On the Accuracy of Transmembrane Segment Prediction of Helical Integral Membrane Proteins

Shin Tanimoto[1]
shin-ta@is.aist-nara.ac.jp

Gautam Basu[2]
gautam@is.aist-nara.ac.jp

Takeshi Kawabata[1]
takawaba@is.aist-nara.ac.jp

Nobuhiro Go[1,2]
ngo@nyc.odn.ne.jp

[1] Graduate School of Information Science, Nara Institute of Science and Technology, Nara 630-0192, Japan
[2] CCSE, Japan Atomic Energy Research Institute, 8-1 Umemidai, Kizu-cho, Souraku, Kyoto 619-0215, Japan

Keywords: transmembrane helix prediction, hydrophobicity, topology

1 Introduction

Integral membrane proteins play a vital role in a number of essential biological functions. Although abundant, about 30% of genes are known to code for membrane proteins, the number of solved structures in the pdb is less than 1%. Thus, structure prediction of membrane proteins is an essential tool for understanding their functions. A fundamental characteristic of the predicted structure is the topology – identification of trans-membrane segments and the overall orientation with respect to the membrane (intra- or extra-cellular). Several prediction methods have been developed for this purpose, both knowledge-based and residue hydrophobicity-based. Although the performances of almost all of these methods are rather high, short loops and long helices are predicted less accurately [1]. One of the problems of estimating accuracy of different prediction methods is the absence of experimentally reliable trans-membrane annotations to compare with. Thus, one is forced to compare prediction versus prediction, where the assumed transmembrane segments typically are identical to entire helix lengths of putative membrane-spanning segments in the pdb structure. Adding hydrophobicity information to the structural data, and by optimizing the resulting membrane-protein mismatch energy, we provide alternate assignments of trans-membrane segments of membrane proteins whose structures are known. This new assignment is used to revaluate the prediction accuracy of established prediction methods and hydrophobicity-based methods. The implications of our results are discussed.

2 Method

The membrane bilayer is modeled as a rectangular slab of 30Å thickness with the z-axis being perpendicular to the bilayer plane. The protein is treated as a rigid body with three degrees of freedom (Fig. 1). A mismatch energy, E(z, θ, ϕ), given by the product of residue hydrophobicity Hi and residue fractional solvent accessibility Ai, summed over all residues lying within the bilayer for a given conformation, is calculated. Finally the conformation with the lowest energy is chosen to be the most likely orientation representing the protein-membrane complex. The 3D MPTopo data set [2] is used for topology assignment and subsequent prediction. The assigned topology (called OM: Orient Membrane) is compared with transmembrane segments predicted by published knowledge-based methods and an in-house hydrophobicity-based program.

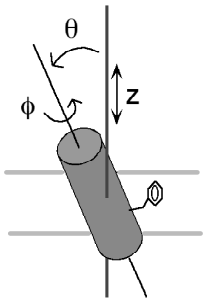

Figure 1:

3 Results and Discussion

Fig. 2a shows a typical energy surface (calcium transporting ATPase; pdb: 1eul), the corresponding MPTopo trans-membrane helix annotations and the OM-annotations are shown in Fig. 2b and Fig. 2c (the transmembrane segments are shaded). This example demonstrates the essence of the OM assignment against the rather arbitrary nature of MPTopo assignments. The difference is also clearly reflected in the length distribution of OM and MPTopo assignments, computed over the entire MPTopo database (Fig. 2d). In general the OM-assigned transmembrane segments are shorter, and indeed, as shown in Fig. 2d, they compare better with predicted length distributions, either KD [3] hydrophobicity-based or by a typical knowledge-based method, TMHMM [4].

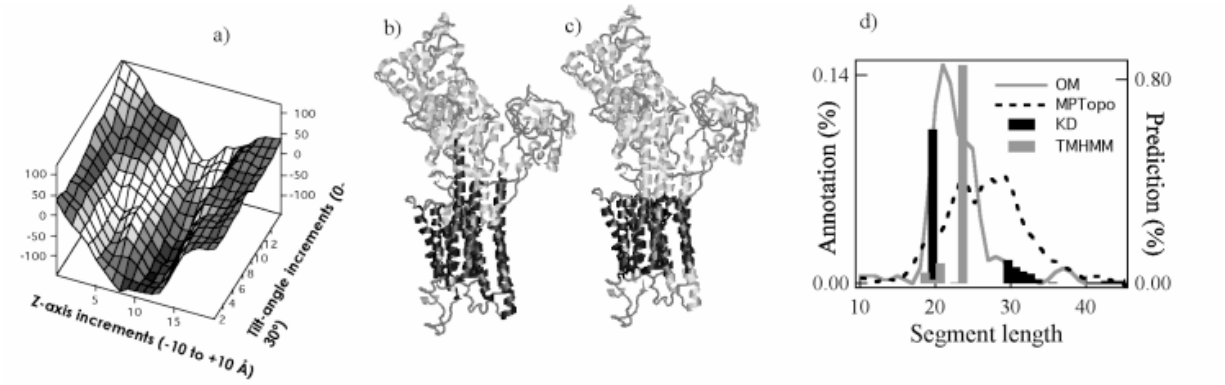

Figure 2:

In conclusion, by employing structure/hydrophobicity-based alternate annotation for actual transmembrane segments, we have shown that transmembrane helix prediction methods are not poor predictors of long helices, rather, a more consistent method of transmebrane segment annotation, as employed here should replace the currently used annotation method. Further, based on the observed merits and limitations of hydrophobicity-based methods, we propose that rather than expecting hydrophobicity-based methods to precisely predict transmembrane segments, such predictions should be taken at their face value and used to gain insights into the nature of the resulting complex.

References

[1] Chen, C.P. and Rost, B., Long membrane helices and short loops predicted less accurately, *Protein Sci.*, 11:2766–2773, 2002.

[2] Jayasinghe, S., Hristova, K., and White, S.H., MPTopo: a database of membrane protein topology, *Protein Sci.*, 10:455–458, 2001.

[3] Kyte, J. and Doolittle, R.F., A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.*, 157:105–132, 1982.

[4] Sonnhammer, E.L.L., von Heijne, G., and Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences, *6th International Conference on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, CA, 175–182, 1998.