Journal of Physical Science and Application 2 (10) (2012) 399-408

An Approximated Voxel Approach for the Identification and Modelling of Ligand-Binding Sites

Ling Wei Lee[1] and Andrzej Bargiela[2]

1. School of Computer Science, The University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

2. School of Computer Science, The University of Nottingham Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, UK and Institute of Informatics, Cracow Technical University, Poland

Received: September 14, 2012 / Accepted: September 26, 2012 / Published: October 15, 2012.

Abstract: Most protein-ligand interactions take place on surfaces and include but not limited to factors such as chemical composition, hydrophobicity, electronegavitiy and shape complementarity. Past studies showed that protein-protein interactions occur on comparatively flat regions whereas protein-ligand bindings involve crevices. In the search for such sites various approaches have been designed and developed each of which is algorithmically unique. The use of grid units or voxels has been demonstrated in early studies with relatively good results obtained. We present here an approximated approach comprising of the use of voxels and computer vision methods in the search for ligand-binding areas. Each test protein is modelled and analysed in 2D with all corresponding residues graphically presented for successfully identified sites. The study was carried out on 2 sets of proteins: FK506-bound proteins and heme-bound proteins with promising results obtained for all test cases.

Key words: Binding sites identification, ligand-binding, voxel space, voxelisation, grid units, protein surface atoms.

1. Introduction

Proteins constitute of combinations of amino acids which collectively define all associated functions. The process of binding to external agents or ligands usually take place on surfaces and each site can be defined by a host of factors which contribute to its reactivity including (but not limited to) hydrophobicity, electronegativity, chemical composition, shape complementarity, etc [1-6].

Atomic arrangement is one of the key factors contributing to binding site characteristics. A bind often results in the activation or inhibition of certain atoms. In comparison to sequential or structural studies, surface-based analyses return more concise protein surface information. A publication by Via, et al. [7] stated that “protein surface comparison is a hard computational challenge and evaluated methods allowing the comparison of protein surfaces are difficult to find”. This affirms the difficulties associated with protein surface studies. There are cases in which selected proteins bind to the same ligand therefore suggesting the possibility of conserved features within the dock sites. Such scenario indicates that the proteins may have descended from the same ancestors. On the contrary it is also likely that proteins of different origins undergo mutations resulting in the proteins bearing similar features [8].

Many methods are available today for the investigation of protein surface properties. The methods were designed using a diverse selection of algorithms ranging from simple mathematical models to energy functions and probes equipped with chemical information. Early programs such as POCKET [9] and LIGSITE [10] employed the use of grid environments and experimental spheres for the

Corresponding author: Ling Wei Lee, Ph.D. candidate, research fields: multi-resolution modelling, systems biology, image and vision processing. E-mail: leelingwei@yahoo.co.uk.

detection of pockets on surfaces. Both the methods constructed 3-dimensional grid spaces for the study of proteins, experimental spheres of specified radii are used to examine the surfaces. The former includes 3-directional scannings to locate potential sites but it was found that orientation issues induced through the use of grid units have not been resolved. To reduce the severity of this problem LIGSITE increased the directional scanning to 7 with the additional directions being the 4 diagonals. The additional scans were shown to improve the performance of the algorithm, and the program is computationally fast.

Another concept associated to binding sites is that of surface patches and hot spots which were introduced in several works. Jones, et al. [11] proposed surface patches for the detection of interaction sites on a protein. Each patch is defined by a series of parameters including but not limited to solvation potential, hydrophobicity, planarity, accessible surface area with rankings carried out based on these values. 'Hot spots' were proposed by Bogan, et al. [12] which correlate as well to regions of interest. It was found that binding energy does not distribute evenly across the surfaces of proteins but is highly concentrated on dock sites. An example of work focusing on hot spots is the optimal docking area (ODA) program [13] which aims to identify patches through experiments of different atomic solvation parameters. Chakrabati, et al. [14] discovered that larger interfaces are generally made up of multiple patches with at least a pair of patches equivalent in size to a single patch interface.

Good understanding of the factors contributs to a binding site aids in the development of methods for successful detection of these regions. Shape complementarity is prioritised in grid-based implementations. In a binding process, a crevice has to be sufficiently large (and/or deep) to accommodate a ligand. An approximated method is therefore presented here which comprises of a combination of computer-vision techniques and voxel-based environments for the identification and modelling of potential binding sites. All associated atoms and their corresponding residues are extracted in the process.

The surface of a protein is uneven and consists of both concave and convex areas. As demonstrated by POCKET and LIGSITE, grid-spaces offer a fast and comparably robust solution to solving protein-related challenges. In this study, a cubic grid-space large enough to contain the entire protein is first constructed. This experimental space is then tessellated into smaller units with the smallest unit having a size of 4.0 Å. Justification for the selection of this value is given in the next section. All data sources for the test proteins are obtained from the RCSB Protein Data Bank and downloaded in PDB format. The required information are then extracted from the files including the spatial coordinates of the atoms, the residual information, atom element and type, with the van der Waals radii of the atoms separately introduced. These data are compiled into a new file which will then be used as input to the algorithm.

