A divide and conquer approach to fast loop modeling in proteins

S. C. E. Tosatto (1), A. Cestaro (2) and R. Männer(1)

(1) ICM and LS Informatik V, Universität Mannheim, Germany
(2) CRIBI and Dept. of Biology, Università di Padova

Determination of protein structures that have not been solved experimentally is frequently done by comparative modeling techniques [1]. Copying parts of the target structure, which are assumed to be superimposable, from a known protein structure serves as a framework. Structurally variable regions, referred to as loops, have to be treated separately. Because loops often show the greatest variation in amino acid sequence and are usually less restrained in conformation than the core regions, they cannot easily be taken from the parent structure. Their prediction remains one of the main problems in comparative protein modeling [1]. One problem is the generation of a good set of alternative structures for evaluation with a scoring or energy function. To be used in typical comparative modeling situations, the loop modeling method should be fast enough to allow the rapid prediction of all loops in the protein.

We describe the newer developments in our fast ab initio method for modeling local segments in protein structures [2]. The algorithm is based on a divide and conquer approach and uses a database of precalculated look-up tables, which represent a large set of possible conformations for loop segments of variable length. The target loop is recursively decomposed until the resulting conformations are small enough to be compiled analytically. The algorithm, which is not restricted to any specific loop length, generates a ranked set of loop conformations in 10 to 90 seconds on a desktop PC. The prediction quality is evaluated in terms of global root-mean-square deviation (RMSD). Depending on loop length the top prediction varies between 1.06 Å RMSD for three residue loops and 3.72 Å RMSD for eight residue loops. Due to its speed the method may also be useful to generate alternative starting conformations for complex energy-based simulations.

A preliminary version of the loop modeling tool has been used in the CASP-4 structure prediction experiment to model insertions and deletions in comparative modeling and fold recognition targets. The tool has been integrated in our modeling package and both will soon be available as servers for web-based predictions.

[1] Moult J, et al. Proteins (1999), Suppl 3.
[2] Tosatto SCE, Bindewald E, Hesser J, Männer R. A Divide and Conquer Approach to Fast Loop Modeling. Protein Engineering (2002), in press.