Binding sites identification in protein structures using ligand-aspecific structural motifs

Bianchi V[(1)], Ausiello G[(1)], Gherardini PF[(1)], Helmer-Citterich M[(1)]

[(1)]Department of Biology, University of Tor Vergata, Roma

Motivation

Identifying the location of the functional sites in protein structures makes it possible to increase the efficiency of applications such as molecular docking, de novo drug design, and structural identification and comparison of functional sites. The aim of functional site prediction is to detect functional sites on uncharacterized proteins. Actually only few methods seem to be able to successfully predict binding sites in unbound proteins.

Methods

Here we describe a new method for ligand binding sites identification in protein structures. The method assigns a propensity value to each aminoacid in a protein structure based on the local similarity of the residue and its neighbours with portions of other binding pockets of known structure (independently of the type of bound ligand). In order to calculate this value we created a non-redundant dataset of all the protein-ligand binding pockets present in the PDB and a complementary nr-dataset of all residues not belonging to any binding pocket. The structure to be analyzed is then used as a query in a search for small structural motifs (3 residues) using the Query3D local comparison program in both datasets. The propensity value of each aminoacid is calculated as the ratio between its occurrence in structural motifs matching the binding dataset vs. the non-binding one.

Results

We tested the rationale of the method by trying to discriminate between solvent exposed regions of protein structures and buried ones. To do so we substituted the 2 datasets with datasets of solvent accessible and non-accessible residues made from 1091 nr PDB chains. The method permitted us to discriminate between residues in these two sets with an AUC of 83.4 ± 5.2 %. This proves that the geometric information (in the form of the type of aminoacids packing) can be used to distinguish between residues buried in the core or exposed on the surface of a protein structure. To test if this information can also be used to discriminate between residues belonging to binding pockets and other non-binding surface residues, we are now applying the method to the LigAsite dataset of protein binding pockets in the holo and apo forms.

Contact : gabriele.ausiello@uniroma2.it