Two groups of proteins have been selected for the study. The first set consists of FK506-binding proteins [PDB: 1FKF, 1BKF, 1YAT, 3VAW, 3UF8, 1C9H] which are molecules having a single active binding site each for one substrate. The protein 1FKF has been studied in two other experiments, the first being a wet lab approach [15] and the second a geometrical-based search integrated with geometric hashing [16]. With proven results this protein serves as a good benchmark subject for comparison of the obtained results. The second set consists of heme-bound proteins [PDB: 4MBN, 4HHB, 4FWX, 3TGA, 4DY9, 1B7V]. Each file first undergoes pre-processing followed by processes for identification of binding sites. The results are graphically presented with comparisons carried out between the obtained output against screenshots of the proteins as projected in the Jmol Viewer from RCSB PDB.

2. Experiments

Each file first undergoes pre-processing for compilation of all required data for the study. 6 fields are collected from the RCSB downloaded files, namely the X, Y and Z coordinates, the atom element type, the atom number and the residue name. All van der Waals radii are then separately introduced. Each atom entry for a chosen protein is represented by the above information.

An input protein is first projected into the 3D grid environment followed by a tessellation of the experimental space. The tessellation terminates when the smallest unit voxel has been achieved, i.e., 4.0 Å in this study. This value was selected based on analysis of protein atoms sizes. It was found that most protein atoms range between ~1.0 Å to 2.0 Å in radius (diameter of ~2.0 Å to 4.0 Å), therefore a unit size of 4.0 Å is sufficient to accommodate all smaller atoms as well as all larger ones. Due to the use of grid units, each protein may be represented as a cluster of voxels. Information such as rough estimates of the volume of the protein, total surface area and descriptions of overall size may be obtained as well. This cluster is 3-dimensional in nature and to process the 3D space induces much complexity. As such steps are taken to reduce the dimensionality of the state by converting the 3D space into a series of 2-dimensional images. The concept is very much similar to the Z-buffer algorithm in computer graphics. More details of the implementation can be found in a previous publication [17]. Fig. 1 shows a graphical representation of the conversion algorithm. An axis is selected for 'slicing' wherein an image is generated for every 4.0 Å interval.

Once all 2D images have been generated from the conversion process, the next step is to process each image using simple image processing techniques to identify protein-related voxels. By default a voxel is selected when a single pixel belonging to the protein is detected. This proved to be infeasible at later stages of the study when all surface atoms are required to be extracted. Fig. 2 shows such a scenario wherein the surface voxels have been highlighted and the circled areas show regions where surface atoms have not been shortlisted. As such the first constraint was introduced in the form of a voxel occupancy value. Rigorous experiments have been carried out to determine the optimum value for this constraint and it was found that a value of 40%-100% occupancy for each voxel returns the highest coverage of atoms in the lowest possible number of voxels [18]. Fig. 3 shows a graph displaying the count of commonly extracted surface atoms versus applied voxel occupancies. A total of 13 different orientations of the same protein were used with the common atoms selected if they were found to exist in the extractions of all 13 cases. The graph follows a Gaussian distribution curve with the highest extractions occurring between 40% and 50%. Fig. 4 shows a comparison of output from non-application of the constraint against output with the constraint applied. It is clearly shown that the constraint gives better definition to the exterior region of the protein [19, 20].

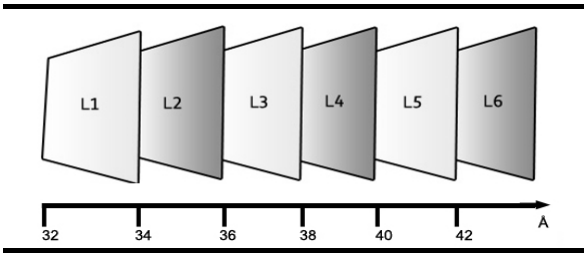


Fig. 1 Illustration of the 'slicing' process.
An image layer is produced for every 4.0 Å interval along the selected axis. L1 consists of atoms between coordinates 34 and 36, L2 of coordinates 36 to 38 and so forth.

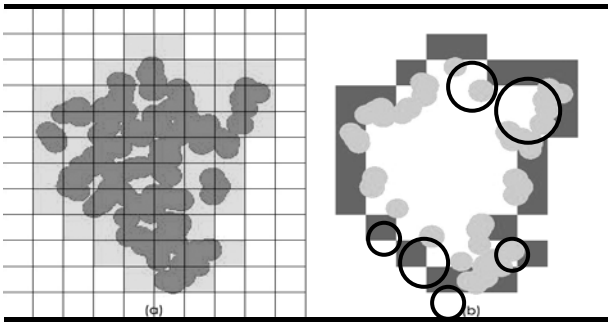


Fig. 2 (a) Highlighting of all voxels containing atoms of the protein with no constraint applied. (b) Selection of only the surface voxels and their associated atoms.
The circled areas show regions where surface atoms have not been extracted.

