Success of sequencing projects and major advances in protein structure determination have led the structural biology community to undertake the large scale mapping of protein structure space. Since structural genomics projects aim at high-throughput delivery of protein structures regardless of the state of their functional annotation, bioinformatics support is required to provide tools to structure-based prediction of molecular function. We propose to contribute to this task by detecting and annotating ligand binding sites.

Using representative proteins, we automatically generate 3D motifs of binding sites of interest. This is achieved using the Nestor3D software which creates 3D motifs from consensus atom positions. Each motif is also associated to a CASTp-based cavity descriptor.

The identification of ligand binding sites is achieved by, first, detecting in proteins putative cavities using CASTp. Using cavity descriptors, protein cavities can then be associated to potential motifs. Finally, a matching process searches for the relevant ligand 3D motifs within cavity surfaces, which are also represented by atoms. Our methodology is evaluated on a set of holo and apo flavoprotein structures, which demonstrates its potential for binding site detection and annotation.