Introduction

The extent to which an amino acid is accessible to the surrounding solvent is determined by the type and number of amino acids. In general, hydrophobic amino acids tend to be located near the solvent-accessible surface and hydrophilic amino acids toward the core of the protein. To measure this effect, several solvent exposure measures have been proposed.

Recently, a new promising solvent exposure measure, called HSE-exposure (HSE), has been proposed. While the CN measure uses a single sphere centered at a residue, the HSE measure considers two hemispheres. Two values, an up and a down value, are therefore associated with each residue corresponding to the upper and lower hemispheres. Hamelryck showed that the up and down values are almost uncorrelated and well conserved and could therefore be considered independently.

Protein Model

To render and decorate the protein conformation space, we require the coordinates of the amino acids to be positioned in a 3D lattice. A lattice can be defined by a set of basis vectors corresponding to the directions to the neighboring nodes. The basis vectors of a cubic lattice (SCC) are the cyclic permutations of \((\pm 1, \pm 1, \pm 1)\). The length of an edge between neighboring nodes is \(\sqrt{3}/2\), which is also the average distance between two consecutive SCC atoms in protein. A high coordination lattice has an underlying cubic lattice with unit length less than \(\sqrt{3}/2\). For some N, the high coordination lattice used in the experiments has parameters \(N = 8, \epsilon = 0.2\), which gives \(90\) basic vectors.

Experimental Results: Comparison of Heuristics

For both TS and MCM, 20 searches starting at random conformation are optimized in 20 minutes using the HSE potential for Proteolipid 1 (P129, length 16 amino acids).

Conclusion

We have developed a framework for minimizing the CNS-HSE potential. The search heuristic is based on TS with a novel tabu definition and it performs significantly better than MCM for this problem. The results of TS and MCM comparison show that structures found using the HSE potential are generally much closer to the native structure than structures found using the CNS potential. These results are found using short proteins (the largest protein has 36 amino acids). When using longer proteins, it becomes very difficult to find near-optimal solutions. Future research could therefore consider a more detailed potential function using secondary structure information.

References