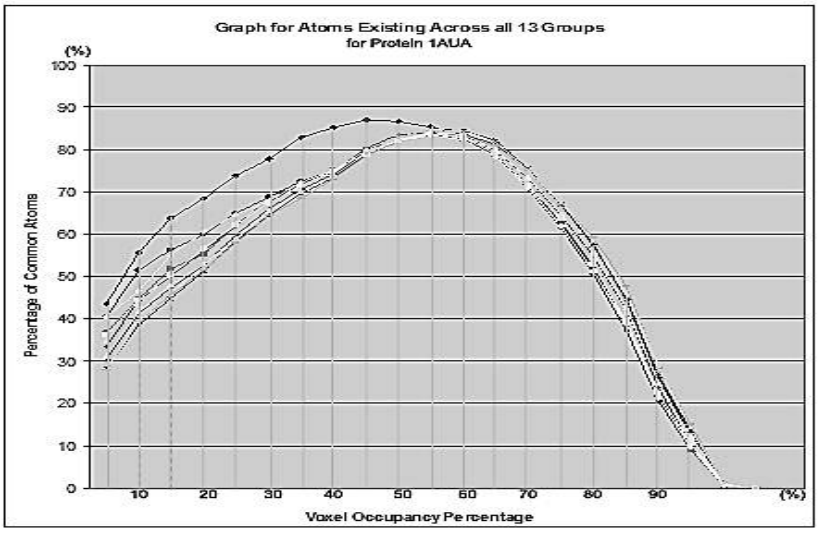


Fig. 3 The percentage of common atoms extracted against the voxel occupancy percentage applied in a set of 13 different orientation test cases of the same protein.

The highest count of common extractions occur between 40% to 50% of voxel occupancy. Therefore it is justifiably suitable to maintain a percentage of 40%-100% as the value of the constraint.

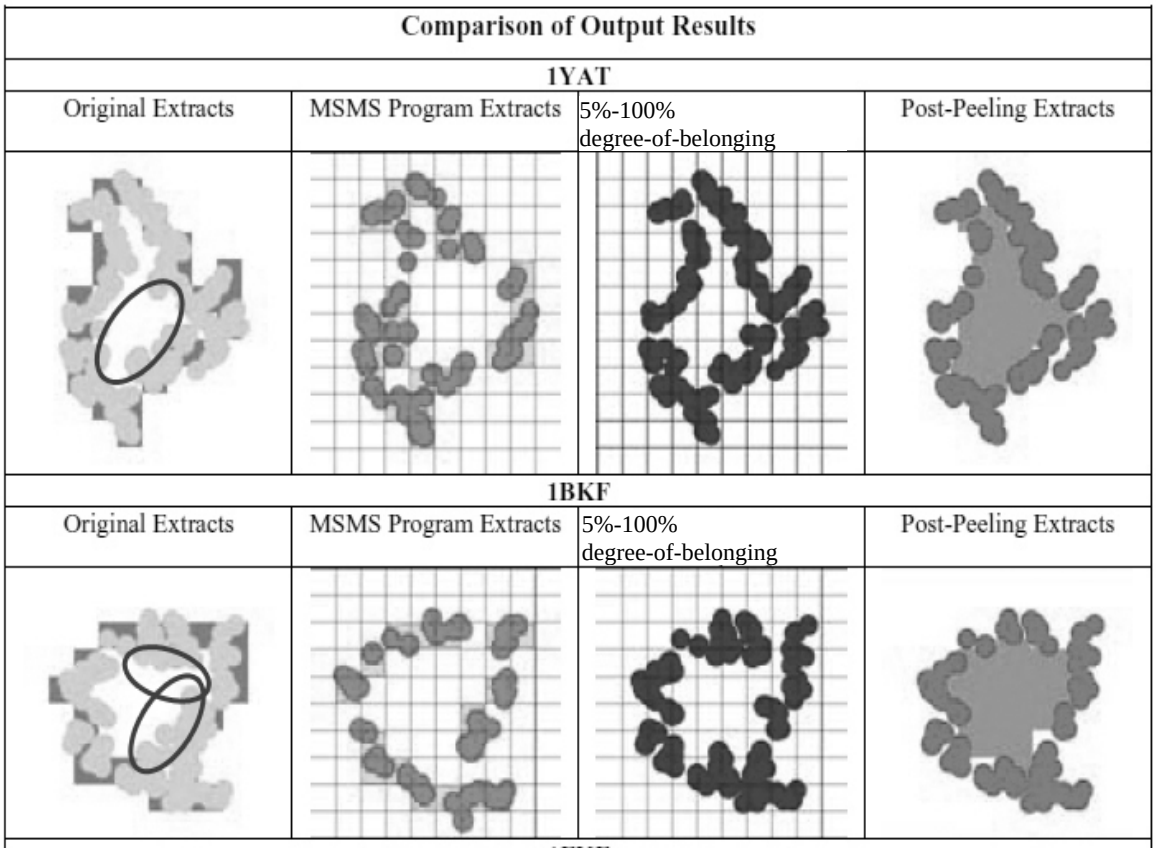


Fig. 4 Comparison of output obtained from different stages of the implementation and results from the MSMS program.

