

Sequence Alignment Tool of Membrane Protein

Takahiko Muramatsu^{1,2}

taka-mu@is.naist.jp

Makiko Suwa^{1,2}

suwa@cbrc.jp

¹ Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), 8916-5, Takayama-Chuman Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

² Computational Biology Research Center (CBRC), Aomi Frontier Building 17F, AIST, 2-43 Aomi, Koto-ku, Tokyo 135-0064, Japan

Keywords: sequence alignment, pair hidden markov model, membrane protein

1 Introduction

One of the major challenge in the bioinformatics field functional analysis of membrane proteins, which consist of one fourth of all proteins. Number of membrane protein 3-D structure has gradually been increasing with the great advances in experiment techniques, but . Effective use of their structures is to predict desirable structure (Target) from given structure (Template).

One of the most important steps in comparative modeling is computation of template-target alignment. It is not so easy to perform their automatic alignment, because of the feature that membrane protein have two different types of domains, hydrophobic Transmembrane domain and hydrophilic globular domain. Sequence alignment method is developed based on protein structure, transmembrane topology. In this work, we propose sequence alignment tool for alpha helical transmembrane protein, combining sequence alignment and prediction of transmembrane topology. Our method refines membrane protein alignment and perform more accurate prediction of structure.

2 Method

The concept of our method is to predict membrane topology and sequence alignment at the same time. In order to realize it, we use one of probabilistic models, Pair Hidden Markov Models (PHMMs)[3], which emits symbol pair. Figure 1. is shown as this model. The model generates more than single pair of symbol based on transmembrane length distribution in membrane states, and fixed 4 pairs of symbol in cytoplasmic or extracellular cap states. In this model there are no gaps in the membrane and cap regions. In each cytoplasmic and extracellular loop the model have three states. Match state can emit amino acid pair, and Insertion (Ins) and Deletion (del) states can emit one amino acid and one gap pair. On this model we calculate the optimal state sequence to observed two sequences by dynamic programming, and perform sequence alignment.

3 Result and Discussion

Table 1. shows the results of the benchmark test. We evaluate the performance of pairwise alignment with pairwise score. Two benchmark test set, family dataset and superfamily dataset, based on structural alignment are used. Family dataset has 26 sequence pairs represents each protein and species level in the same membrane protein Family and are less than 30 percents sequence similarity. Next, Superfamily dataset has 18 sequence pairs represents each protein in the same membrane protein

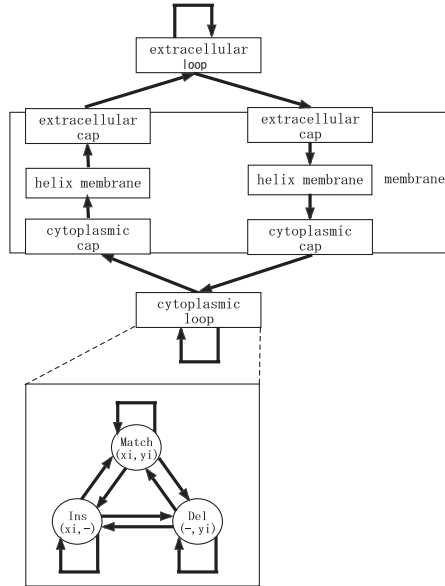


Figure 1: Pair hidden markov model for membrane protein sequence alignment

superfamily and are less than 30 percents sequence similarity. Our method is compared to standard global alignment with several amino acid substitution matrix. JTT is PAM based matrix and PHAT is based BLOSUM (BL) for membrane regions.

The results show that performance of our method improves standard global alignment with any substitution matrix for both datasets. This refinement of sequence alignment enhances accuracy of the homology model quality.

Table 1: Evaluation of pairwise sequence alignment method

| | PHMM | standard global alignment | | | | |
|--------------|-------|---------------------------|-------|---------|--------|-------|
| | | BL62 | BL50 | PAM 250 | PHAT75 | JTT |
| Family | 0.705 | 0.618 | 0.568 | 0.572 | 0.582 | 0.576 |
| Surperfamily | 0.661 | 0.557 | 0.510 | 0.530 | 0.540 | 0.529 |

References

- [1] Pachter, L., Alexandersson, M. and Cawley, S., Applications of generalized pair hidden Markov models to alignment and gene finding problems, *J. Comput. Biol.*, 9(2):389–399, 2002.
- [2] Jones, D.T., Taylor, W.R. and Thornton, J.M., A mutation data matrix for transmembrane proteins, *FEBS Lett.*, 339(3):269–275, 1994.
- [3] Ng, P.C., Henikoff, J.G. and Henikoff, S., PHAT: a transmembrane-specific substitution matrix. Predicted hydrophobic and transmembrane, *Bioinformatics* , 16(9):760–766, 2000.