Identification of phosphate binding sites in protein structures

Parca L, Ausiello G,Gherardini PF, Helmer-Citterich M

Department of Biology, University of Tor Vergata, Roma

Motivation

The most represented and fundamental protein ligands involved in biochemical processes, like nucleotides and other cofactors, contain one or more phosphate groups. Consequently, it is important to be able to predict whether a protein is capable of interacting with this molecule. Many methods have been developed to predict metal binding sites in protein structures but no one is available so far for the identification of phosphate binding sites. Here we present a new method whose aim is to predict phosphate binding sites, either as a single ion or as a functional group, in an unbound protein structure.

Methods

The method we developed is based on the identification of small sets of aminoacids, in a 3D conformation that allows them to interact with a phosphate ion. For this purpose Query3D, a local structural comparison program, is used to search in the protein structure to be analyzed for groups of residues (structural motifs) similar to known conformations of aminoacids able to bind the phosphate chemical group. The template motifs have been obtained from a manual selection and analysis of binding motifs having two specific features: being present in at least 2 different SCOP folds and interacting with a phosphate functional group (also in the context of larger ligands).

Results

The method has been tested on a dataset composed of 55 high quality protein-nucleotide complexes. Each test structure has been compared with a list of 232 phosphate binding motifs. Motifs extracted from structures homologous (sequence identity higher than 30%) to the test structure have not being used. A prediction is considered correct whenever a phosphate is placed within a 1.7 Å radius from the original position in the complex. Structural matches with residues having a total relative accessibility lower than 25% have been excluded. 42 out of the 55 complexes obtained at least one correct solution among those proposed. The PPV (Positive Predictive Value) of the method is actually 8.76% but can be consistently increased by considering only phosphate positions that are predicted outside the solvent accessible surface of the proteins.

Contact : gabriele.ausiello@uniroma2.it