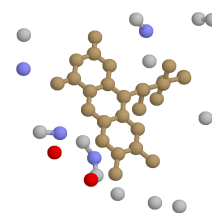


Detection and annotation of ligand binding sites based on atomic descriptors



Jean-Christophe Nebel

Faculty of Computing, Information Systems & Mathematics, Kingston University,
Kingston-Upon-Thames, KT1 2EE, UK

j.nebel@kingston.ac.uk

Abstract

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. Methodology is evaluated on a set of holo and apo flavoprotein structures, which demonstrates its potential for binding site detection and annotation.

Principles

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.

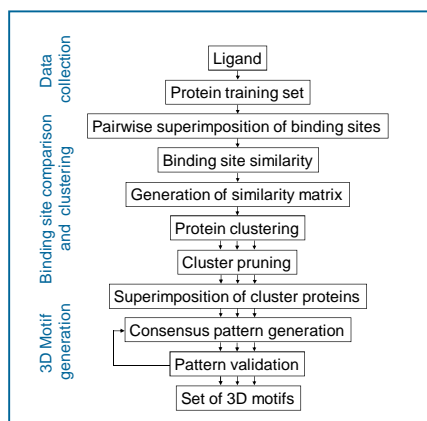
We propose to contribute to this task by detecting and annotating ligand binding sites.

This is achieved by:

- Generation of 3D motifs of protein structure binding sites based on consensus atom positions [1].
- Detection of 3D motifs from cavity surface atom representation as provided by CASTp [2].

Automatics generation of 3D motifs*

General pipeline



Data collection

- For each ligand of interest, a representative set of protein structures with bound ligand is extracted from the PDB [3].

Binding site comparison and clustering

- Nestor3D software performs pairwise superimposition of binding sites [4]: scores allow production of a similarity matrix.
- Similarity matrix is used to produce clusters of similar binding sites.

3D Motif generation

- For each cluster, 3D motif is produced based on consensus atom positions.

*More details in [1].

Detection of 3D motif in cavity and ligand position prediction

3D Motif descriptor

- 3D motifs are produced from atoms belonging to relevant cavity surfaces as provided by CASTp.
- 3D motifs are associated to a CASTp-based cavity descriptor: average molecular surface and area of the cavities of site representatives.

Pre-processing

- Protein structures of interest are processed using CASTp.
- Cavities compatible with cavity descriptors are selected for 3D motif detection.

3D Motif detection

- For cavity of interest, create all possible triplets of atoms composed of exactly 1 Nitrogen, 1 Oxygen and 1 Carbon atoms.
- Similarly, create all atom triplets in 3D motif.
- Find matching triplets.
- For each matching triplet, Nestor3D software scores superimposition of 3D motif with cavity.

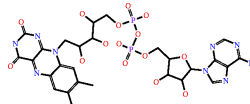
Ligand position prediction

- If superimposition score is below threshold, ligand is positioned in cavity according to its relative position in the detected 3D motif.
- Ligand position is accepted if no collision between ligand and cavity is detected.

Application to flavoproteins

Dataset

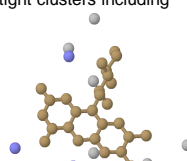
- Ligand of interest: Flavin-adenine dinucleotide (FAD).



- Representative set of protein structures with bound FAD: 122 structures in PDB30%.

Generated 3D motifs

- 8 motifs produced from 8 tight clusters including 86 structures.
- Example: 11-atom motif generated from 5 representatives of a 12 structure cluster

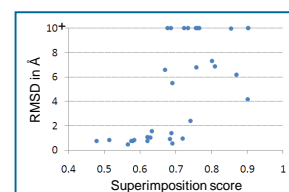


(here, rigid part of FAD is displayed in 3D motif)

Some results

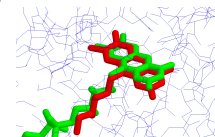
Holoproteins

- Accuracy of ligand location prediction:



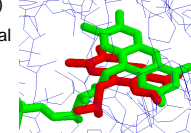
- Good correlation between score and accuracy.
- Ligand location prediction for 3dme:

RMSD 0.74Å



Apo protein 2oam (Biosynthetic protein/flavoprotein)

- 2oam is the apo form of 2oal (RMSD: 0.19Å):
- 3D motif is detected with a score of 0.63.
- No collision between FAD and 2oam's cavity. (closest atom: 2.20Å)
- RMSD between 2oal ligand position and prediction in 2oam: 1.74Å



Future work

Investigation of other flavin ligands (e.g. FMN)

Use of 'thicker' cavity surfaces

- currently only 50% of atoms present in 3D motifs are present in motifs generated from cavity data.

Refine ligand location prediction using force field

References

- J.-C. Nebel et al. (2007) Automatic generation of 3D motifs for classification of protein binding sites, *BMC Bioinformatics*, 8:32
- J. Dundas et al. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues, *Nucl. Acids Res*, 34:W116-118
- H. Berman et al. (2000) The Protein Data Bank, *Nucleic Acids Research*, 28, 235-242
- J.-C. Nebel (2006) Generation of 3D templates of active sites of proteins with rigid prosthetic groups, *Bioinformatics*, 22(10): 1183-1189