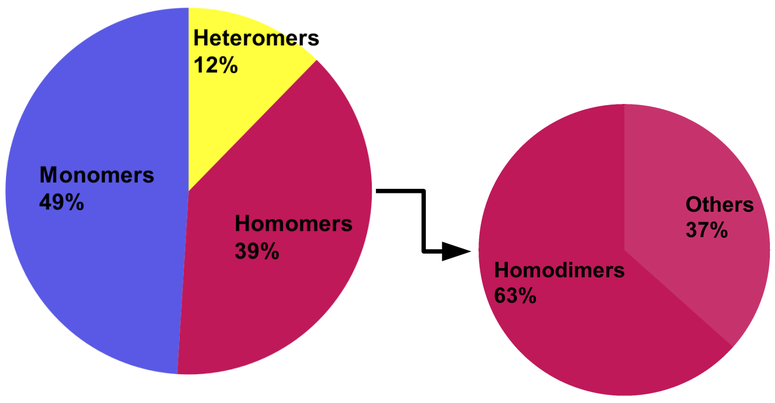
HOMODIMER PREDICTION
Pattern Recognition of Symmetrical Protein-Protein Interaction Sites

Akanksha Srivastava and Jean-Christophe Nebel
Department of Computing, Information Systems & Mathematics, Kingston University
akanksha.iitkgp.2012@gmail.com, J.Nebel@kingston.ac.uk

Kingston University London

Rational

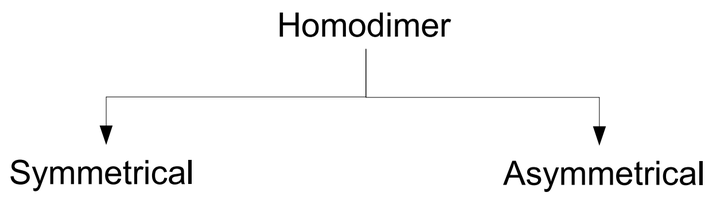
Homodimers constitute 39% of PDB [1] proteins[1].



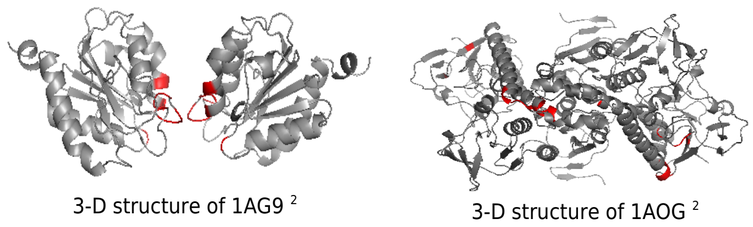
Homodimer proteins play important role in cellular functions and biological processes.

Pattern Recognition could help in detecting potential homodimer interactions, from a protein sequence, which would allow modelling proteins and aid in drug discovery.

Homodimer Classification

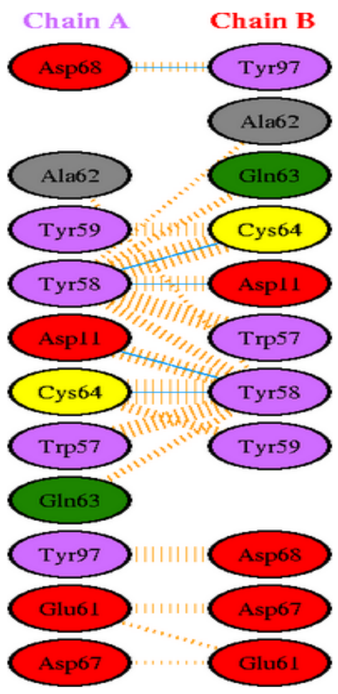


Symmetrical Homodimers refers to the homodimers in which the interface region consists of same amino acids present at same positions in both chain A and chain B.



