Algorithms for Computing an Optimal Protein Threading with Profiles and Distance Restraints

Tatsuya Akutsu[1] takutsu@kuicr.kyoto-u.ac.jp
Morihiro Hayashida[1] morihiro@kuicr.kyoto-u.ac.jp
Etsuji Tomita[2] tomita@ice.uec.ac.jp
Jun'ichi Suzuki[2] jsuzuki@ice.uec.ac.jp
Katsuhisa Horimoto[3] khorimot@ims.u-tokyo.ac.jp

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
[2] Graduate School of Electro-Communications, The University of Electro-Communications, Chofu-city, Tokyo 182-8585, Japan
[3] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: protein threading, constraint, maximum clique, protein structure

1 Introduction

Protein threading is one of the powerful methods for protein structure prediction. Some advances are recently made by the significant utilization of distance restraints. Xu *et al.* proposed a protein threading method in which distance constraints obtained from NMR experiments were taken into account [5]. Young *et al.* recently developed a novel experimental method to increase the predictive accuracy of the fold recognition method by using chemical cross-linking and time-of-flight (TOF) mass spectrometry to identify LYS-LYS cross-links [6].

Though Xu *et al.* considered protein threading with pairwise contact energy and constraints [5], threading with profiles is also known as a powerful method for protein structure prediction. In particular, PSI-BLAST is widely used both for sequence similarity search and protein structure prediction [3]. Therefore, we have recently developed two algorithms for computing optimal threadings with profiles and constraints [1]. In this short abstract, we show performances of these two algorithms.

2 Method and Results

We developed two algorithms: CLIQUETHREAD and BBDPTHREAD [1]. CLIQUETHREAD reduces constrained threading to the maximum edge weight clique problem. In CLIQUETHREAD, we employ the maximum edge-weight clique algorithm developed by the authors, which was shown to be one of the fastest clique algorithms using DIMACS benchmark data [4]. BBDPTHREAD combines a DP (dynamic programming) algorithm and a branch-and-bound procedure, where the DP algorithm is developed based on a former work by the authors [2].

We performed computational experiments on CLIQUETHREAD and BBDPTHREAD in order to evaluate practical computation time and usefulness for improving the accuracy of profile threading. We used a PC cluster with Intel Xeon 2.8GHz CPUs and the LINUX operating system. But, only one CPU was used for each case. In order to obtain constraints for target proteins, we first calculated distances between C^{α} atoms of Lys-Lys residues, and then Lys-Lys pairs with the distances less than 24Å were only taken into account as constraints [6].

Table 1: Comparison of CPU times (sec.).

Target (#res)	1bbn (133)	1bla (155)	1xyzA (320)	1atnA (372)
Template (#res)	1cnt1 (150)	1hce (118)	8timA (247)	1atr (383)
CLIQUETHREAD	1.5	1.1	3279	NA
BBDPTHREAD	8.3	7.0	59.9	1101

Table 2: Comparison of threading results.

Target (#res)	1bbn (133)	1bla (155)	1xyzA (320)	1atnA (372)
Template (#res)	1cnt1 (150)	1hce (118)	8timA (247)	1atr (383)
Unconstrained	10.35Å (119)	15.06Å (106)	16.68Å (214)	14.41Å (314)
Constrained	8.97Å (127)	4.12Å (118)	13.15Å (197)	9.01Å (299)

Table 1 compares CPU times (seconds) of CLIQUETHREAD and BBDPTHREAD, where NA means that the execution of the program did not finish within 10 hours. In all cases except NA, feasible threadings were obtained and both algorithms output the same threading results. As seen in Table 1, CLIQUETHREAD is faster than BBDPTHREAD for small proteins, while BBDPTHREAD shows better performance than CLIQUETHREAD for large proteins (e.g., up to proteins with 300-400 residues).

The accuracies of obtained threadings are summarized in Table 2, where the number of superimposed residues and RMSD (*Root Means Squares Deviation*, Å) between the superimposed C^{α} atoms are shown for each case. As seen in Table 2, the accuracies obtained by constrained threading are in general better than those by unconstrained threading. It should be noted that much better results were obtained by constrained threading for 1bla/1hce and 1atnA/1atr pairs.

References

[1] Akutsu, T., Hayashida, M., Tomita, E., Suzuki, J., and Horimoto, K., Protein threading with profiles and constraints, submitted for publication.

[2] Akutsu, T. and Miyano, S., On the approximation of protein threading, *Theoretical Computer Science*, 210:261–275, 1999.

[3] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25:3389–3402, 1997.

[4] Suzuki, J., Tomita, E., and Seki, T., An algorithm for finding a maximum clique with maximum edge-weight and computational experiments, *Technical Report*, MPS-42-12:45–48, Information Processing Society of Japan, 2002.

[5] Xu, Y., Xu, D., Crawford, O.H., and Einstein, J.R., A computational method for NMR-constrained protein threading, *Journal of Computational Biology*, 7:449–467, 2000.

[6] Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., Taylor, E.W., Kuntz, I.D., Gibson, B.W., and Dollinger, G., High throughput protein fold identification by using experimental constraints derived from intermolecular cross-links and mass spectrometry, *Proc. Natl. Acad. Sci. USA*, 97:5802–5806, 2000.