Evaluation of the best Features of a Fragment Library for Fragment-based Method for Protein Structure Prediction

Trevizani, R.¹; Custódio, F.L.¹; Dardenne, L.E.¹

¹Grupo de Modelagem Molecular de Sistemas Biológicos, LNCC, RJ, Brazil

Fragment-based protein structure prediction methods allow to import empirical data into the model without the dependency homologous structures. This study aims to assess the ability of distinct fragment libraries to correctly reproduce native structures. We developed a program to generate fragment libraries from a non-redundant set of non-homologous PDB structures. For each aminoacid of the target sequence a list of 200 fragments ranked according to its likelihood (given by sequence and secondary structure similarities) to reproduce the local structure. The ProFragger software was used to create fragments of 9, 6 and 3 residues long. Additionally, a mixed library with fragments of all three sizes were used in the attempt to rebuild the native structure. Separately, fragments were clustered according to structural similarity allowing a single fragment to represent a larger group. Random positions were sampled throughout the query sequence, as the computational cost of verifying every combination makes an exhaustive approach prohibitive. Fragments that better fit local and global DME criteria are inserted in their respective position. The mixed libraries consistently yielded better models as they benefit from the advantages of larger and smaller fragments: the former reproduces secondary structures more accurately and are used in the early stages of the algorithm while the latter improves loops and coils regions later in the simulation. Clustering the libraries by structural similarity had no substantial negative effects on the quality of the model and might be an interesting feature for Ab initio methods, as it reduces search space.

Word Keys: Ab initio prediction, Fragment library, Protein structure, Supported by: CAPES