The first column shows the initial extracts collected based on associations with the surface voxels using a voxel occupancy constraint of 40%-100%. The second column gives results generated from the MSMS program (with probe radius of 2.0 Å). In the third column the extracts have been applied with a 5%-100% degree-of-belonging value whereas in the final column internal atoms have been 'peeled' off from within the protein.

The first column of Fig. 4 shows only the surface voxels highlighted alongside all associated atoms with the 40%-100% voxel occupancy constraint applied. Observation shows that some internal atoms have been included in the extracts and as such, steps were introduced to reduce the interference of internal entries thus 'cleaning up' the extracted list. The second constraint in this study is therefore introduced at this stage. Termed as degree-of-belonging of an atom to a voxel, the constraint attempts to filter out entries with minimal associations to the surface voxels. However analyses of the filtered output showed that a small number of internal atoms still remain within the group. The final step of enhancement is the use of a 'peeling' method to check for the exposure of each atom to the external environment. Any atom with low external exposure is consequently 'peeled' off therefore leaving only atoms with high probabilities of participating in externally induced interactions [19, 20]. The last two columns of Fig. 4 show sample output from application of the constraints to the original list of atoms.

After the process of removing internal atoms, the next stage is to proceed with the identification of binding sites on surfaces. As each protein is enclosed within a grid space, one can only obtain 6 views of the protein due to the characteristics of voxels. In a visual sense, crevices of an object may be discerned through a combination of factors of depth, clarity and intensity. In this study attempts are taken to identify dock sites based on the depth attribute. Due to the use of voxels and the size of 4.0 Å, the protein's resolution becomes generalised with loss of information to a certain degree. This data loss is often negligible as it does not affect the properties of the protein as a whole. The generalised model of each test protein combined with the use of voxels result in an approximated method of predicting binding sites. Although the clarity of details may not as competitive as wet lab or NMR approaches, however experiments in this study proved that the method is capable of approximating the location of potential binding sites based on just the depth factor alone. This eliminates the need to conduct rigorous scans over the protein's surface to determine sites of interest.

An internal cuboid is first grown inside the protein until a plateau is hit in each of the 6 faces. By plateau it means a plane that is fully filled with voxels. Any plane wherein a single voxel (or more) is empty results in the termination of the algorithm. Fig. 5 shows an illustration of the cuboid-growing method. Each of the plateaus defines the starting plane for visual projections executed from the plateau itself and headed outwards to the external environment. Voxels are built layer upon layer for each investigated face. In the final outcome crevices can be clearly discerned by observing the lowest-level voxels which are still externally exposed. Voxels on the plateau itself are assigned values of '1', the next higher level are assigned values of '2' and so on. A depth count is finally carried out and any crevices wherein the count is smaller than or equals to a user-specified value is shortlisted as a potential dock site. The final step is to assign and visualise all associated atoms and residues for the identified sites. Details of the algorithm are published in Ref. [21].

A colour-map is generated for each face with different levels of voxels distinctly colour-marked. The lowest levels are coloured red, followed by orange, yellow and so on. The choice of colours is optional and dependent on individuals' preferences. An integer matrix is instantiated as well alongside this colour map

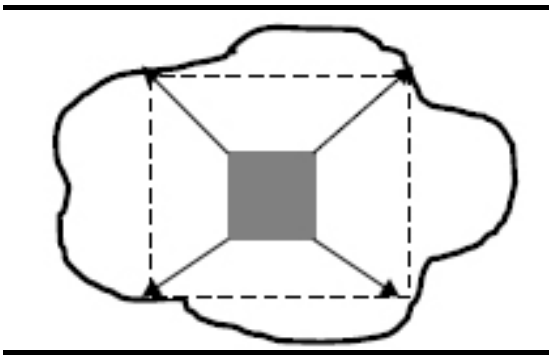


Fig. 5 The internal cuboid growing process.
An infant cube is first initiated in the center of the protein and subsequently stretched until the largest possible fully-filled cuboid is attained.

