Side-Chain Clusters in Protein Structures

N. Kannan kannan@mbu.iisc.ernet.in
S. Vishveshwara sv@mbu.iisc.ernet.in
Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India-560 012

Keywords: clusters, protein folding, protein-protein recognition, eigenvalues, cluster centers

1 Introduction

Identification of side-chain clusters in protein structures is important from protein stability, function and folding point of view. Specific side-chain interactions in the protein are important for the stabilization of the tertiary structure [2]. Hydrophobic and charged clusters on the protein surface are important in protein-protein recognition and protein-DNA interaction [1, 6]. Often a network of charged side-chains is found near the metal binding site and active-site of the protein [5].

A method for detecting such side-chain clusters using a Graph spectral method [3] is described here. In a protein structure, the side-chain interactions are represented by a weighted graph (as mentioned in the methods) and the constructed graph is represented by a Laplacian matrix. The clustering information is obtained from the vector components of the second lowest eigenvalue and the cluster centers are obtained fron the vector components of the top eigenvalues [4]. This method uses global information for clustering and is computationally efficient as a single numeric computation of required to identify clusters of interest.

2 Methods

A graph for a protein structure is constructed by considering the C_β atoms of the side-chains as nodes of the graph and the nodes of the graph are connected with and edge weight corresponding to $1/d_{ij}$ (where d_{ij} is the distance between nodes i and j). The Laplacian matrix (B) for the constructed graph is obtained from the adjacency matrix (A) and the degree matrix (D) of the graph [3]. The Laplacian matrix B is given by $B = D - A$. On diagonalizing the Laplacian matrix, the vector components of the second lowest value with a constant vector component value form a cluster [4]. For example in the case of Lysozyme molecule the vector components of the second lowest eigenvalue is shown in column 4 of Table 1. The residues Tyr 20, Lys 97 and Arg 101 with a constant vector component value of -0.310 form a cluster. Similarly, the vector components of the top 7 eigenvalues are shown in columns 7-13. The vector components of the higher eigenvalues have information on only one of the clusters. The information regarding the 1st cluster is found in the vector components corresponding to the 5th highest eigenvalue. Tyr 20 forms the center of this cluster as the magnitude of its vector component is the highest (0.814)(Table 1).

3 Results

Cluster analysis using the Graph spectral method was applied for a dataset of proteins which were well studied from protein structure, function and folding point of view. The detected clusters were found to emanate from different secondary structural regions of the protein, stabilizing the tertiary fold. In most of the proteins studied, clusters on the protein surface were also identified. Clusters

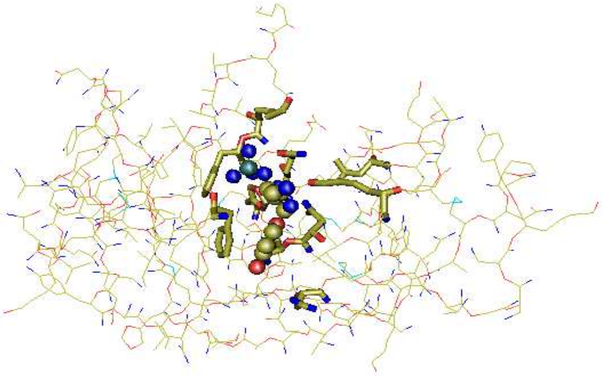

Figure 1: Cluster near the active site in RNase A molecule. The cluster residues are shown in BONDS representation and the ligand is shown in VDW representation.

close to the active and binding site of the protein was detected. In Fig. 1 is shown the cluster detected near the active site of the protein Ribonuclease A. The identified clusters were also found to be conserved in topologically similar proteins. The detected clusters show a good correlation with the folding intermediates as probed by hydrogen exchange experiments. At present this algorithm is being used to predict the active and binding sites of the protein from its native structure.

References

[1] Anderson, J.E., Ptashne, M. and Harrison, S.C., Structure of the repressor-operator complex of bacteriophage 434, *Nature*, 326(6116):846–852, 1987.

