Trends in Bioinformatics

ISSN 1994-7941

science alert

ANSInet
an open access publisher
http://ansinet.com

Trends in Bioinformatics 8 (2): 63-74, 2015
ISSN 1994-7941 / DOI: 10.3923/tb.2015.63.74

Detection of Active Site Residues in Bovine Rhodopsin Using Network Analysis

[1]I. Arnold Emerson and [2]Preeti Tabitha Louis
[1]Bioinformatics Division, School of Bio Sciences and Technology,
[2]School of Social Sciences and Languages, VIT University, Vellore, 632014, Tamil Nadu, India

Corresponding Author: I. Arnold Emerson, Bioinformatics Division, School of Bio Sciences and Technology, VIT University, Vellore, 632014, Tamil Nadu, India

ABSTRACT

Topological determinants of the protein structures can be obtained by modeling them as undirected network, which may relate to functionally important residues. In this study, we aim to analyze the Bovine rhodopsin structure (PDB ID: 1U19A) a GPCR family protein. We modeled the protein structure as, an undirected network and network parameters were calculated and compared with the control random networks. Our findings show that the protein contact network possesses a small world property. The functionally important residues in the protein contact network were identified using residue centrality. The statistical significance of the central residues was determined using the Z-score values of the residue centrality. Results show that residues with high Z-score values are highly conserved, are in close proximity with the ligand and are also situated closer to the centre of mass of the protein.

Key words: Residue centrality, small world property, center of mass

INTRODUCTION

Proteins are biological macromolecules synthesized in the cell as a linear chain of amino acids, which then folds into a three-dimensional structure comprising of different secondary structural elements, such as helices, sheets and coils, by making short and long contacts between the amino acid residues along the chain. They perform diverse biochemical functions and also provide structural basis in living cells. It is important to understand how proteins fold into their native-state structures and the relevance of these structures to their function. Network analysis of protein structures is one such attempt to understand possible relevance of various network parameters to protein structure and functioning.

Biological systems have been studied to understand protein interaction network (Jeong *et al*., 2001), networks of interacting components of living cells possessing scale-free properties (Wolf *et al*., 2002), metabolic and biochemical pathway networks (Ravasz *et al*., 2002; Holme *et al*., 2003) and gene regulatory networks (Shen-Orr *et al*., 2002). Network representation of protein structures has also been useful in analyzing interactions between amino acids in studies on protein folding (Heringa and Argos, 1991; Vendruscolo *et al.,* 2002), identifying the functional residues (Amitai *et al*., 2004) and in understanding protein dynamics (Atilgan *et al*., 2004). It has further been shown that protein structures can be represented as graphs corresponding to small-world networks (Greene and Higman, 2003; Bagler and Sinha, 2005) and how residues contributed to protein-protein binding free energy in given complexes (Del Sol and O'Meara, 2005). These

networks are usually highly clustered with a few links connecting any pairs of nodes (Watts and Strogatz, 1998) and recently, there have been many reviews on protein contact networks (Greene, 2012; Hu *et al*., 2013; Khor, 2014).

For our study, we are interested in analyzing the G-protein coupled receptor. They are the integral membrane proteins characterized by seven membrane-spanning (transmembrane, TM) regions. They are involved with signal transduction across cell membranes. Many medically and pharmacologically important proteins are included in this super-family: e.g., Acetylcholine receptors, Dopamine receptors and Opioid receptors. The structure/function relationships for G-Protein Coupled Receptors (GPCR) are of vital importance (Wilson and Bergsma, 1999) as, they regulate a wide range of cellular processes, including the senses of taste, smell and vision and control a myriad of intracellular signaling systems in response to external stimuli. Many diseases are linked to GPCRs deficiency. Due to the difficulty in crystallizing GPCRs for X-ray crystallography the available X-ray crystallographic analyses on bovine rhodopsin are reviewed as the only available high-resolution structures for any GPCR.

Rhodopsin, a well known member of the G-protein-coupled receptor family, is located in the disk membranes of the outer segment of rod photoreceptor cell, where it is responsible for the visualization of dim light. Rhodopsin is the most extensively studied G-protein-coupled receptor, (Klabunde and Hessler, 2002; Sakmar, 2002) and knowledge about its structure serves as a template for other related receptors such, as light-sensitive compounds, odors, pheromones, hormones and neurotransmitters. A new high-resolution structure is reported for bovine rhodopsin. The new structure completely resolves the polypeptide chain and provides further details of chromophore binding site including the configuration about the C6-C7 single bond of the 11-cis retinal. The new X-ray structure has been improved to 2.2 Å (Okada *et al*., 2004).

