A geometrical model for generating mainly-alpha protein conformations

ID - 160

Mucherino Antonio[1], Di Serafino Daniela[1,2], D'apuzzo Marco[1,2], Costantini Susan[2,3], Facchiano Angelo M.[2,3], Colonna Giovanni[2]

[1]Department of Mathematics, Second University of Naples, via Vivaldi 43, I-81100 Caserta, Italy

[2]Department of Biochemistry and Biophysics and Research Center of Computational and Biotechnological Sciences (CRISCEB), Second University of Naples, via Costantinopoli 16, I-80138 Naples, Italy

[3]Laboratory of Bioinformatics and Computational Biology, Institute of Food Science, CNR, via Roma 52 A/C, I-83100 Avellino, Italy

Motivation

Recent studies have shown that the protein folding is widely influenced by the topology of the tertiary structure and have suggested that symmetry and geometry properties determine a limited spectrum of ground state conformations that a protein may assume as its native state [1,2]. On the basis of these considerations, we have developed a computational model taking into account only geometrical features of the proteins, which seems to provide a new way toward the protein fold prediction. The main steps along this way are: a) to use this model for generating a large set of C-alpha conformations potentially able to represent the traces of real mainly- alpha proteins, b) to build from this set full-atom structures corresponding to a given amino-acid sequence, and c) to evaluate them on the basis of structural and energetic criteria, in order to select few structures that may actually represent conformational states of the chosen sequence.

Methods

Our model has been obtained starting from the tube model proposed by Banavar et al. [3,4], which represents the protein backbone as a flexible tube whose axis passes through the C-alpha atoms. This tube has the maximum allowed thickness under many constraints, that account for steric effects, hydrogen bonds and hydrophobic interactions, thus preventing self intersections and enforcing compactness. We have modified the tube model to obtain a simpler protein representation, focusing on the simulation of mainly-alpha conformations. By analysing a set of mainly-alpha monomeric protein structures available in the PDBSelect database, we have identified few main geometric properties that these structures must satisfy, including the range of values that the thickness of a generic alpha-helix may assume. We have then described a mainly-alpha conformation as an ensemble of secondary protein structures, which maximizes the number of alpha-helices having such thickness. To get compactness we have also imposed that as many C-alpha atoms as possible are contained in an ellipsoid approximating the protein shape. This model has been formulated as a constrained global optimization problem and a Metropolis Monte Carlo Simulated Annealing has been applied to compute its solutions [5]. The simplicity of this model allows to generate a large set of conformations at a low computational cost. On the other hand, due to the few constraints imposed, this set may contain many non-physical traces, that must be identified and discarded. In our study we have applied the MaxSprout program to build full-atom structures from the generated traces and have evaluated these structures by checking their stereo-chemical quality and their suitability to resemble globular proteins in terms of H-bonds, voids, solvent-accessible surface area and water molecules in a layer of 5 Angstroms [6], and by using other methods for 3D model quality assessment.

Results

Our computational model has been applied to build about 10000 traces consisting of 65 C-alpha atoms. Full-atom structures have been generated from all the computational traces, by using MaxSprout and the sequences corresponding to four reference protein structures available in PDB, i.e. 2CRO, 1UTG, 1COKA and IJ8A. By applying the structural and globularity criteria four subsets of conformations have been selected, consisting of 896 models for 2CRO, 652 for 1UTG, 597 for 1COKA and 872 for 1JE8A. By using structural fitting to evaluate each subset against the corresponding reference template, simulated conformations with RMSD values lower than 7 have been found.

Further selection criteria have reduced the number of models to 13 for 2CRO, 11 for 1UTG, 6 for 1COKA and 11 for 1JE8A, which are the only simulated conformations satisfying both structural and energetic criteria. These results show that our approach is able to generate conformations that may represent native states of real mainly-alpha proteins.

Email: antonio.mucherino@unina2.it