PROSPECT: A Computer System for Globally-Optimal Threading

Ying Xu and Dong Xu
Computational Biosciences Section, Life Sciences Division
Oak Ridge National Laboratory, Oak Ridge, TN 37831-6480, USA
Email: {xyn, xud}@ornl.gov

May 28, 1999

This poster presents a new computer system, PROSPECT, for protein threading. PROSPECT employs an energy function that consists of three additive terms: (1) a singleton fitness term, (2) a distance-dependent pairwise-interaction preference term, and (3) an alignment gap penalty; and currently uses FSSP or DALI domain library as its threading template database. PROSPECT uses a divide-and-conquer algorithm to find an alignment between a query protein sequence and a protein fold template consisting of only core residues, which is guaranteed to be globally optimal for its energy function. Mathematically, PROSPECT finds a globally-optimal threading between a query sequence of n residues and a fold template of m residues and M core secondary structures in $O(nm + MnN^{1.5C-1})$ time and $O(nm + nN^{C-1})$ space [1], where C, the *topological complexity* of the template fold as we term, is a value which characterizes the overall "structure" of the considered pairwise interactions in the fold; and N represents the maximum number of possible alignments between an individual core of the fold and the query sequence when its neighboring cores are already aligned. Practically, it takes from several hours to several days on a workstation for a query sequence with up to 400 amino acids to finish threading against DALI domain template library. PROSPECT also allows a user to incorporate known biological constraints about the query sequence into the threading process. The system finds a globally-optimal threading which satisfies the constraints. PROSPECT can deal with restraints which reflect the geometrical relationships among residues of disulfide bonds, active sites, or determined by the NOE restraints from (low-resolution) NMR spectral data. Finally we will present the assessments on the performance of alignments accuracy, ranking of similar fold among the templates in the database, and the computing time using a threading test set [2].

References

[1] Xu, Y., Xu, D., and Uberbacher, E. C. An Efficient Computational Method for Globally Optimal Threading. *J. Comp Biol.*, **5**(3):597–614, 1998.

[2] Holm, L. and Sander, C. Decision support system for the evolutionary classification of protein structures. *ISMB*, **5**:140–146, 1997.