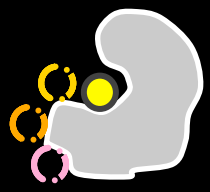
In Search of Ligand-Binding Regions in Proteins using Voxel Clusters with Applied Approximated Measures

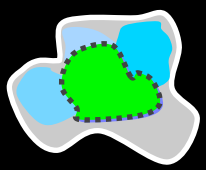
Andrzej Bargiela, Ling-Wei Lee
The University of Nottingham

INTRODUCTION

- Proteins – combinations of amino acids.
- Many factors contribute to binding sites, including but not limited to hydrophobicity, electronegativity, chemical composition, and shape complementarity.
- The behaviour of proteins in binding to ligands can be attributed to atomic arrangements on surfaces.



Probe-based Methods



Surface Patches/Hot Spots

BACKGROUND

- Geometrical criteria have been the focus of many dock site detection algortihms.

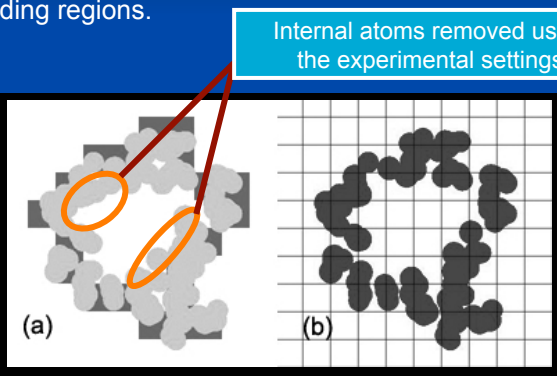
Examples of Dock Site Detection Programs	
PASS	PASS (Putative Active Sites with Spheres) locates dock sites by checking for the 'burial count' of atoms placed on the surface. Each potential region is represented by the center of the site and is known as an Active Site Point (ASP).
SURFNET	SURFNET fits virtual spheres into accessible spaces. Each sphere has two opposing atoms on its surface.
CAST	CAST computes protein surfaces using alpha shapes derived from Voronoi diagrams and Delauney triangulations. Smaller triangles are grouped into neighbouring larger ones known as sinks.

THE ALGORITHM – PREPARATION OF FILES

- The algorithm is demonstrated on a selected group of 32 bound/ unbound structures (a subset of the dataset used in PocketPicker and LIGSITE^{csc}).
- Only files of <300kb in size were selected – however this is still a good sampling set. Residues count range from 64 (smallest in dataset) to 501 (largest in dataset).
- Files were downloaded from the RCSB PDB in PDB format.
- Each file undergoes pre-processing for extraction of spatial coordinates and atom types.
- Each protein is contained within a large enough cubic grid space. The space is tessellated into equal units of 4.0 Å. Therefore the protein is represented in terms of voxel units.

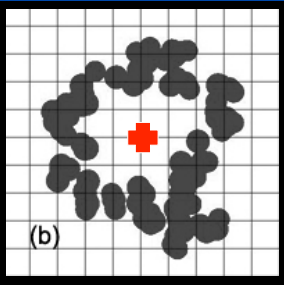
THE ALGORITHM – PREPROCESSING THE PROTEIN

- Surface atoms are first extracted from the protein using a series of experimental settings, i.e. voxel occupancy and degree of belonging of atoms to voxels.
- The list of surface atoms is then used as input data in the search for ligand-binding regions.



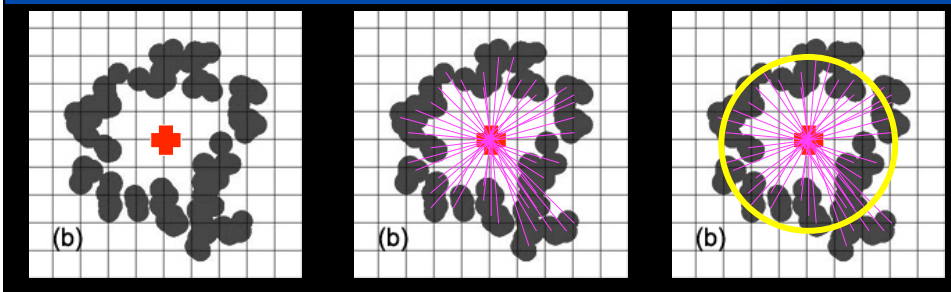
THE ALGORITHM – IDENTIFYING POTENTIAL BINDING SITES

- Due to the voxel-based representation, properties such as voxels count and cluster size may be used to represent the protein.
- Potential dock sites are identified based on the depth attribute.
- Firstly the averaged center of the protein, C_p, is calculated using the spatial values of all surface voxels.



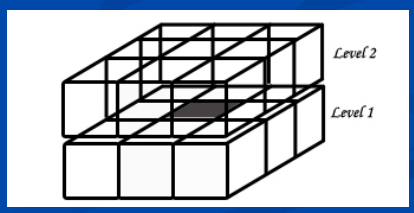
THE ALGORITHM – IDENTIFYING POTENTIAL BINDING SITES

- Based on this average point C_P, the distances of all atoms to Cp are obtained followed by averaging the sum of all the distances.
- This average distance, D_A – is used to create a 'bubble' around the protein. The 'bubble' divides all extracted surface atoms into two regions – the outer group and the inner group. The former is associated with protruding regions while the latter with crevices.



