PDBBinder: a New Method for Binding Site Prediction in Protein Structures

Bianchi V, Gherardini P F, Helmer-Citterich M, Ausiello G
Department of Biology, University of Tor Vergata, Roma

Motivation
The identification of ligand-binding sites in protein structures is a key task in the annotation of proteins with known structure but uncharacterized function. Computational methods for the detection and characterization of functional sites on protein structures have increasingly become an area of interest (Gherardini and Helmer-Citterich, 2008), largely due to the many newly solved structures that have poorly characterized biochemical functions or molecular interactions. Faced with a rapidly increasing number of known protein structures, it has become more important to have analytical tools that identify functional sites. In addition, functional site detection is important for targeting specific sites in structure-based drug design.

Methods
Here we describe a knowledge-based method exploiting the observation that unrelated binding sites share small structural motifs that bind the same chemical fragments irrespective of the nature of the ligand molecule as a whole. We compare a query protein structure against a library of binding and non-binding protein surface regions derived from the whole PDB (Berman, et al., 2000) using Superpose3D (Ausiello, et al., 2005; Gherardini, et al., 2010), a local structural comparison method. The number of similarities identified in the two sets is then used to derive a propensity value for each residue in the protein of interest. Surface areas with high propensity values denote the position of the predicted binding site.

Results
PDBinder was trained on a non-redundant set of 1356 high-quality structures of protein-ligand complexes derived from the whole PDB, and tested on LigASite (Dessailly, et al., 2008), a dataset comprising 267 apo/holo structure pairs. The results achieved by this method are evaluated in terms of Matthew’s Correlation Coefficient (MCC), Sensitivity, Specificity and Positive Predictive Value (PPV). We obtained an average MCC value of 0.33 with an average sensitivity of 0.31, an average specificity of 0.98 and a PPV of 0.44 using the Holo Test set while on the Apo Test set of unbound structures the average MCC values calculated from the distribution was 0.29 with a sensitivity of 0.27, a specificity of 0.98 and a PPV of 0.40. PDBinder is able to assigns a residue to the correct class, binding or non-binding one, on average in the 77% of the cases both for apo and holo protein structures. These results show that geometry alone gives a strong
contribution to the predictive power of PDBinder. Moreover, this method suffers only a modest decrease when moving from holo to apo structures, indicating that it is suitable for real-world applications where the location of binding site is unknown.

Contact email
valerio.bianchi@uniroma2.it