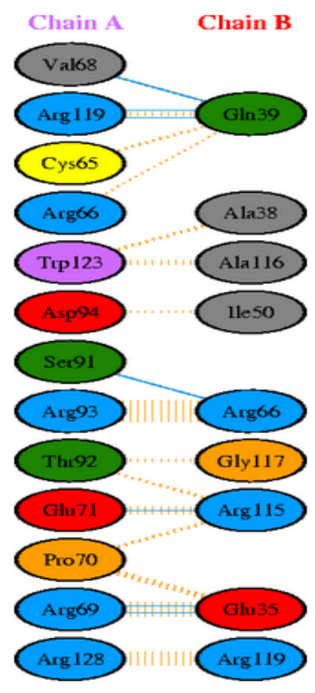
3-D structure of 1AG9 [2]

3-D structure of 1AOG [2]



PDBsum Entry for 1AG9

Eg. of symmetrical homodimer



PDBsum Entry for 1AOG

Eg. of Asymmetrical homodimer

Formula used for calculating score

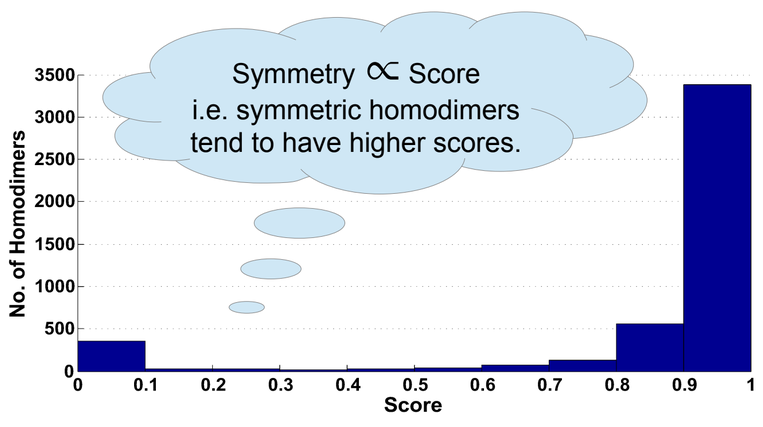
Score = ½ * (A∩B) * (1/A+1/B)

where,

A is the no. of residues in the interface region of chain A,

B is the no. of residues in the interface region of chain B,

A ∩ B is the total no. of same residues at same position in Chain A and Chain B.



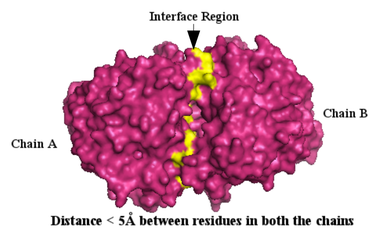
Experiment 1 : Binding Site properties of Symmetrical Homodimers

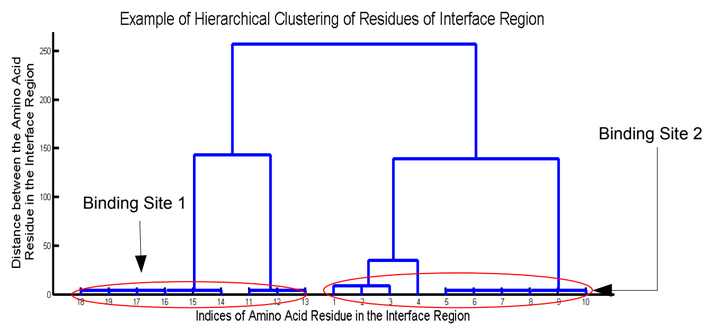
Step 1 : Query search in PDB Proteins.

Query	Structures Found
Homologue Removal - 90% Identity Cutoff of Chain Type: there is a Protein chain but not any DNA or RNA or Hybrid and Stoichiometry in biological assembly: Stoichiometry is HOMOMER and Sequence ID is 0.95 and Stoichiometry in biological assembly: Stoichiometry is A2 and Sequence ID is 0.95 and Oligomeric state Search : Min Number of oligomeric state=2 Max Number of oligomeric state=2	8757

Step 2 : Hierarchical Clustering of amino acid residues of interface region of each 8757 PDB proteins.

Clustering Parameters used:
**Minimum 4 residues*
**Cluster density>50%*





Step 3 : Estimate the distribution of number of Binding Sites. See Figure 1.

Step 4 : Clustering of all homodimers having one binding site by using Needleman Wunsch Algorithm and scoring matrix based on AA Index. See Figure 2.

