Jordan University of Science and Technology

Improving Protein 3D Structure Prediction Accuracy using Dense Regions Areas of Secondary Structures in the Contact Map

Authors: Amer Al-Badarneh, Mohammad Khalil, and Mo'taz Al-Hami

Abstract: The protein folding problem is a fundamental problem in structural molecular biology. This problem describes how a protein is transformed from its primary sequence (i.e., amino acid sequence) into the three dimensional structure (3D structure) for this sequence that determines the function of the protein. The 3D structure of a protein can be represented using a square symmetrical binary matrix called contact map. The concept of contact map facilitates the transformation of the folding problem into a computational one, so various computational approaches use the contact map to predict protein secondary structures. Correlation mutation analysis is an approach that tries to study the mutated patterns that appear in the multiple sequence alignments, this approach predicts every pair of protein residues to be in contact or not independently of the other pairs. Approach: This study proposed an improvement over correlation mutation analysis to predict the secondary structures that exist in the contact map. The proposed method uses regions of the secondary structures instead of independent pairs as in the typical correlation mutation analysis; also it applies the analysis on the dense regions rather than the whole contact map. Results: The proposed method was implemented on proteins related to different classes (i.e., mainly alpha, mainly beta, mixed alpha beta and low secondary structures). The test proteins are extracted from the Protein Data Bank (PDB) of solved structures. The results show improvements of dense regions accuracy over correlation mutation accuracy and random accuracy. Conclusion: According to the amount of wrongly predicted contacts, the results show a large decrease in the wrongly predicted contacts in the dense regions analysis over correlation mutation analysis.