# Statistical Extraction Of Protein Surface Atoms Based On A Voxelisation Method

Ling Wei Lee, Andrzej Bargiela

The University of Nottingham Malaysia Campus

Proteins – What Are They?

- ❖ One of the fundamental building blocks of all living beings.

- ❖ Maintains bodily functions. And your health.

- ❖ Made up of sequences of atoms like C, N, O etc.

- ❖ Reactivity of proteins – key and lock.

Protein   Ligand
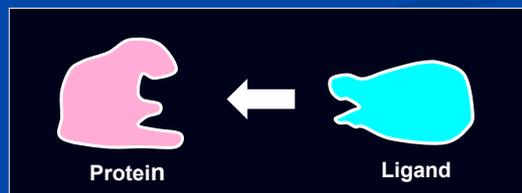
- ❖ The right binding agent fits to the right binding site.

So We're Talking About The Surfaces of The Proteins?

❖ Arrangement and composition of surface atoms.

❖ The need to understand better what lies on a protein's surface that leads it to bind to some agents and not to others.

❖ Extract the surface atoms.

❖ Challenging task – not many have attempted this.

❖ Current research heavily involves in :-

　* Sequence comparison and analysis

　* Structural comparison and analysis

Protein Sequence Comparison

❖ Strings of sequences

RRFDKYLAQP CKLSIACLAYSPKEYLGFWHRVYEKFVVRCQIR G

CKLSIACLAYSPKEYLGFWHRVYEKFVVRCQIR GAAMKKWPDH

❖ Determines the longest common subsequence

❖ Smith-Waterman [1], Needleman-Wunch [2] algorithms

❖ Computationally efficient

❖ But can it provide information on which atoms are surface atoms?

Protein Structural Comparison

❖ Analysis in 3D – spatial coordinates required

Common Substructure

❖ Extracts common sub-structures within proteins

❖ Computationally exhausting

❖ Higher level of processing – more signatures obtained

❖ But the question remains – can it identify the surface atoms responsible for protein-ligand interaction?

The Significance of Protein Surfaces

❖ Let's take two proteins as example.

❖ First scenario – both proteins stemming from same ancestor

❖ Second scenario – both proteins from same ancestor but with MUTATION

Protein A
Protein B
Protein B
Protein A
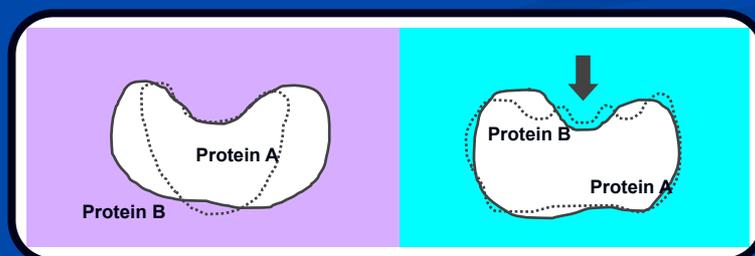
Past Research and Studies on Protein Surface Atoms

❖ Water molecule sized probe for surface inspection – the Connolly method [3]

❖ Cube representations for soft docking of proteins – Jiang and Kim [4]

❖ Ease of implementation and manipulation of cubic grids.

❖ Challenge in using cubic grids – orientation problem.

Before Rotation   After Rotation
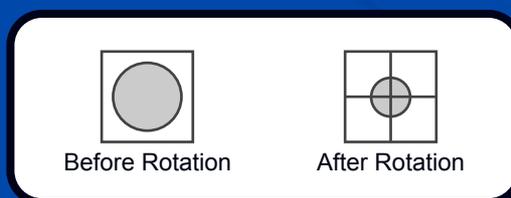
The Algorithm – Step 1

❖ The obtained PDB file first undergoes pre-processing.

❖ Vital information including the spatial coordinates, the atom element and all residues data are extracted.

❖ Data stored into new file with Van der Waals radius.

ATOM   2 CA GLN A 4   -5.941 24.227 39.819 1.00 96.72   C

ATOM 139.35 310.78 541.72 17.0 C 2.5 CA GLN A 4

❖ Of course, the spatial coordinates are all transformed to the positive domain.