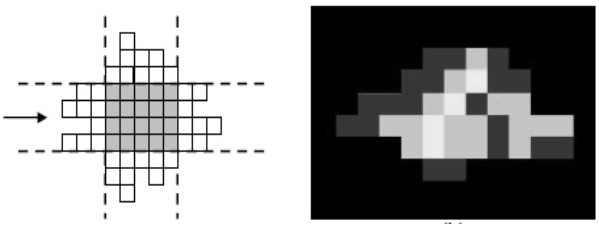
for the checking of potential sites. Fig. 6a shows a sample voxel representation of a protein and the colour map for the arrow-designated face is given in Fig. 6b. Fig. 7 depicts the integer matrix for the color map of Fig. 6b. All 0s represent blank spaces, 1 for voxels lying on the plateau, 2 for the next higher level, etc. A rule-based 3 × 3 window is introduced for use in the scans. The use of an integer matrix reduces computation time from processing a color map of a face from the voxel. Scanning each color map requires comparably higher computational time as each voxel consists of 1600 pixels (40 pixels for the length of 4.0 Å, and 40 × 40 for each voxel) and there are many voxels in a face. However a matrix requires only N × N integer representations therefore effectively reducing the computational time, a process many times faster than processing a single color map. Rules for the filter window are given in Fig. 8. The threshold value is dependent on the user, and in most cases a value equivalent to the average of the depth levels suffices. The final steps in the process are to obtain all associated surface atoms and check against projections from the RCSB PDB as well as benchmarks from past publications [15, 16].

3. Experimental Results

The results are divided into two main sections. The first section presents output for all selected FK506-bound proteins, and includes additional comparisons for protein 1FKF to published results. The second section presents output for all selected heme-bound proteins. The identified site of each protein is presented in images and compared against projections from RCSB PDB. Screenshots are captured for the latter and presented here. Note that all PDB projections include solvent molecules whereas these molecules were omitted in the implementation (in the tradition of past publications).

3.1 FK506-Bound Proteins

A total of 6 FK506-bound proteins were arbitrarily selected. These proteins are given as [PDB: 1FKF, 1BKF, 1YAT, 3VAW, 3UF8, 1C9H]. Each protein first undergoes processing as described in Experimental Setup and the potential sites consequently identified. With the exception of protein 1FKF wherein additional comparisons were carried out against published benchmarks, the remaining proteins are presented as pairs of images to illustrate the actual binding site and the identified region by the implemented method. Fig. 9 shows a comparison of the obtained output for protein 1FKF against a screenshot from RCSB PDB.



(a) (b)

Fig. 6 (a) Sample voxel representation of a protein. The grey cubes show the internal cuboid and how its faces act as plateaus for the internal-to-external exploration of each of the 6 surfaces, (b) the colour-map for the arrow-designated face in (a).

0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	2	1	0	0	0
0	0	0	0	1	1	2	3	1	1	0	0
0	0	1	1	1	2	3	1	2	2	0	0
0	1	1	2	2	3	2	2	1	2	2	0
0	0	0	0	2	3	2	2	1	2	1	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0

Fig. 7 The integer matrix corresponding to the colour-map in Fig. 5b.

A	B	C
D	X	E
F	G	H

Rule 1
If any neighbouring voxel of X is found to be of value '0', then X is set to '-1'. This implies that the current voxel is located at the outermost edge of the protein.

Rule 2
If at least a neighbour voxel has value >0 while all other neighbours are valued at either '-1' or >0, and X is valued at '1', then X is shortlisted into the dock site cluster.

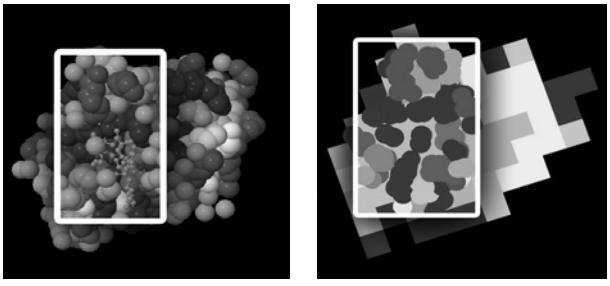
Fig. 8 The 3 × 3 filtering window and the implemented rules.

Based on the tabulated results (Table 1), it was shown that the voxel-based method correctly identified all residues associated to the binding site. The obtained excess residues count is high as well although this may be considered a minor problem as atoms from unconcerned residues may be found contained within the surface voxels as well. This is a compromise that has to be made from using a more generalised representation of the protein in voxel units. The excess residues are listed as: ILE90, TYR80, PRO45, ASP79, ARG40, GLN53, PRO78, ASP41, PHE48, ASN43, GLY58, SER39, GLY83, LYS44, PRO88, HIS25 and LYS47. Figs. 9-14 show all output obtained from the selected test set of proteins.

Table 1 Comparison of extracted residues from a wet lab experimentation, a geometric hashing-based approach and the implemented method for protein 1FKF.