[2] Chou, K.C., Nemethy, G., and Scheraga, H.A., Energetics of interactions of regular structural elements in proteins, *Accts. Chem. Res.*, 23:134–141, 1990.

[3] Hagen, L. and Kahng, A. B., New Spectral Methods for Ratio Cut Partitioning and Clustering, *IEEE Trans. Computer-Design*, 1074–1084, 1992.

[4] Kannan, N. and Vishveshwara, S., Identification of side-chain clusters in protein structures by Graph spectral Method, *J. Mol. Biol.* , 292:441–464, 1999.

[5] Ng, K. K., Drickamer, K., and Weis, W.I., Structural analysis of monosaccharide recognition by rat liver mannose-binding protein, *J. Biol. Chem.*, 271: 663–274, 1996.

[6] Young, L., Jernigan, B.L., and Covell, D.G., A role for surface hydrophobicity in protein-protein recognition, *Protein Science,* 3:717–729, 1994.

Table 1: Cluaters and eigenvector components in Lysozyme (1LZ1)

Cl[a] No	Residue Number	Residue Name	Eigenvector of 2nd lowest eigen value	% ASA[b]	SS[c]	*Magnitude of vector components of the top eigenvalues*						
						1	2	3	4	5	6	7
1	20	TYR	-0.310	25.739	S2	0.000	0.000	0.000	0.000	**0.814**	0.000	0.000
	101	ARG	-0.310	50.130	T11	0.000	0.000	0.000	0.000	**0.461**	0.000	0.000
	97	LYS	-0.310	24.702	H4	0.000	0.000	0.000	0.000	**0.354**	0.000	0.000
2	3	PHE	-0.272	2.671	C2	0.000	0.000	0.000	0.000	0.000	**0.814**	0.000
	7	GLU	-0.272	37.944	H1	0.000	0.000	0.000	0.000	0.000	**0.466**	0.000
	8	LEU	-0.272	0.000	H1	0.000	0.000	0.000	0.000	0.000	**0.348**	0.000
3	54	TYR	-0.024	10.498	S6	**0.795**	0.000	0.000	0.000	0.000	0.000	0.000
	67	ASP	-0.024	5.018	C7	**0.560**	0.000	0.000	0.000	0.000	0.000	0.000
	81	CYS	-0.024	1.372	H3	**0.235**	0.000	0.000	0.000	0.000	0.000	0.000
4	112	TRP	0.022	6.819	H6	0.000	**0.861**	0.000	0.000	0.000	0.000	0.000
	117	GLN	0.022	33.679	T13	0.000	**0.289**	0.000	0.000	0.000	0.000	0.000
	107	ARG	0.022	51.161	H5	0.000	**0.362**	0.000	0.000	0.000	0.000	0.000
	106	ILE	0.022	2.796	H5	0.000	**0.210**	0.000	0.000	0.000	0.000	0.000
5	28	TRP	0.030	0.000	H2	0.000	0.000	**0.816**	0.000	0.000	0.000	0.000
	17	MET	0.030	0.000	C3	0.000	0.000	**0.434**	0.000	0.000	0.000	0.000
	23	ILE	0.030	7.904	S3	0.000	0.000	**0.382**	0.000	0.000	0.000	0.000
6	99	VAL	0.199	1.493	H4	0.000	0.000	0.000	**0.817**	0.000	0.000	0.000
	64	TRP	0.199	14.441	T8	0.000	0.000	0.000	**0.415**	0.000	0.000	0.000
	109	TRP	0.199	7.465	C12	0.000	0.000	0.000	**0.407**	0.000	0.000	0.000
7	58	GLN	0.349	3.191	T7	0.000	0.000	0.000	0.000	0.000	0.000	**0.811**
	53	ASP	0.349	23.420	S6	0.000	0.000	0.000	0.000	0.000	0.000	**0.488**
	35	GLU	0.349	16.328	H2	0.000	0.000	0.000	0.000	0.000	0.000	**0.323**

a-Cl no: Cluster number,b-ASA: Accessible surface area
c-SS: Secondary structure (S, sheet; H, helix; T, turn; C, coil)