The Algorithm – Step 2
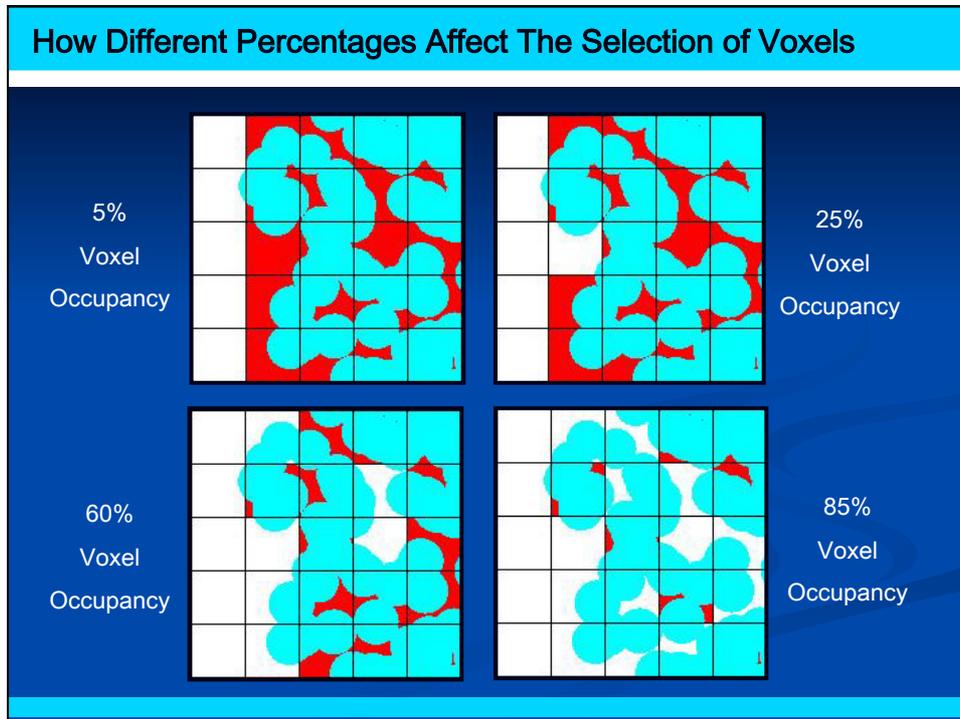
❖ The protein is rotated based on our 13 test cases.

	0-0-0	
20-0-0	0-20-0	0-0-20
30-0-0	0-30-0	0-0-30
45-0-0	0-45-0	0-0-45
60-0-0	0-60-0	0-0-60

❖ Affine transformation is used for the 3D rotation of the protein.

❖ Atoms experience the largest change of orientation at the above angles in the grid space domain.

❖ Processed information rechecked to ensure coordinates are in positive domain.

The Algorithm – Step 3

❖ Protein cast into experimental grid space.

❖ Tessellation carried out until the smallest unit is reached (40 units) in 2D space.

❖ Different occupancy percentages are used for the shortlisting of voxels (>5% to >95%).

❖ Definition of occupancy percentage:

$$\frac{\text{Number of protein pixels}}{\text{Size of voxel in pixels}} \quad \text{X} \quad 100\%$$

How Different Percentages Affect The Selection of Voxels

5%
Voxel
Occupancy

25%
Voxel
Occupancy

60%
Voxel
Occupancy

85%
Voxel
Occupancy

The Algorithm – Step 4

❖ Surface atoms are extracted for all the different voxel occupancy percentages.

❖ Extraction of surface atoms are based on the exposure value of surface voxels.

Exposure Value = 5   Exposure Value = 3

3/5/13
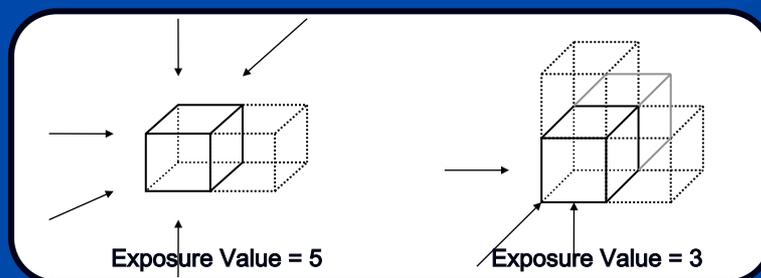
Results

❖ It was found that the optimum extractions (efficiency vs extraction accuracy) was determined at 40%-50% voxel occupancy.