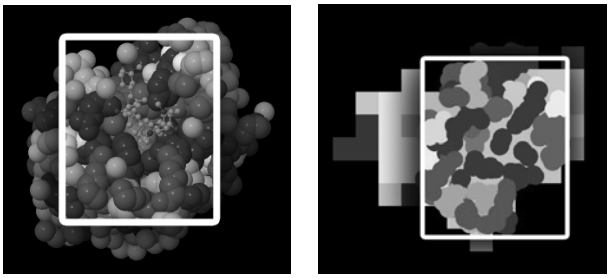
	Residue	WL*	GH**	Voxel
1	TYR26	Y	Y	Y
2	PHE36	Y	Y	Y
3	PHE46	Y	Y	Y
4	VAL55	Y	Y	Y
5	ILE56	Y	Y	Y
6	ARG57	-	Y	Y
7	TRP59	Y	Y	Y
8	ALA81	Y	Y	Y
9	TYR82	Y	Y	Y
10	PHE99	Y	Y	Y
11	ASP37	Y	-	Y
12	ARG42	Y	-	Y
13	GLU54	Y	-	Y
14	HIS87	Y	-	Y
15	ILE91	Y	-	Y

*WH-Wet Lab; **GH-Geometric Hashing.



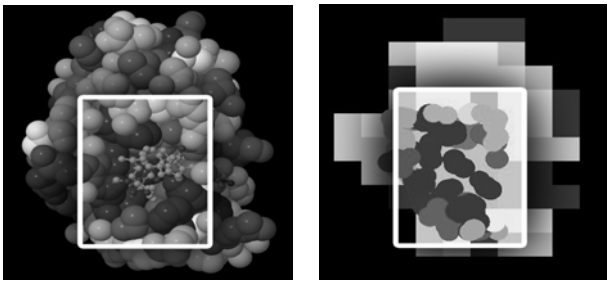
(a) (b)

Fig. 9 (a) Screenshot of FK506-bound protein 1FKF from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



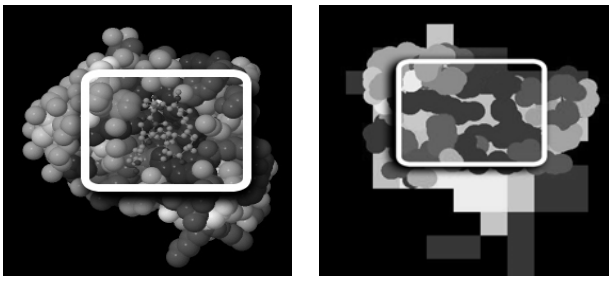
(a) (b)

Fig. 10 (a) Screenshot of FK506-bound protein 1BKF from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



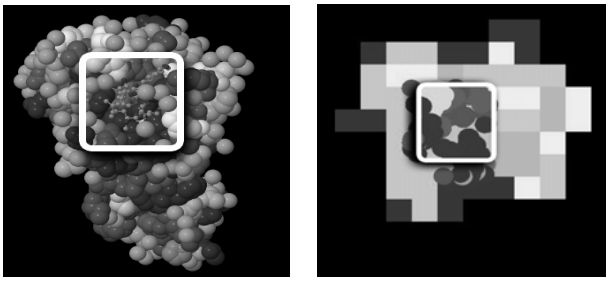
(a) (b)

Fig. 11 (a) Screenshot of FK506-bound protein 1YAT from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



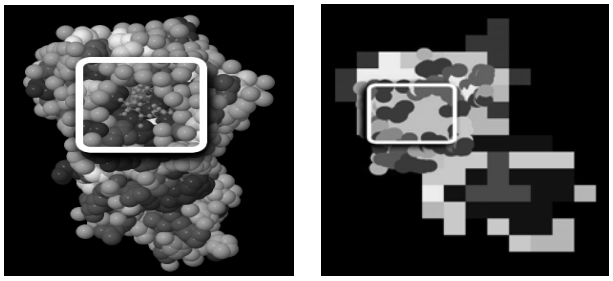
(a) (b)

Fig. 12 (a) Screenshot of FK506-bound protein 1C9H from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



(a) (b)

Fig. 13 (a) Screenshot of FK506-bound protein 3VAW from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.

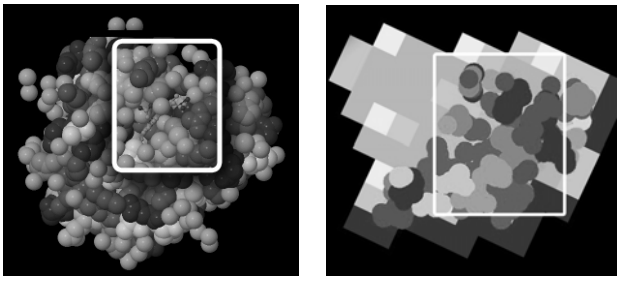


(a) (b)

Fig. 14 (a) Screenshot of FK506-bound protein 3UF8 from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.

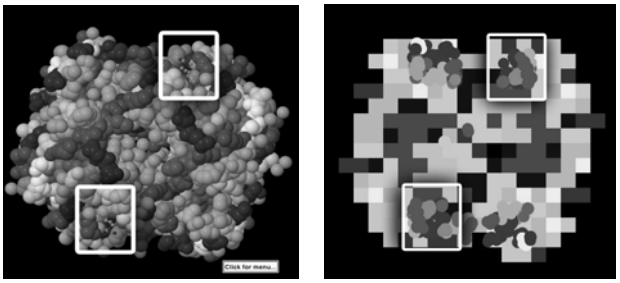
3.2 Heme-Bound Proteins

A total of 6 HEM-bound proteins were arbitrarily selected from RCSB PDB for the second part of this study. The proteins are listed as [PDB: 4MBN, 4HHB, 4FWX, 3TGA, 4DY9, 1B7V]. Similar to Section 3.1, the output for each of the proteins are presented as voxel method-RCSB projection pairs (Figs. 15-20).



