A computer system for predicting membrane protein structure

Takatsugu Hirokawa
hirokawa@cc.tuat.ac.jp

Seah Boon Chieng
seah@cc.tuat.ac.jp

Makiko Suwa
suwa@cc.tuat.ac.jp

Shigeki Mitaku
mitaku@cc.tuat.ac.jp

Faculty of Technology, Tokyo University of Agriculture and Technology
Nakacho 2-24-16, Koganei, Tokyo 184, Japan

There are two kinds of approaches to the analysis of amino acid sequences. One is to relate particular sequences to various characteristics of proteins, such as structure and function, by the statistical calculation of biological databases. The other approach is the simulation of physicochemical processes such as the protein folding and the functioning. The present system is adopting the latter approach, particularly focusing to membrane proteins, because the mechanism of the folding of membrane proteins is simpler than that of soluble ones.

There are two stages in the folding of membrane proteins: the penetration of polypeptide segments into membrane, simultaneously forming transmembrane helices, and the formation of tertiary structure in membrane. The two processes in membrane proteins are driven by characteristic nonbonded interactions. The penetration of polypeptide segments is mainly due to the hydrophobic interaction, and the segments burried in membrane simultaneously form a-helices. The tertiary structure in membrane is formed by several factors: the polar interactions, the length of loop segments and the effect of lipids. Therefore, physicochemical method for the prediction of membrane protein structure should be based upon the dominant interactions.

Here, we are developing a new software system for the structure prediction of membrane proteins, which comprises of three parts. First part deal with the discrimination of membrane proteins from other types of proteins and the secondary structure prediction of membrane proteins. Second part is the tertiary structure prediction, which is equivalent to the prediction of the positioning of transmembrane helices. Third part connects the molecular structure to the function. The first two parts are already complete, and the discrimination of membrane proteins, the secondary structure prediction and the determination of helix positioning are possible for amino acid sequences by the present system.

The first part is the analysis of the hydrophobicity of amino acid sequences [1, 2]. It comprises of two modules: (1) the discrimination of membrane proteins from other kinds of proteins and (2) the secondary structure prediction. An amino acid sequence is first converted to the sequence of hydrophobicity, from which two parameters are calculated: the average value of the total sequence and the power spectral density for the period beyond 30 or 20 residues. When many protein are plotted in a dispersion diagram of the average hydrophobicity versus the power sepectral density beyond 30 residues, membrane proteins are located at the area of high hydrophobicity and high long periodicity.

広川貴次，謝 文清，諏訪牧子，美宅成樹：東京農工大学工学部，〒184 東京都小金井市中町 2-24-16

The system correctly distinguishes multihelical membrane proteins with the reliability of 99 %. The second module of the secondary structure also uses the averagy hydrophobicity and the power spectral density. In this case, however, the window of 7 residues was used for the calculation of average. The power sepctrum is calculated by a maximum entropy method with the window of 15 residues, and the helix periodicity around 3.6 residues is used for the secondary structure. The predictability of the region of transmembrane helices is as good as 90 %.

The second part of the prediction system deals with the positioning of transmembrane helices. When amino acid sequences of transmembrane segments are given, they may be folded into a-helices in a graphic workstation. Then, the interaction between the helices is calculated by using the energy calculation software, QUANTA and CHARMm, on a workstation, IRIS 4D/35. However, the exact calculation of all pairs of helices is too elaborate that the procedure for the estimation of the binding energy between helices was simplified from two aspects. First, it is assumed that the polar interactions are dominant, which is physically sound because the dielectric constant in membrane is very low. Second, the polar interaction fields are calculated by a probe helix method, and the binding energy between helices are estimated from the product of the polar interaction field of two helices. In the probe helix method a transmembrane helix and a probe helix of poly-serine-poly-alanine are constructed on the graphic workstation and the polar interactions are calculated. Rotating the transmembrane helix by 15 degrees and changing the location of poly-serine in the probe helix, 48 numbers were obtained for the polar interaction field. The product of the fields of two transmembrane helices was roughly proportional to the binding energy by the exact calculation. Consequently, the use of the polar interaction energy field by the probe helix method greatly reduces the computational time of the binding energy.

The software module for the helix positioning was developed for two cases. When we have a two-dimensional electron density map of low resolution, the problem becomes the combination of transmembrane segments in amino acid sequence and the electron density peaks that lead to the free energy minimum of the system [3]. On the other hand, the positioning of helices has to be completely simulated, if we have no information about its structure. Two factors are taken into account in the former case, the polar interactions and the length of loop segments, whereas three factors are necessary for the simulation adding the interaction between helices due to the entropy decrease by the boundary lipids. The applicability of the methods was tested for bacteriorhodopsin, whose three dimensional structure has been reported [4]. The module of the former problem gave good results. When the effect of a prosthetic group, retinal, was considered, the lateral and the rotational positioning of the transmembrane helices of bacteriorhodopsin could be reconstructed by the theoretical calculation [3]. Therefore, our algorithm is expected to be applicable to other membrane proteins. The latter module for the complete simulation is not yet finished.

We will show the system for the structure prediction of membrane proteins at the workstation session of the Genome Informatics Workshop. This work was partly supported by the Grant-in-Aid for the Priority Area "Genome Informatics" from the Ministry of Education, Science and Culture.

References

[1] Yanagihara, N. Suwa, M. and Mikaku, S, Biophys. Chem, 34, 69–77, 1989.

[2] Suwa, M. Mikaku, S. and Kuroda, Y, Biochem. Biophys. Research Communications, 191, 782–789, 1993.

[3] Suwa, M. Hirokawa, T. and Mitaku, Proteins, 1994 submitted.

[4] Henderson, R., Baldwin, J. M. Cheka, T. A., Zemlin, F., Beckman, E., Downing, K. H., J. Mol. Biol, 213, 899–929, 1990.