whose locations of transmembrane segments are determined experimentally. The method was applied to them as test sequences.

Acknowledgment

The authors thank Dr. Kenta Nakai for helping with collecting the test data and for helpful discussions. 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 in Japan.

References

[1] J. Kyte, R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein." *J.Mol.Biol.*, 157, pp. 105-132, 1982.

[2] P. Klein, M. Kanehisa, C. DeLisi, "The detection and classification of membrane-spanning proteins." *Biochim. Biophys. Acta*, 815, pp. 468-476, 1985.

[3] K. Hoffmann, W. Stoffel, "TMbase - A database of membrane spanning protein segments." *Biol. Chem. Hoppe-Seyler*, 374, pp. 166, 1993.