FINDING RESIDUES RESIDING AT THE PROTEIN CORE AND PROTEIN SURFACE USING GRAPH THEORETICAL TECHNIQUES

R.Mageswari
Assistant Professor, Department of Mathematics
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya
Kanchipuram - 631 561, Tamilnadu, India

V.Meena
Scholar, Department of Mathematics
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya
Kanchipuram - 631 561, Tamilnadu, India

Abstract: Graph of the protein can be obtained from the protein contact map. In this paper, residues residing at the core of the protein with protein id 2WPO can be identified clearly from the minimal connected dominating sets.

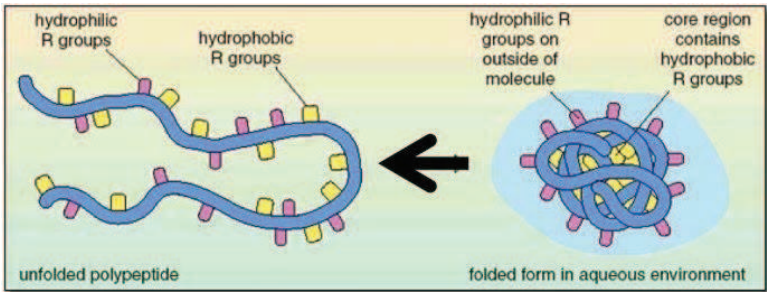

Figure 1: Hydrophobic Core and Hydrophilic Surface of A Protein

Keywords: Contact Map, Residues.

1. Introduction: Amino Acids are organic compounds containing amine (-NH2) and carboxyl (-COOH) functional groups, along with a side chain (R group) specific to each amino acids. There are twenty amino acids. Proteins are large bio molecules or macro molecules consisting of one or more long chains of amino acid residues. There are thousands of different proteins and the characteristic of each type of protein is determined by a different gene.

Proteins make up the cellular structure of organisms and proteins structure determines its biological function. Understanding the structure of protein is significantly useful in understanding its function. Around one lakh protein 3D structures were identified by the experimental methods such as X-ray crystallography, NMR and Electron Microscopy and are available in the Protein Data Bank [1]. But, experimental methods are more expensive, time consuming and the process is also tedious. Hence, mathematical methods are introduced in the analysis of proteins.

Among the different mathematical methods for protein structure identification, graph theoretical methods are easy to estimate. Three dimensional structure of a protein is well determined by its continuous folds. Protein core is occupied by the amino acids with the higher average degree [2-8]. Hydrophobic amino acids reside at the core of the protein to avoid water [9] (Figure 1).

Contact map plays an important role in understanding proteins structure [10]. Using contact map of the protein, the hydrophobic and hydrophilic amino acids are predicted by its total average degree using the HYDCORE algorithm [11]. Also, specific folds [12] in the protein are identified using contact map. In [13],

a new method is introduced to identify the hydrophobic residues from the minimal connected dominating set [14] of the graph of the protein. In this paper, hydrophobic residues of the protein with PDB id 2WPO are obtained using the concept of dominating sets.

2. Preliminaries

2.1 Graph: Let G = <V, E> be a graph in which V denotes the set of vertices, E denotes the set of edges and |V (G)| is the number of vertices of G. Two vertices u and v of G are called adjacent or neighbors if u-v is an edge of G. The degree d(v) of a vertex v is denoted by d(v) and is defined as the number of neighbors of v.

2.2 Connected Graph: A graph G is said to be connected if every pair of its points are connected.

2.3 Dominating Set of a Graph: Let G = (V,X) be a graph. A subset S of V is called a dominating set if every vertex in V-S is adjacent to a vertex in S.

2.4 Minimal Connected Dominating Set: A Minimal Connected Dominating Set is a connected dominating set from which no vertex can be removed without destroying its dominance property.

2.5 Contact Maps: The distance between two residues may be defined by the distance between two alpha carbon (C_α) atoms or between two beta carbon (C_β) atoms or it may be the minimum distance between any pair of atoms belonging to the side chain or to the backbone of two residues. A contact between two given atoms (or residues) exists when a certain distance is below a given threshold.

