

ABSTRACT

Specification of the three dimensional structure of a protein from its amino acid sequence, also called a "Grand Challenge" problem has eluded a solution for over six decades. A modestly successful strategy has evolved over the last couple of decades based on the premise that an ensemble of decoys could be generated as plausible candidates for the native and a scoring function which mimics free energy could then capture the native or native-like structure. The scoring function must discriminate the native from unfolded/misfolded structures, must be fast enough and must have been validated on a large dataset to generate sufficient confidence in the score.

Addressing these issues, we report here a method to detect true native structure in the top 5 to 93% accuracy from an ensemble of candidate structures. If we eliminate the native from ensemble of decoys then pcSM is able to capture near native structure (RMSD $\leq 5\text{\AA}$) in top 10 with 86% accuracy. The parameters considered in pcSM are C-alpha euclidean metric, secondary structural propensity, surface areas and an intramolecular energy function. pcSM has been tested on 415 systems consisting 142698 decoys (public and CASP - largest reported hitherto in literature). The average rank for the native is 2.38, a significant improvement over that existing in literature. In-silico protein structure prediction requires robust scoring technique. Therefore pcSM is easily amenable to integration into a successful protein structure prediction strategy. The tool is freely available at <http://www.scfbio-iitd.res.in/software/pcsm.jsp>.