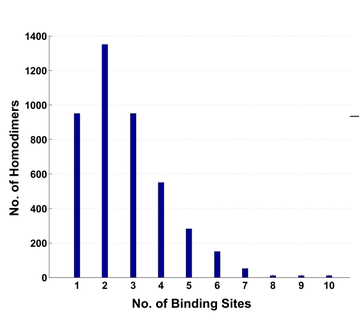


Figure 1

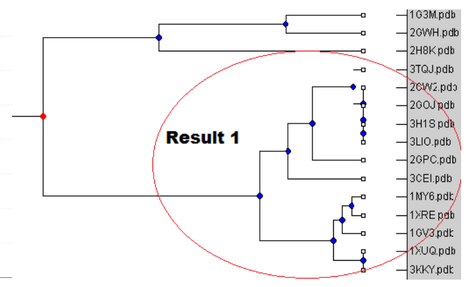


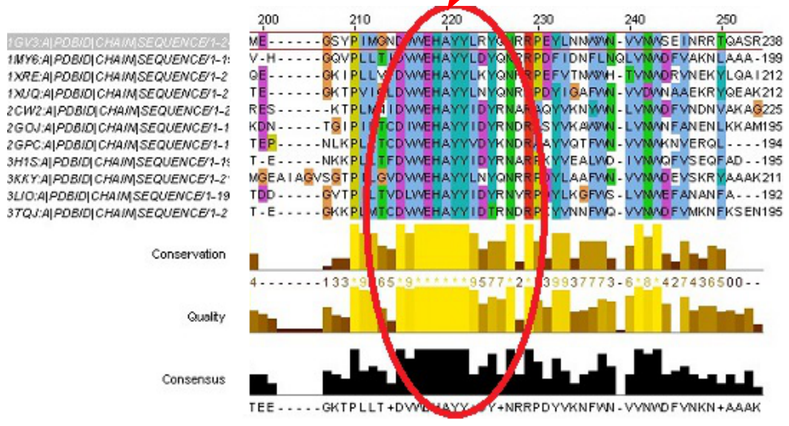
Figure 2

Results of Experiment 1

The reasons why the proteins in the said cluster appear as they do:

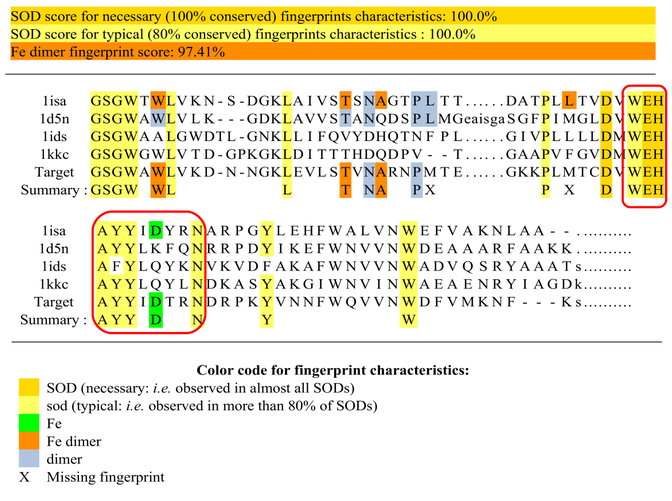
(a) Predicted family is Fe/Mn SOD family.

(b) Predicted binding site is: WEHAYYIDYRN



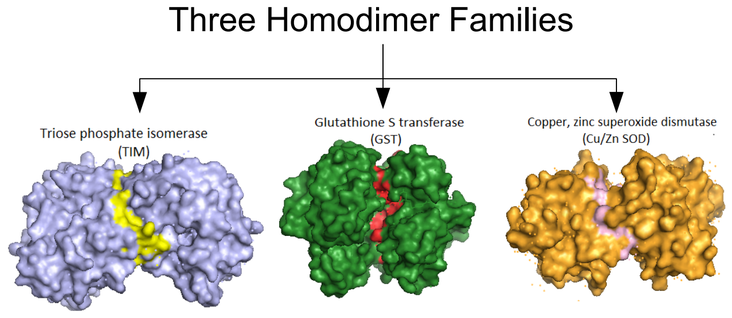
Multiple Alignment of sequences from Cluster highlighted in step 4 by using clustalW

This result is in line with SODa [3] which shows that Binding Site predicted is necessary for SOD family.