Let P be a protein with n atoms which are labeled 1,2,3,....n. We define the contact map of the protein as a matrix,

$T = (t_{ij})_{1\le i,,j\le n}$

where $t_{ij} = 1$ if $i \neq j$ and d_{ij} (Euclidean distance between i & j) $\le$ 6Å

= 0 otherwise.

2.6 Graph of the Protein: A protein can be considered as a graph G = <V,E> for which each vertex $v_k \in$ V represents a residue of the protein and each v_i-$v_j \in$ E represents a contact between two residues v_i and v_j. On the other hand, there is an edge v_i-$v_j \in$ E if $t_{ij} = 1$.

3. Methodology: Coordinates of the alpha carbons of the Protein with PDB id 2WPO is taken from the Protein Data Bank and the three dimensional structure of the protein for the required alpha carbons can be seen using the software Molsoft. The distances between the alpha carbons have been noted and the contact map, graph of the protein have been obtained as given below.

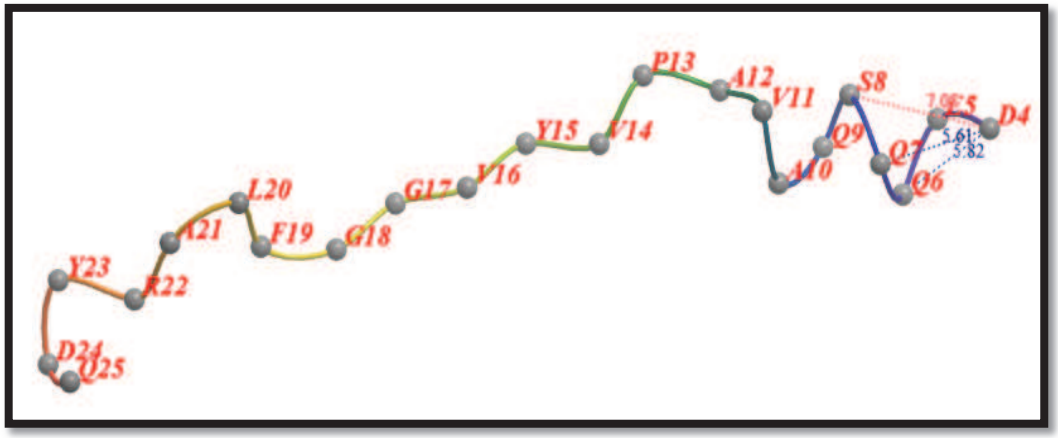

Figure 1: Wire Representation of the Protein 2WPO

Table 1: Residues and its Adjacent Residues with 6 Å

Residue	Vertex Number	Adjacent Residues
D_4	1	E_5,Q_6,Q_7
E_5	2	D_4,Q_6,Q_7,S_8,Q_9
Q_6	3	D_4,E_5,Q_7,S_8,Q_9
Q_7	4	D_4,E_5,Q_6,S_8,Q_9,A_{10}
S_8	5	L_5,Q_6,Q_7,Q_9,A_{10},V_{11}
Q_9	6	E_5,Q_6,Q_7,S_8,A_{10},V_{11},A_{12}
A_{10}	7	Q_7,S_8,Q_9,V_{11},A_{12}
V_{11}	8	S_8,Q_9,A_{10},A_{12}
A_{12}	9	Q_9,A_{10},V_{11},P_{13}
P_{13}	10	A_{12}, V_{14}
V_{14}	11	P_{13},Y_{15}
Y_{15}	12	V_{14},V_{16}
V_{16}	13	Y_{15}, G_{17}
G_{17}	14	V_{16},G_{18}
G_{18}	15	G_{17},F_{19}
F_{19}	16	G_{18}, L_{20}, A_{21}
L_{20}	17	F_{19}, A_{21}
A_{21}	18	F_{19}, L_{20}, R_{22}
R_{22}	19	Q_{21}, Y_{23}, D_{24}, Q_{25}
Y_{23}	20	R_{22}, D_{24}, Q_{25}
D_{24}	21	Y_{23}, R_{22}, Q_{25}
Q_{25}	22	R_{22}, Y_{23}, D_{24}

Contact Map of the protein is obtained as a matrix given below.

$$\begin{pmatrix}
0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&1&1&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&1&1&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&1&1&1&1&0&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&1&1&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&1&1&1&0&1&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&0
\end{pmatrix}$$

Using the contact map shown above, the graph of the protein is obtained as below.

