Abstract

Computational Protein Design (CPD) is an emerging technology in the field of pharmaceutical and biotechnology. Through this technique, proteins can be engineered to have novel folds and functionality with better thermo stability and affinity towards their interacting partners. Moreover, in the field of disease biology, it helps to detect the responsible residues causing the abnormality in the cellular processes. Given a target structure, CPD tries to find out one or more sequences that will fold to the target tertiary protein structure. A sequence search procedure and an energy function assessing the sequence-structure compatibility are two key components of a CPD program.

Although deterministic and non-deterministic algorithms are well explored to solve the CPD problem, there have been only a few attempts to utilize parallel algorithms to solve this problem. At first, we have utilized the independent modular structural organization of the protein structure in designing the parallel sequence search algorithm. A divide-and-conquer approach has been incorporated where the protein is split into protein units (PU) and each PU is explored in parallel. This shared memory implementation of modularity-based parallel sequence search leads to better search space exploration compared to the case of traditional full protein design. On analyzing the design sequences on the benchmark data set, it shows 39.7% sequence similarity with an average root-mean-square-deviation (RMSD) and average TM-score of 1.17Å and 0.89, respectively.

An efficient search space exploration technique in CPD helps to get the optimum solution in a considerable amount of time. Here, we propose a greedy simulated annealing-based Monte-Carlo parallel search algorithm for better sequence-structure compatibility probing in protein design. The guidance provided by the evolutionary profile, the greedy approach, and the cooling schedule adopted in the simulation ensures sufficient exploration and exploitation of the search space leading to faster convergence. On evaluating the proposed algorithm on the benchmark dataset, we report an average RMSD of 1.07Å and an average TM-score of 0.93 with the modeled designed protein sequences. A high (93.4%) intra-group recapitulation of hydrophobic residues in the buried region indicates that the proposed protein design algorithm preserves the core residues in the protein and provides alternative residue combinations in the solvent-accessible regions of the target protein.

Protein-protein interactions are critical for most of the cellular processes ranging from immune defense, intercellular signal transmission to growth and repair of cells. Protein interface engineering is one of the major applications of CPD which opens up numerous possibilities for enhancing or inhibiting such cellular processes where the protein complex structure serves as the starting point. In this work, an automated computational framework has been developed for modeling functionally related protein complex structures utilizing GO-based semantic similarity technique and co-evolutionary information of the interaction sites. The framework considers both protein sequence and structure information as input and employs rigid-body docking and template-based modeling together based on an integrated protein-complex structure prediction approach. Our framework combines geometric as well as physicochemical features for re-ranking the docking decoys. The proposed framework has an 83% success rate when tested on a benchmark dataset while considering Top1 models for template-based modeling and Top10 models for the docking pipeline.