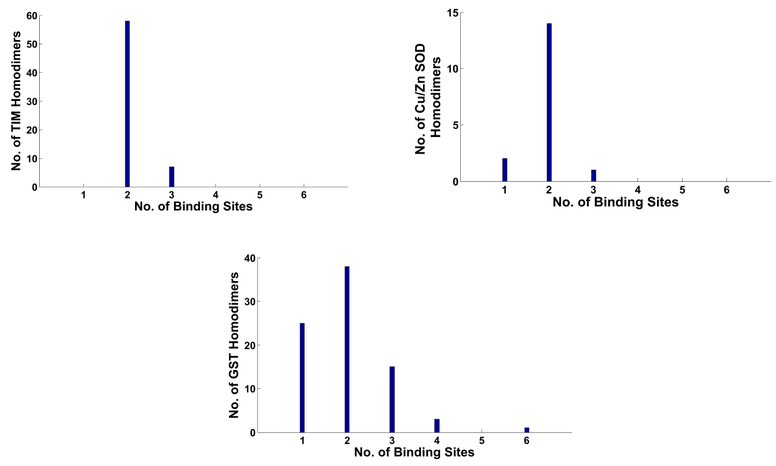
Experiment 2 : Analysis of Specific Homodimer Families

Out of six homodimer families, mentioned by Valdar and Thornton [2], three are analysed in this experiment.

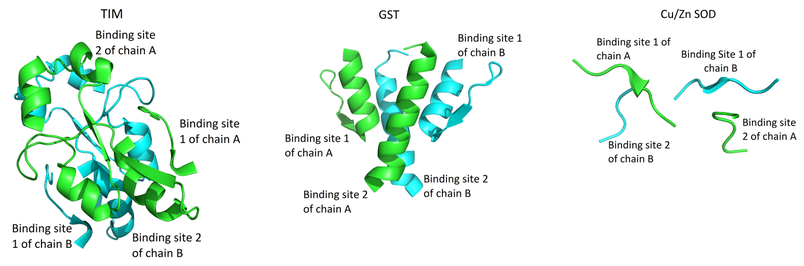


Step 1 : Hierarchical Clustering of amino acid residues of each family using clustering parameters as mentioned in Experiment 1.

Step 2 : Histogram distribution of homodimers of each family in the Binding Sites (BS).



Step 3 : Analysing the binding pattern with maximun number of Homodimers, for example, 2 for each family in this case.



Results of Experiment 2

The binding pattern predicted for TIM family and Cu/Zn SOD family are shown in Figure 3 and Figure 4 , respectively.

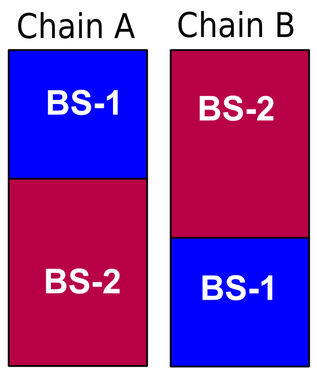


Figure 3

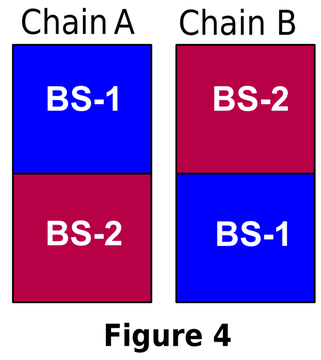


Figure 4

Binding Site of GST family needs further analysis.

Future Work

- Pattern recognition of Fe/Mn SOD family.
- Finding the fingerprints for the pattern from the primary squences of the above studied families.
- Looking for other homodimer familes having same fingerprints.

References

[1] Berman,H.M. et al. (2000) The Protein Data Bank, Nucleic Acids Research, 28, 235-242.

[2] Valdar,W.S.J and Thornton,J,M.(2001) Protein–Protein Interfaces: Analysis of Amino Acid Conservation in Homodimers, PROTEINS: Structure,Function,andGenetics,42:108–124.

[3] Kwasigroch,J.M. et al.(2008) SODa: An Mn/Fe superoxide dismutase prediction and design server.URL: http:// babylone. Ulb. ac. be/soda.

[1] All data based on the PDB release of 11thMarch 2013.

[2] Generated by PyMOL.