Identification and analysis of conserved pockets on protein surfaces

M. Cammisa[1] ✉, A. Correra[1], T. Fioriello[1], G. Andreotti[1], M.V. Cubellis[2]

[1]Istituto di Chimica Biomolecolare - CNR, Pozzuoli, Italy
[2]Istituto di Biostrutture e Bioimmagini - CNR, Napoli, Italy

Motivations

The interaction between proteins and ligands occurs at pockets which are often lined by conserved aminoacids. In order to make the research of new drugs as economic as possible, it is necessary to exploit "in silico" techniques, high throughput and fragment-based screenings, which require the identification of pockets on the surface of proteins, active sites or not, which might be the targets of low molecular weight drugs.

Methods

We developed a tool to evaluate the conservation of each pocket detected on the protein surface by CastP. This tool was named DrosteP because it recursively searches for optimal input sequences to be used to calculate conservation. DrosteP uses a descriptor of statistical significance, Poisson p-value, as a target to optimize the choice of input sequences. To benchmark DrosteP we used monomeric or homodimer human proteins with known 3D-structure whose active site had been annotated in UNIPROT. DrosteP is able to detect the active site with high accuracy because it nearly always coincides with the most conserved pocket. We extended our analysis to all the pockets found on the surface of human proteins.

Results

Several methods for predicting ligand binding sites on protein surfaces have been proposed which combine 3D-structure and evolutionary sequence conservation, but any method relying on conservation depends critically on the choice of the input sequences. DrosteP chooses how deeply distant homologs must be collected to evaluate conservation and thus optimize the identification of active site pockets. Moreover it recognize conserved pockets other than those coinciding with the sites annotated in UNIPROT which might represent useful druggable sites. Amino acid composition of conserved pockets differs significantly from that of non conserved pockets. This finding provides useful hints on the fundamental principles underlying protein-ligand interaction.

Availability

http://www.sbcentrostorico.unina.it/cammisa