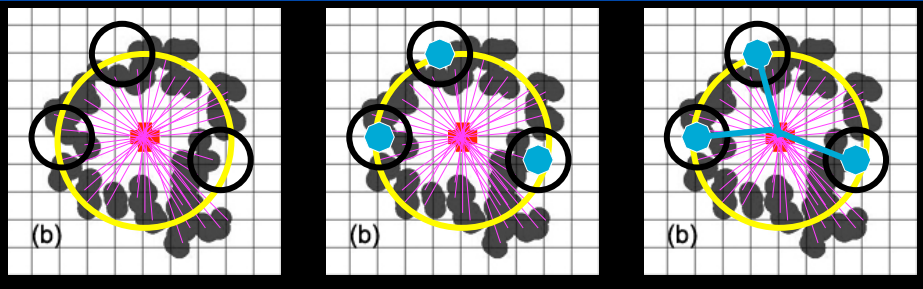
THE ALGORITHM – IDENTIFYING POTENTIAL BINDING SITES

- Only atoms of the inner group are considered in the search for potential dock sites.
- Binding sites are determined based on the depth attribute and the criteria that they have to be sufficiently large.
- All surface voxels are first sorted from the units closest to C_P to the furthest.
- The list is then iterated and a cluster constructed about each voxel.
- A two-level voxels boundary about each iterated unit is created which includes a counter for the second level of the boundary.



THE ALGORITHM – IDENTIFYING POTENTIAL BINDING SITES

- The cluster is checked for acceptibility by the following criteria :-
  I. Each cluster must contain 5 or more voxels.
  II. The count of voxels in the second level of boundary must not be zero.
- The average points of all the clusters, C_A, are calculated.
- The cluster returning the largest value based on 1/(distance between C_A and C_P) is then selected.



RESULTS

- Comparisons are made between the output from the implemented method to those of LIGSITE^{csc} and PocketPicker.
- These benchmarks are chosen as they have proven to return good results.
- Evaluation for the voxel-based method is carried out on a visual basis. All identified residues are compared against projections from the RCSB PDB JMol Viewer.
- An identification is accepted if >5 matching residues are found.
- As the implementation returns only the best matches, therefore only the best hits for other methods (values of '1') are used for comparisons.

RESULTS

Bound	Unbound	Voxel	Pocket-Picker	LIGSITE^{csc}
1BID	3TMS	1	1	1
1DWD	1HXF	1	1	1
1HEW	1HEL	1	1	1
1HYT	1NPC	1	1	1
1INC	1ESA	1	(1)	1
1RBP	1BRQ	1	1	1
1ROB	8RAT	1	1	1
1ULB	1ULA	1	(1)	0
2IFB	1IFB	1	1	1
3PTB	3PTN	0	3	1/2
4PHV	3PHV	1	2	1
1A6W	1A6U	0	3	1/3
1APU	3APP	1	1	0
1HFC	1CGE	1	1	1
1IDA	1HSI	1	1	1
1MRG	1AHC	1	(1)	1
1MTW	2TGA	1	4	1/5
1OKM	4CA2	0	1	1
1PSO	1PSN	1	1	1
1QPE	3LCK	0	1	2
1RNE	1BBS	0	1	1
1SNC	1STN	1	1	1
1SRF	1PTS	0	1	1
2CTC	2CTB	1	1	1
2H4N	2CBA	1	1	1/2
2PK4	1KRN	1	1	1/2
2SIM	2SIL	0	2	1/2
2TMN	1L3F	1	1	0
3GCH	1CHG	1	2	10
3MTH	6INS	1	2	9
5P2P	3P2P	0	1	1
6RSA	7RAT	1	1	1/4
Percentage (%)		75.00	68.75	59.38

Notes

- All numbers indicate rankings.
- Numbers within brackets indicate identifications located out of defined radius, i.e. 8 Å.
- Fractional numbers for LIGSITEcsc indicate re-rankings from a previous resultset. These are not included in the comparisons due to the re-rankings carried out.

SUMMARY

- Although voxels and grid spaces may impose a degree of rigidity and inflexibility, however with proper rules and measures they can be used as tools for the study of subjects.
- In this study the method of locating possible ligand-binding sites is an approximated approach but has proven to successfully locate the regions of interest.
- It has to be noted that due to the size of the unit voxels used each voxel may carry one or more atoms from any residue. It is highly possible for neighbouring residues with little or no contribution to interactions to be extracted when their atoms is found contained in the two-level voxels boundary.
- The study affirmed that most ligand-binding sites are found in crevices. In the event of the failed cases the sites were either located on comparably flat regions or in crevices insufficiently deep.

THE END

THANK YOU

QUESTIONS?