Construction of a Membrane Protein Database and an Evaluation of Several Prediction Methods of Transmembrane Segments

Toshio Shimizu [1] slsimi@si.hirosaki-u.ac.jp

Kenta Nakai [2] nakai@nibb.ac.jp

[1] Faculty of Science, Hirosaki University
3 Bunkyo-cho, Hirosaki 036 Japan

[2] National Institute for Basic Biology
Myodaiji, Okazaki 444 Japan

Abstract

How reliable and useful are predictions of transmembrane segments(TMSs) of membrane proteins from the amino acid sequences? It remains still under debate. Kyte and Doolittle proposed a simple scheme for the prediction of TMSs [1]. It is based on the hydropathy plot and is widely accepted as a basic and standard method. Since then, a large number of more sophisticated predictive algorithms have been proposed, which are improved varieties of the Kyte-Doolittle's approach. Although these methods have been considered to give rather good results, their abilities are still not enough to predict the number and positions of TMSs precisely; they often give totally different predictive results with proteins having many TMSs, in particular [2, 3]. One reason for this situation can be attributable to the low quality of the information on TMSs described in general amino acid sequence databases. The information included within the SWISS-PROT database, for example, is mostly not based on any experimental evidence but on predicted models; there is often no explicit description about whether the data comes from experiments or calculations in databases. Higher quality of information on TMSs from experimental evidence only is essential to evaluate existing prediction methods more precisely and to develop an algorithm overcoming their problems.

We have collected 128 references reporting the membrane topology of proteins, and are continuing our efforts to triple this number. From them, we selected 54 topology models based on experimental evidence, at least partially. Combining these data with the sequence information from the SWISS-PROT database, we have constructed a membrane protein database in the form of relational database. Current version includes 54 proteins which are classified into 3

[1]清水 俊夫：弘前大学理学部，〒036 弘前市文京町 3

[2]中井 謙太：岡崎国立共同研究機構 基礎生物学研究所，〒444 岡崎市明大寺町字西郷中 38

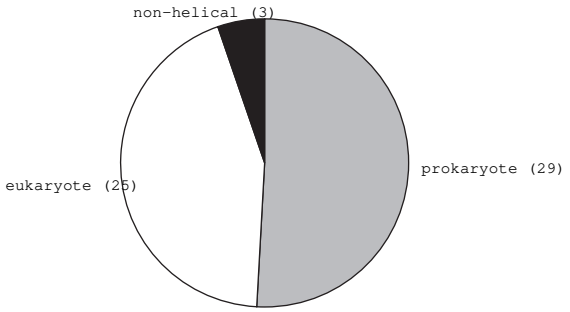

Figure 1: Content of Membrane Protein Database. Number of data in parentheses

groups (eukaryotic proteins, prokaryotic proteins, and the proteins with non-helical segments) as shown in Figure 1. Using this database we evaluated the predictability of the algorithms of following authors: Eisenberg [4]; Klein, Kanehisa and DeLisi(KKD method) [5]; von Heijne(TopPred method) [6]; and Persson and Argos [7]. The KKD method and the TopPred method predicted the exact number of TMSs for 59% and 67% of proteins in our database, respectively. These values could be increased to 63% and 74% by optimizing respective parameter values. The KKD method tends to predict fewer number of TMSs than the correct number, while the TopPred method shows the opposite tendency. We are now testing our previous idea to use different cut-off parameters for one TMS proteins and multiple TMS proteins in the KKD method and are also trying to develop a new predictive algorithm, by taking more precise position-dependent information on TMS into account.

References

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

[2] Goffeau, A., Slonimski, P., Nakai, K. and Risler, J.-L., "How many yeast genes code for membrane-spanning proteins?," *Yeast*, Vol. 9, pp. 691-702, 1993.

[3] Goffeau, A., Nakai, K., Slonimski, P., and Risler, J.-L., "The membrane proteins encoded by yeast chromosome III genes," *FEBS Lett.*, Vol. 325, pp. 112-117, 1993.

[4] Eisenberg, D., Weiss, R. M. and Terwilliger, T. C., "Three-dimensional structure of membrane and surface proteins," *Ann. Rev. Biochem.*, Vol. 53, pp. 595-623, 1984.

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

[6] von Heijne, G., "Membrane protein structure prediction: hydrophobic analysis and the positive-inside rule," *J. Mol. Biol.*, Vol. 225, pp. 487-494, 1992.

[7] Persson. B. and Argos, P., "Prediction of transmembrane segments in proteins utilizing multiple sequence alignments," *J. Mol. Biol.*, Vol. 237, pp. 182-192, 1994.