

Classification and Analysis of Inter-Helical Loop Segments in Membrane Proteins

Takatsugu Hirokawa

t-hirokawa@aist.go.jp

Makiko Suwa

m-suwa@aist.go.jp

Yutaka Akiyama

akiyama-yutaka@aist.go.jp

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koutou-ku, Tokyo 135-0064, Japan

Keywords: membrane protein, inter-helical loop segment, transmembrane helix, sequence analysis

1 Introduction

The discovery and high-throughput structure modeling of specific membrane protein families, such as G-protein coupled receptors (GPCRs) and channel proteins from complete genome sequences, are important issues in pharmacogenomics and structure-based drug design. Membrane protein families from other protein sequences have been detected and classified with high accuracy by means of homology search, hidden Markov model, physicochemical profiles, and neural networks [5].

Structure modeling of membrane proteins, by (1) comparative modeling of GPCRs using the structural template of bovine rhodopsin and (2) several non-statistical approaches has also been proposed [2, 3, 8, 7]. Transmembrane (TM) regions in membrane proteins play a major role in structure modeling due to the functional insights they provide and their simple architecture, which consists of a membrane spanning α -helix composed of approximately a 20-30 amino acids. On the other hand, inter-helical (IH) loop segments that link TM helices other than the large loop segments formed in a compact soluble domain, have been omitted from the structure modeling process. However, a couple of recent crystallography and diffraction studies of membrane proteins, such as potassium and water channel proteins, have revealed that the specific medium loop segments between TM helices fold back onto membrane positions and play an important role in membrane protein folding and biological function [1, 6]. Previous methods of TM helix prediction have not considered the contribution of these IH loop groups. Consequently, specific medium loop segment is often predicted as a TM helix. This false-positive prediction causes serious problems in TM topology prediction and structure modeling.

Here, we analyzed the IH loop segments in proteins having multi-TM helices for classification and detection of the specific medium loop segment from amino acid sequences. The results show that the specific loop length and extra-cellular environment are dominant factor for the classification of IH loop types.

2 Method and Results

The amino acid sequences and three-dimensional coordinates of IH loop segments were derived from 11 membrane proteins having multiple TM helices in a chain (PDB ID: 1qhj, 1hw0, 1aijL, 1aijM, 2occA, 2occB, 2occC, 1f88, 1bl8, 1eul and 1be3C). The total number of IH loop segments was 57.

First, we classified IH loop segments into two types, surface buried and exposed loops, based on the number of atom-atom contacts at the IH loop segment-TM regions interface in membrane proteins. The tentative classification boundary between buried and exposed types was composed of 70 atoms in the present work. Each IH loop segment was characterized by amino acid length, intra/extra-cellular

topology classification from experimental studies, and the average hydrophobicity and the intensity of amphiphilicity of helices, which we obtained using Kyte-Doolittle scale [4]. An additional two analyses were also carried out for classification of IH loop types: a statistical analysis of the position-specific distribution of amino acid residues in each IH loop type and a calculation of the average hydrophobicity of TM helix segments in a chain. A total of seven classification factors for buried loop segments were obtained, and the relationship between these factors and buried/exposed IH loop types were analyzed on dispersion diagrams.

The factors are the upper and lower limit of length (15-38 residues), extra-cellular loop segments, a strong average hydrophobicity and intensity of amphiphilicity, relatively hydrophobic TM helices for covering buried loops, and the position-specific distribution of tyrosine and proline residues. The first and second factors, length and topology, are useful in classifying the buried/exposed IH loop types. Disulfide bonding between a loop segment and TM helix in the rhodopsin resulting from three-dimensional coordinate also seems like a strong interaction for folding the IH loop segment at the membrane position. A more structural determination of membrane proteins in the rhodopsin as well as other families is necessary in order to elucidate the contribution of disulfide bonding to loop folding.

References

- [1] Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R., The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity, *Science*, 280:69–77, 1998.
- [2] Du, P. and Alkorta, I., Sequence divergence analysis for the prediction of seven-helix membrane protein structures: I. Comparison with bacteriorhodopsin, *Protein Eng.*, 7:1221–1229, 1994.
- [3] Hirokawa, T., Uechi, J., Sasamoto, H., Suwa, M., and Mitaku, S., A triangle lattice model that predicts transmembrane helix configuration using a polar jigsaw puzzle, *Protein Eng.*, 13:771–778, 2000.
- [4] Kyte, J. and Doolittle, R.F., A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.*, 157:105–132, 1982.
- [5] Möller, S., Croning, M.D., and Apweiler, R., Evaluation of methods for the prediction of membrane spanning regions, *Bioinformatics*, 17:646–653, 2001.
- [6] Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., and Fujiyoshi, Y., Structural determinants of water permeation through aquaporin-1, *Nature*, 407:599–605, 2000.
- [7] Sansom, M.S. and Davison, L., Modeling transmembrane helix bundles by restrained MD simulations, *Methods Mol Biol.*, 143:325–347, 2000.
- [8] Taylor, W.R., Jones, D.T., and Green, N.M., A method for alpha-helical integral membrane protein fold prediction, *Proteins*, 18:281–294, 1994.