

"Development of data mining tools for identifying structural determinants that dictate protein-ligand interactions".

Abstract

Modelling binding sites of enzymes is a fundamental but rather demanding task, of increased complexity since the residues forming these sites are not rigid. Similarly, binding studies of a ligand at such a site and complex formation raises difficulties mainly because most of the structural determinants that control binding are not known. Using a combination of sampling algorithms and statistical analysis techniques, we shall contribute towards developing much more accurate binding affinity predictions for macro-molecular docking. To this end, we study benchmark protein families with known 3D structure with the aim to identify specific geometric parameters for modeling their binding cavities. This will be possible by studying the boundaries within which every residue in those cavities can move, in 3D Euclidean or conformational space. Key methods employed include structural alignment of secondary structure elements, RMSD heat-maps, sampling (e.g. in the space of rotamers), standard scoring functions, and (generously) allowed regions as defined in Ramachandran plot. Our algorithmic tools involve powerful methods, such as nearest-neighbor search and clustering, which shall be adapted to the specific context. The developed methods were tested on a subset of protein kinases with known 3D structure, which offer a number of target sites for one or several ligands and phosphorylases, which provides a single target for many ligands. Different conformers were first produced based on the simulation of chi angles rotations and then clustered in a 2-level hierarchical process. For each conformer cluster, representative polygonal shapes were produced which can be thereafter exploited in ligand screening approaches. To further aid protein analysis a series of bioinformatics tools were developed and their usage is discussed.

Keywords: Conformers, hierarchical clustering, structural determinants, conformation space, data mining