Topology Prediction of Membrane Proteins Based on a Modified "Positive-Inside" Rule

Koji Kadota kadota@cc.tuat.ac.jp
Hiroyuki Sasamoto sasamoto@cc.tuat.ac.jp
Takatsugu Hirokawa hirokawa@cc.tuat.ac.jp
Shigeki Mitaku mitaku@cc.tuat.ac.jp

Department of Biotechnology, Tokyo University of Agriculture and Technology
Nakacho 2-24-16, Koganei, Tokyo 184-8588, Japan

1 Introduction

It is difficult to determine the 3D structure of integral membrane proteins by experimental techniques. However, theoretical prediction of the secondary structures from amino acid sequences appears much easier for intrinsic membrane proteins than soluble proteins, because of steric constraints from membrane structure and weak hydrophobic interaction within the lipid bilayer. Then, the next step is to predict their transmembrane topology. It has been established that positively-charged amino acids (Lys and Arg) tend to be more prevalent in the cytoplasmic region of membrane proteins than in the external regions, which is called a positive-inside rule.

In order to improve the accuracy of the topology prediction, we have incorporated another generally accepted rule to the problem, in which long loop-segments of membrane proteins tends to be in the extra-cytoplasmic space rather than in the cytoplasmic space (say, the long-loop-outside rule). In previous works [1,2], loops and tails longer than 60 residues were excluded, since they do not conform to the positive-inside rule. As a complementary approach, we have normalized the magnitude of positive charges by the total length of loops and tails.

As a result, a score as good as 95% for the topology prediction was achieved by calculating the composition of positively charged residues (Lys and Arg) of the protein segments exposed at each side of the membrane, for the criterion (See, Methods). In the dataset, 81 out of 120 membrane protein sequences have actually longer extra-cytoplasmic segments, in total, than cytoplasmic ones.

2 Method

The 120 amino acid sequences, of membrane proteins with known topology and TM regions were collected from the dataset by Rost et al. [3]. Concerning each protein in the dataset, we calculated the quotinent of the number of positively-charged residues (Arg and Lys) [1], divided by the corresponding length of loops and tails. The topology was determined by the comparison between the normalized number of the positively charged residues at the cytoplasmic side and the corresponding value at external side. When the normalized number of positively charged residues is larger at the N-terminal side, this side is located at the cytoplasmic side. Every topology of the 120 membrane proteins was judged according to this criterion.

3 Results and Discussion

The positive-inside rule is well established for topology prediction of membrane proteins. In this paper, we proposed a new method for the topology prediction based on a modified positive inside rule. By contrast, the difference of the composition of positively charged residues at the both sides of the membrane distinguished 95% of the topologies of the membrane proteins (114 out of 120) with known TM regions and topology. Since the previous method [1,2] for prediction of the topology is coupled with prediction of the TM regions concurrently, it is intrinsic hard to compare its score for the prediction with ours directly. However, our prediction accuracy is pretty good, and this might indicate that the composition of positively-charged residues in all the extra-membrane segments poises the topology of an integral membrane protein, with relation to the longer loops and tails on the outside of the membrane.

Although there is no eminent characteristic and difference of total length of the extra-membrane segments among the incorrectly predicted 6 amino acid sequences, one should notice that the three proteins have a signal peptide. Accordingly, our method could be useful to predict the topology of membrane proteins without signal peptides, for the present time. Proteins with signal peptides, and signal-anchor (SA-II) membrane proteins in endoplasmic reticulum have been extensively studied on the formation of the topology [4]. Such approaches will modify the topology prediction in the future.

A practical topology prediction should be carried out with prediction of the TM regions, from the amino acid sequence of a membrane protein. We have recently developed the system SOSUI [5], to find a TM segment and determine the TM regions in an amino acid sequence, but it is imperfect to detect all the TM regions for the present method with high accuracy. Therefore, better prediction of the topology by our method will be opened up if prediction of the TM regions by the SOSUI algorithm is improved.

4 Acknowledgments

The authors thank Dr S. Shimizu for critical reading and revision of the manuscript. 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] von Heijne, G., The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology, *EMBO J.*, 5:3021–3027, 1986.

[2] Sipos, L. and von Heijne, G., Predicting the topology of eukaryotic membrane proteins, *Eur. J. Biochem.*, 213:1333–1340, 1993.

[3] Rost, B., Fariselli, P., and Casadio, R., Topology prediction for helical transmembrane proteins at 86% accuracy, *Protein Sci.*, 5:1704–1718, 1996.

[4] Sakaguchi, M., Characterization of topogenic sequences which regulate the integration and membrane topology of proteins in the endoplasmic reticulum membrane, *Seikagaku*, 69(1):1–15, 1997.

[5] Hirokawa, T., Boon-Chieng, S., and Mitaku, S., SOSUI: Classification and Secondary Structure Prediction System for Membrane Proteins, *Bioinformatics*, 14(4):378–379, 1998.