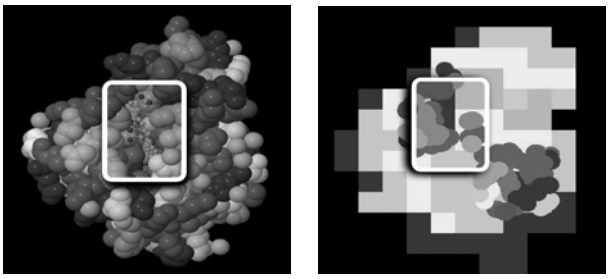
(a) (b)

Fig. 15 (a) Screenshot of heme-bound protein 4MBN from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



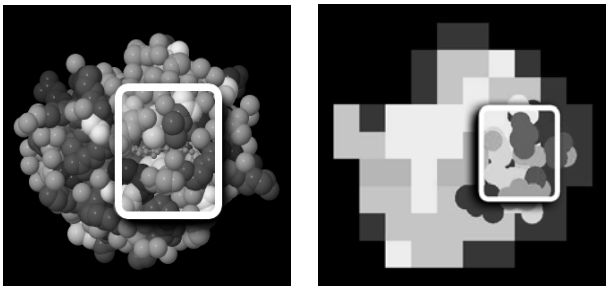
(a) (b)

Fig. 16 (a) Screenshot of heme-bound protein 4HHB from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



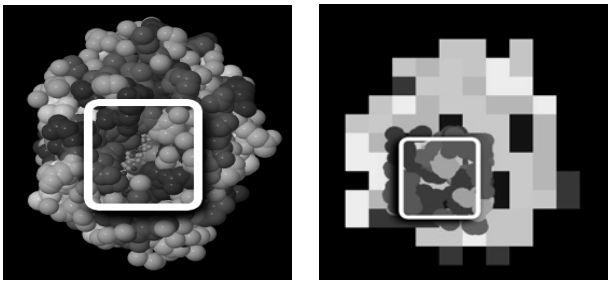
(a) (b)

Fig. 17 (a) Screenshot of heme-bound protein 4FWX from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



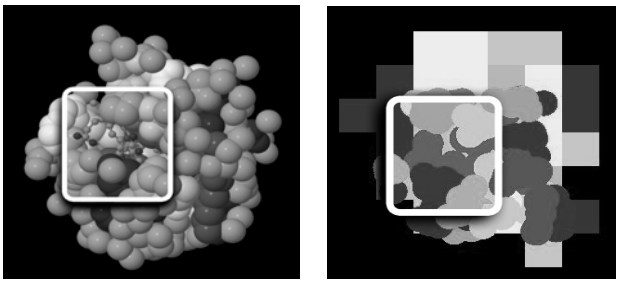
(a) (b)

Fig. 18 (a) Screenshot of heme-bound protein 4DY9 from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



(a) (b)

Fig. 19 (a) Screenshot of heme-bound protein 3TGA from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.



(a) (b)

Fig. 20 (a) Screenshot of heme-bound protein 1B7V from the RCSB PDB with ligand site shown, (b) screenshot of the identified site from the voxel-based approximated method.

3.3 Discussion of Results

The results obtained show that a voxel-based approach integrated with computer vision concepts is capable of identifying binding sites on proteins' surfaces. The algorithm was shown to correctly identify residues associated to the sites of interest, and is reasonable in terms of computational efficiency. By using processed datasets of shortlisted surface voxels alongside their atoms, the algorithm requires on average 10 to 15 seconds of processing time on a standard PC to identify and produce the output images from all 6 faces. The computational complexity is linearly dependent on the size of the protein as well as the area and depth for each of the 6 faces. Typically a larger and deeper area produces a higher number of voxels and this affects as well the time required for generating the colour map.

However the method has its limitations as well. The algorithm works well when binding sites are located within the boundaries of any of the 6 viewing platforms. The challenge comes when a binding site is located on any of the edges resulting in the site being 'split' between two (or more) viewing platforms. One way to tackle this issue is to carry out pre-rotations for realignment of the protein such that all binding sites are contained within the faces. Another limitation is when the sites are not sufficiently large and deep. The algorithm is targeted at identification of deep crevices and if a crevice appears to be shallow in comparison to another non-docking region, the latter gets identified leading to false positives of potential binding areas.