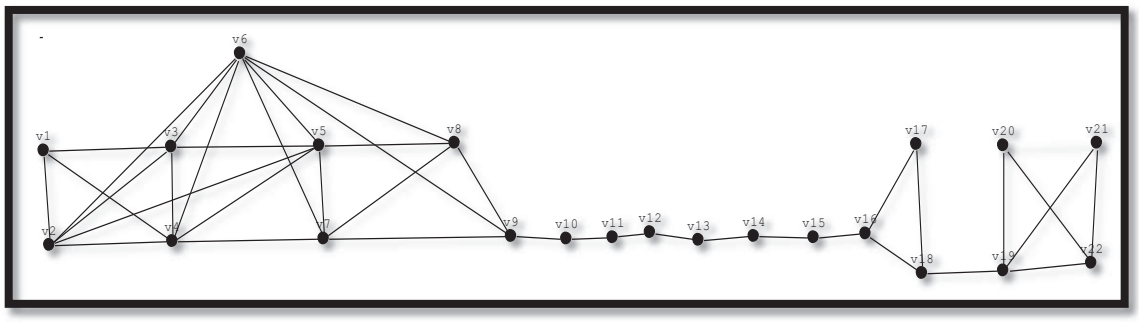

Figure 2: Graph of the Protein 2WPO

The minimal connected dominating sets of the above graph are,
$D_1 = \{ v_4, v_5, v_6, v_7, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{22}\}$
$D_2 = \{v_6, v_9, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{22}\}$
$D_3 = \{v_4, v_5, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{22}\}$

Conclusion: Residues residing at the core of the protein 2WPO are listed as $v_4, v_5, v_6, v_7, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{22}$ and the remaining residues v_1, v_2, v_3 and v_8 are hydrophilic.

References:

1. The Protein Data Bank: Exploring Biomolecular Structure. Nature Education 2010, 39.
2. Cornette J L, Cease K B, Margalit H, Spouge J L, Berzofsky J A, DeLisi C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. J Mol Biol. 1987, 195, 659–685.
3. Eisenberg D, Schwarz E, Komaromy M, Wall R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol. 1984, 179, 125–142.
4. Engelman D M, Steitz T , Goldman A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem.1986, 15, 321–353.
5. Hopp T P, Woods K R. A computer program for predicting protein antigenic determinants. Mol Immunol. 1983, 20, 483–489.
6. Janin J. Surface and inside volumes in globular proteins. Nature. 1979, 277,491–492.
7. Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982,157,105–132.
8. Rose G D, Geselowitz A R, Lesser G J, Lee R H, Zehfus M H. Hydrophobicity of amino acid residues in globular proteins. Science. 1985,229,834–838.
9. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Prot Chem. 1959,14, 1–63.
10. Mageswari R, Srinnivasa Rao K, Sivakumar K: Protein similarity/dissimilarity using contact map and matching index, *Mathematical Sciences International Research Journal*, 2014,3 672-676.
11. Mahnaz Habibi, Changiz Eslahchi, Mehdi Sadeghi , Hamid Pezashk: The interpretation of protein structures based on graph theory and contact map Open Access Bioinformatics 2010 ,2, 127-137.
12. Dahiyat .B I; Mayo .S L.: Probing the role of packing specificity in protein.. Natl Acad Sci U S A. 1997, 94, 10172–10177.
13. Mageswari.R, SrinivasaRao.K, Sivakumar.K, Prediction of Hydrophobic Core using Contact Map and Minimal Connected Dominating Sets, Indian Journal of Science, 2015, 13(37), 24-28.
14. Narsingh Deo: Graph Theory with Applications to Engineering and Computer Science, Prentice Hall of India Private Limited, New Delhi 2008.

***