A Method to Predict Transmembrane Helix Arrangements in Receptor Proteins by Analyzing Size-Compensated Substitution Pairs

Kiminobu Sato [1] **Toshio Shimizu** [2]
xsat@cc.hirosaki-u.ac.jp slsimi@si.hirosaki-u.ac.jp
Department of Information Science, Faculty of Science, Hirosaki University
Bunkyocho 3, Hirosaki 036, Japan

Abstract

An algorithm for prediction of transmembrane helix arrangements is presented. Transmembrane helix arrangements are predicted on the basis of size-compensated substitution pairs of amino acid residues. Those pairs are extracted from an multiple sequence alignment for helix regions of an transmembrane protein. The constraint condition is employed with the helical wheel representation of each helix to filter out more inconsistent pairs. The method is applied to amino acid sequences of Bacteriorhodopsin, and its arrangements are predicted.

1 Introduction

Because of the circumstance in membrane, obtaining tertiary structures of transmembrane proteins is more difficult than of soluble proteins, because of the difficulty of crystallization. Only a few tertiary structures are known so far, in spite of their functional importance and the increasing number of sequences. Tertiary structures (helix arrangements) of transmembrane proteins are surrounded by hydrophobic environment, and highly conserved to retain their functions. We assume that size-compensated substitution between amino acid residues occurred to conserve helix arrangements; for example, while one amino acid is substituted for smaller one, the other should be for larger [1]. Size-compensated substitution pairs are identified in a multiple alignment that each sequence corresponds to helix region, and extracted.

2 Method

20 amino acids are classified into three classes based on molecular weight, and an integer is given to each class as the class value : Class I (small amino acids)given '2'; Class II (medium),'4'; Class III (large),'8'. A multiple sequence alignment for each helix is converted to a matrix consists of three integers. Next, the matrix is converted to a new matrix of which element is a substitution pattern value represented by the difference between class values. For instance, when an amino acid of Class I changes into Class III one, the substitution pattern value is given as 6, in the case that Class III into Class I, the value is -6. Size-compensated substitution pairs are identified if the following conditions are satisfied simultaneously:

$$x_{l_{i_1 j}} x_{l_{i_2 j}} < 0, \tag{1}$$

$$M > 0.264K, \tag{2}$$

$$\sigma^2 \leq 0.100, \tag{3}$$

where $x_{l_{i_1}}$ and $x_{l_{i_2}}$ are substitution pattern values at columns l_{i_1} and l_{i_2} respectively for N_{min} out of K rows ($N_{min} \leq K$), K is the combination numbers ,*i.e.*, $K = N_{all}(N_{all}-1)/2$ where N_{all} is the total

number of sequences. M is the number of rows satisfying (1). If inequalities (1), (2) are satisfied at column l_{i_1} and l_{i_2} for M out of K rows ($N_{min} \leq M \leq K$), the standard deviation, σ is calculated as follows:

$$\sigma^2 = \frac{1}{M}\sum_{i=1}^{M}(|x_{l_{i_1}} + x_{l_{i_2}}| - \overline{|x_{l_{i_1}} + x_{l_{i_2}}|})^2.$$

Then, the helix-structural constraint condition is defined as following to filter out inconsistent pairs extracted in the above step.

$$\frac{Z_i + Z_j}{2}|\frac{z_{ik}}{Z_i} - \frac{z_{jl}}{Z_j}| \leq n_d,$$

where Z_i, Z_j are total numbers of amino acid residues in helices i, j respectively, and z_{ik}, z_{jl} are the position numbers counted from the cytoplasmic side for helices i, j respectively. n_d is residues contained in two pitches of helical turn corresponding to $12\mathring{A}$ (8 residues) along the axis of helix.

A prediction algorithm for transmembrane helix arrangements is described as follows:

- Step 0-1: The survived pair is allocated onto its corresponded position on each helical wheel.
- Step 0-2: Select the max position for each helical wheel.
- Set $K \leftarrow 0$.
- Step K: Let $P(P_K, P_{K+1})$, $P(\overline{P}_{K+1}, P_{K+2})$ be SCSPs (Size-Compensated Substitution Pairs) for helix pairs$(K, K+1)$, and $(K+1, K+2)$. Let $\mathbf{mp}$, $\mathbf{p}_{K+1}$ and $\overline{\mathbf{p}}_{K+1}$ be the max position vector, and the SCSP position vectors for positions P_{K+1} and $\overline{P}_{K+1}$, from the center of helical wheel $K+1$, respectively. If $< \mathbf{mp}, \mathbf{p}_{K+1} > \cdot < \mathbf{mp}, \overline{\mathbf{p}}_{K+1} > \geq 0$, $P(P_K, P_{K+1})$ and $P(\overline{P}_{K+1}, P_{K+2})$ are identified as the candidates of contact pairs. Next, calculate the angle between $\mathbf{p}_{K+1}$ and $\overline{\mathbf{p}}_{K+1}$ ($20(P_{K+1} - \overline{P}_{K+1})$ degree).
- $K \leftarrow K + 1$.
- Iterate Step K while $K \leq N - 2$ (N is the total number of helices).

3 Results and Conclusions

Helix arrangements of *Bacteriorhodopsin* are predicted by using the above method. An unique arrangement of helices A, B, C, D and E is predicted, and approximately coincide with the arrangement in the X-ray structure. The arrangement among helices E, F and G, however, is not determined uniquely. This may be caused by neglecting effects such as polar interactions between side chains and the existence of retinal etc., in this prediction procedure. Our results suggest that positions of helices on the two-dimensional plane can be depicted approximately only from the amino acid sequence analysis. It may be possible to apply our method, for example, to identify the position for each helix on the electron microscopic picture of transmembrane proteins such as *Rhodopsin* of which the complete tertiary structure is unknown despite the electron microscopic picture has already obtained.

References

[1] I.N. Shindyalov, N.A. Kolchanov, and C. Sander, "Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations," *Protein Engineering*, vol.7, no.3, pp.349–358, 1994.