The proposed method is simple to implement and this is one of the key advantages of using grid-based environments. It has proven to be effective in approximated identifications of binding sites and produces results which are comparatively competitive. Although the excess residues obtained may be high, however they are mostly neighbouring entries which may help to indicate the location of sites of interest even if they do not directly contribute to the site. This scenario is unavoidable due to the use of voxels which represents the protein in a lower resolution. The main objective is to prove that with well-defined constraints and good understanding of the features of voxels, a grid-based environment is capable of delivering promising results in the context of protein surface studies.

4. Conclusions

An integrated approach for the approximated identification of protein binding sites is presented in this study. Based on the use of voxels and computer vision concepts, an algorithm is designed which scans for sites of interest. The study was carried out on 2 sets of proteins, namely FK506-bound proteins and heme-bound proteins, with each category consisting of 6 subjects. All the dock sites have been correctly identified. Compilation of the list of residues for the binding site of protein 1FKF showed successful extractions which are competitive against both the wet-lab and geometric hashing approaches. All dock sites were identified based on the depth factor. Successful identifications of sites were followed by the visualisation of atoms and residues associated to the sites and these were compared against projections from RCSB PDB. Although excess extractions were obtained, the method showed promise as quantified by the analysis of the results. This study highlights the fact that ligand-binding regions are usually found in sufficiently large and deep crevices.

References

[1] D. Fisher, R. Norel, H. Wolfson, R. Nussinov, Surface motifs by a computer vision technique: Searches, detections, and implications for protein-ligand interaction, Proteins 16 (1993) 278-292.

[2] R. Cheng, Z. Weng, A novel shape complementarity scoring function for protein-protein docking, Proteins 51 (2003) 397-408.

[3] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, Ligand Fit: A novel method for the shape-directed rapid docking of ligands to protein active sites, J. Mol. Graph. Model 21 (2003) 289-307.

[4] E. Kellenberger, J. Rodrigo, P. Muller, D. Rognan, Comparative evaluation of eight docking tools for docking and virtual screening accuracy, Proteins 57 (2004) 225-242.

[5] M. Nayal, B. Honig, On the nature of cavities on protein surfaces: Application to the identification of drug-binding sites, Proteins 63 (2006) 892-906.

[6] M. Weisel, E. Proschak, J.M. Kreigl, G. Schneider, Form follows function: Shape analysis of protein cavities for receptor-based drug design, Proteomics 9 (2009) 451-459.

[7] A. Via, F. Ferre, B. Brannetti, M.H. Citterich, Protein surface similarities: A survey of methods to describe and compare protein surfaces, Cell. Mol. Life Sci. 57 (2000) 1970-1977.

[8] S.L. Kinnings, R.M. Jackson, Binding site similarity analysis for the functional classification of the protein kinase familiy, J. Chem. Inf. Model. 49 (2009) 318-329.

[9] D.G. Levitt, L.J. Banaszak, POCKET: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids, J. Mol. Graph 10 (1992) 229-234.

[10] M. Hendlich, F. Rippmann, G. Barnickel, LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins, J. Mol. Graph. Model 15 (1997) 359-363.

[11] J. Jones, J.M. Thornton, Analysis of protein-protein interaction sites using surface patchs, J. Mol. Biol. 272 (1997) 121-132.

[12] A.A. Bogan, K.S. Thorn, Anatomy of hot spots in protein surfaces, J. Mol. Biol. 280 (1998) 1-9.

[13] J.F. Recio, M. Totrov, C. Skorodumov, R. Abagyan, Optimal docking area: A new method for predicting protein-protein interaction sites, Proteins 58 (2005) 134-143.

[14] P. Chakrabati, J. Janin, Dissecting protein-protein recognition sites, Proteins 47 (2002) 334-343.

[15] G.D. Van Duyne, R.F. Standaert, P.A. Karplus, S.L. Schreiber, J. Clardy, Atomic structures of the human immunophylin FKBP-12 complexes with FK506 and Rapamycin, J. Mol. Biol. 229 (1993) 105-124.

[16] K.P. Peters, J. Fauck, C. Frommel, The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria, J. Mol. Biol. 256 (1996) 201-213.

[17] L.W. Lee, A. Bargiela, Space-partition based identification of protein docksites, in: Proceedings of the 23rd European Conference on Modelling and Simulation, 2009, pp. 848-854.

[18] L.W. Lee, A. Bargiela, Statistical extraction of protein surface atoms based on a voxelisation method, in: Proceedings of the 24th European Conference on Modelling and Simulation, 2010, pp. 344-349.

[19] L.W. Lee, A. Bargiela, Protein surface atoms extraction: voxels as an investigative tool, Engineering Letters 20 (2012) 217-228.

[20] L.W. Lee, A. Bargiela, An investigation on the use of constraints-applied voxels in the search for protein surface atoms, in: Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering, 2012, pp. 655-661.

[21] L.W. Lee, A. Bargiela, Prediction and modelling of ligand binding sites using an integrated voxel method, in: Proceedings of the 26th European Conference on Modelling and Simulation, 2012, pp. 96-102.