❖ Extraction accuracy is defined by

20-0-0

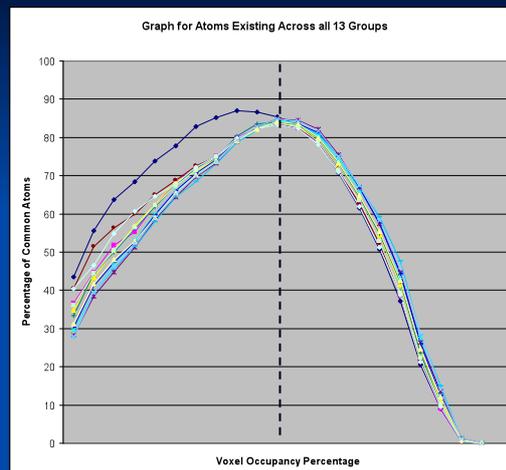Number of common atoms in current extraction
─── X 100%
Total common atoms across all test cases

	0-0-0	
20-0-0	0-20-0	0-0-20
30-0-0	0-30-0	0-0-30
45-0-0	0-45-0	0-0-45
60-0-0	0-60-0	0-0-60

(13 Cases)

Results (Graphs)

Graph showing accuracy of extractions for common atoms existing across all 13 test cases.
Optimum extractions occur between **40% - 50%** occupancy.
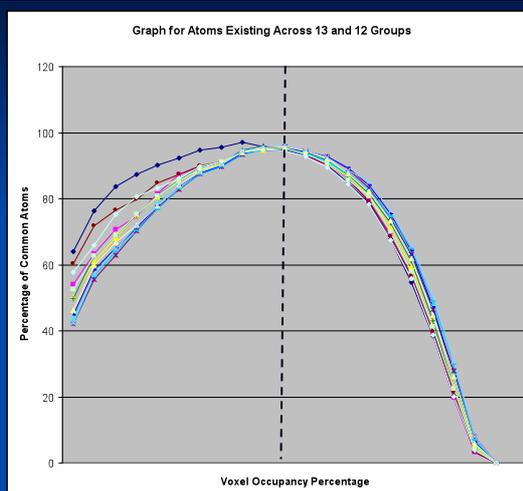
7

Results (Graphs)

Graph showing accuracy of extractions for common atoms existing across 13, 12 test cases.

Optimum extractions occur between **40% - 50%** occupancy.

Results (Tables)

Rotation Case (x, y, z)	Peak Percentage (%)	Extraction Percentage (%)	Rotation Case (x, y, z)	Peak Percentage (%)	Extraction Percentage (%)
(0, 0, 0)	40	87.06	(0, 0, 0)	40	97.02
(20, 0, 0)	50	83.55	(20, 0, 0)	45	95.16
(30, 0, 0)	50	83.96	(30, 0, 0)	50	95.11
(45, 0, 0)	50	84.52	(45, 0, 0)	50	95.41
(60, 0, 0)	50	84.84	(60, 0, 0)	50	95.80
(0, 20, 0)	50	83.55	(0, 20, 0)	50	94.72
(0, 30, 0)	50	83.89	(0, 30, 0)	45	95.93
(0, 45, 0)	50	84.00	(0, 45, 0)	50	95.39
(0, 60, 0)	50	84.24	(0, 60, 0)	50	95.17
(0, 0, 20)	50	83.45	(0, 0, 20)	45	94.85
(0, 0, 30)	50	83.72	(0, 0, 30)	50	95.28
(0, 0, 45)	50	84.03	(0, 0, 45)	50	95.76
(0, 0, 60)	50	84.84	(0, 0, 60)	50	96.18

List of all rotation cases with their peak percentages and extraction percentages for common atoms in all 13 orientations	List of all rotation cases with their peak percentages and extraction percentages for common atoms in 13 and 12 out of 13 orientations

Discussion

❖ Tessellation and usage of voxels = effective and efficient.

❖ Satisfactory results up to 97% accuracy.

❖ Analysis can be carried out on the extracted surface atoms to determine common motifs and patterns.

❖ Next stage of research.

❖ Classification can be carried out on these signatures thus producing a new database.

References

[1] Smith T., Waterman M. 1981. "Identification of Common Molecular Subsequences". *Journal of Molecular Biology*, No.147, 195-197.

[2] Needleman S., Wunsch C. 1970. "A General Method Applicable to The Search for Similarities in The Amino Acid Sequence of Two Proteins". *Journal of Molecular Biology*, No.48, 443-453.

[3] Cao, J., Pham D.K., Tonge L., Nicolau D.V., 2002. "Predicting Surface Properties of Proteins on The Connolly Molecular Surface". *Smart Materials and Structures*, No.11, 772-777.

[4]] Jiang F., Kim S.H., 1991. "'Soft Docking': Matching of Molecular Surface Cubes". *Journal of Molecular Biology*, No.219, 79-102.

Q&A Session

QUESTIONS?

=)

Thank You!