β-sheet Prediction using Inter-strand Residue Pair Propensities

Minoru. ASOGAWA [1]
asogawa@csl.cl.nec.co.jp

[1] Massively Parallel Systems NEC Laboratory, RWCP
4-1-1, Miyazaki, Miyamaeku, Kawasaki, Kanagawa 216, Japann

Abstract

Usually two residues of a pair in a β-sheet separated long distance in a protein sequence. Therefore, to achieve high prediction accuracy for a β-sheet, it is necessary to consider such inter-strand residue interactions. Since widely used methods for a β-sheet prediction are only based on subsequence of residues, generally 20 residues, they cannot achieve high prediction. In this paper, we describe a novel method to predict an anti-parallel and parallel β-sheet, utilizing residue pair propensities, which are calculated from the statistics of inter-strand residue pairs. With primary experiment, it is shown that residue pair propensities are consistent with the result of real protein experiment and can detect parallel β-sheets with high accuracy.

1 Residue Pair Propensities

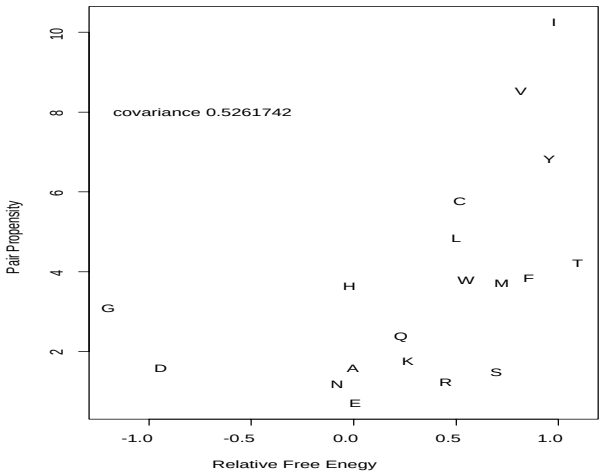
Residue pair propensities in β-sheets are calculated from PDB and HSSP. For the statistics, entries measured in high X-ray resolution (less than 2.5$\AA$) are selected. Although HSSP contains secondary structures for proteins, it is lacking hydrogen bond information. Therefore PDB is utilized to determine hydrogen bonded pairs in an anti-parallel β-sheet.

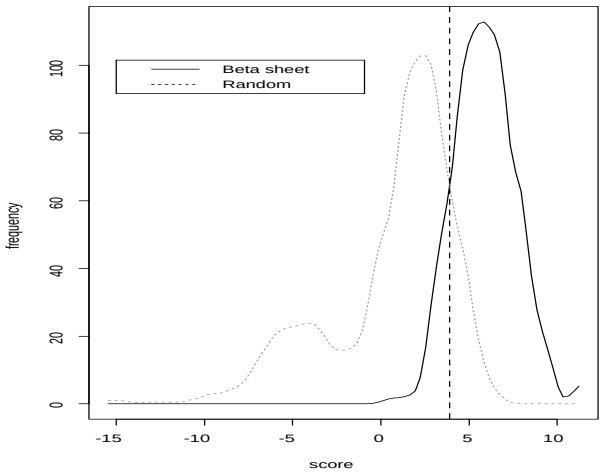
Residue pair propensity in an environment $E(R_1, R_2|H)$ is given as follows;

$$E(R_1, R_2|H) = \log \frac{Prob(R_1, R_2|H)}{Prob(R_1)Prob(R_2)} \tag{1}$$

where $Prob(R_1)$ is a probability of R_1 appears in a whole HSSP database, $Prob(R_2)$ is a probability of R_2, H is an environment, such as an anti-parallel β-sheet with hydrogen bond, an anti-parallel β-sheet without hydrogen bond, or a parallel β-sheet, and $Prob(R_1, R_2|H)$ is a conditional conjunctive probability of R_1 and R_2 is a pair under H.

[1] 麻生川 稔： Real World Computing Partnership 超並列 NEC 研究室 〒 213 神奈川県川崎市宮前区宮崎 4-1-1





2 Comparison with Real Experimental Data

It is assumed that, this propensity is closely related to the free energy of inter-strand residue pair. The left figure depicts the covariance of the pair propensities and relative free energies shown in [Minor 94]. This figure shows that residue pair propensities are consistent with the result of a real protein experiment. In Minor's experiment, IgG-binding domain from protein G is utilized. Relative free energies are measured by substituting residue 53, which is in a β-sheet. The residues adjacent to residue 53 are Ala(6) and Ala(44), Ala(6) is in a parallel β-sheet and Ala(44) is in an anti-parallel β-sheet. In the left figure, the vertical axis indicates the sum of pair propensities, $E(\mathrm{Ala}, \mathrm{Xxx}|H_{\mathrm{parallel}\beta\text{-sheet}})$ and $E(\mathrm{Xxx}, \mathrm{Ala}|H_{\mathrm{anti\text{-}parallel}\beta\text{-sheet}})$, where Xxx correspond to amino acids in the figure.

3 β-sheet Prediction Experiment

The left figure depicts the score of all anti-parallel β-sheets in length 6 (solid line) and that of randomly generated β-sheets in length 6 (dotted line). When a threshold is set to around 3.73, 88.21 percent of anti-parallel β-sheets are correctly classified and 86.84 percent of randomly generated sheets are correctly classified. In the case of length 2, 60.06 percent for β-sheets and 97.05 percent for randomly generated sheets, length 3, 73.53 and 89.26, length 4, 81.85 and 91.09, length 5, 81.23 and 90.64, length 7, 83.56 and 91.89, length 8, 77.46 and 93.44, length 9, 90.63 and 87.94, length 10, 87.93 and 89.39, respectively. Almost same results are obtained in a parallel β-sheets experiment. These accuracies are much superior than the generally used methods and might imply that this method is much suitable for β-sheets prediction. Furthermore, it is possible to utilize pair propensities for a folding simulation, such as in [Asogawa 94], and to predict all β-sheets in a protein, which leads to the tertiary structure prediction.

References

[Asogawa 94] Asogawa M., "Pseudo Folding Simulation for RNA Secondary Structures Prediction", *Proc. of FGCS 94 Workshop: Fusion of Molecular Biology and Knowledge Processing* (1994/December)

[Minor 94] Minor L. M. and Kim P. S., "Measurement of the β-sheet-forming propensities of amino acids", *Nature*, vol. 367, pp. 660-663. (1994).