To apply the small world concept to our protein bovine rhodopsin structure, we modeled our protein structure as an undirected graph, where amino acids are considered to be the nodes and the interaction between them as edges. Residues i and j are considered to be in contact if at least one atom corresponding to residue i is at a distance of ≤5.0 Å to an atom from residue j. This cut-off distance determines the upper limit for attractive London-van-der-Waals forces (Tinoco *et al*., 1995). The network parameters for the protein contact network were calculated and compared with the random controls. The closeness centrality values for each residue were computed and their statistical significance was determined using the z-score values.

MATERIALS AND METHODS

Network representation of protein structure: The crystal structure for Bovine Rhodopsin (PDB ID: 1U19) was obtained from the PDB database. Since, the X-ray structure contains two molecules of bovine rhodopsin, we have considered only the A-chain that comprises of 348 residues. The protein structures were then represented as an undirected graph, where amino acids/residues are denoted as the nodes and the interactions between them as links. Residues i and j are said to be in contact with each other when atoms between the two residues are at a distance of ≤5.0 Å. Distance between residues determine the intensity of London-Vander Waals forces (Atilgan *et al*., 2004). Random Control Networks are used to generate a given number of Nodes (N) and Edges (E) (Bollobas, 2001). They have a bell-shaped degree distribution indicating that the majority of nodes have degrees closer to the average degree <k>. Control random networks help us in identifying network variation by comparing the network parameters and analyzing differences between networks.

For the Protein Contact Network (PCN), 100 random controls were generated while maintaining the same number of nodes and edges. The connectivity distribution as well as the individual connectivity of the PCNs was also conserved. The Network parameters were calculated by using computer programs specifically written in Perl v5.10.1 and by using the Lenovo work station S20. The values obtained were then compared with that of the Protein Contact Network (PCN).

Network parameters

Degree of the network: The most elementary characteristic of a node is its degree (or connectivity), ki, which tells us how many links the node has to other nodes. The average degree K of the network with N nodes is computed as:

$$K = \frac{1}{N}\sum_{i=1}^{N} K_i$$

Degree distribution: The degree distribution, P (k), gives the probability that a selected node has exactly k links. The P (k) is obtained by counting the number o f nodes N (k) with k = 1, 2… links and dividing by the total number of nodes N. The degree distribution allows us to distinguish between different classes of networks. For example, By contrast, a power-law degree distribution indicates that a few hubs hold together numerous small nodes.

Shortest path length and average shortest path length: Distance in networks is measured with the path length, which tells us how many links we need to pass through to travel between two nodes as there are many alternative paths between two nodes, the shortest path is the path with the smallest number of links between the selected nodes. The mean path length represents the average over the shortest paths between all pairs of nodes and offers a measure of a network's overall navigability.

$$L = 1\frac{1}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} L_{ij}$$

where, Lij is the shortest path length between node i and j. We have used Hedetniemi's Algorithm (Arlinghaus *et al*., 1990) for calculating the shortest path distance between every pair of nodes in the network.

Clustering coefficient of the network: In many networks, if node A is connected to B and B is connected to C, then it is highly probable that A also has a direct link to C. This phenomenon can be quantified using the clustering coefficient $C_i = 2n_i/k(k-1)$, where n_i is the number of links connecting the k_i neighbors of node i to each other and where k (k-1)/2 gives the fraction of these possible links that actually exist (Watts and Strogatz, 1998). The average clustering coefficient of the network is calculated using:

$$C = n - 1/\sum_{o \neq k} d(i,k)$$

Where:

C_i = Clustering coefficient for a node i

These measures are 1 if every neighbours connected to n_i also connected to every other node with in the neighbourhood and 0 if no node connected to $n_{i,}$ connects to any other nodes that is connected to n_i.

Closeness centrality: Centrality is a network measure of nodal importance quantifying how prominent a node is relative to others (Wasserman and Faust, 1994). Closeness centrality (Ci) measures how close a node i is to all others in the same network (Jordan *et al*., 2006). The closeness centrality value C_k for residue k is defined as:

Where:
d (i, k) = Shortest path distance between node i and k
n = Number of nodes in the network

Statistical analysis: The statistically significant central residues were evaluated using the z-score values of the residue closeness centrality, defined as:

$$Z_k = Ck - \overline{C} / \sigma$$

Where:
C_k = Closeness centrality of residue k
C = Closeness centrality average value over all protein residues and σ is the corresponding standard deviation (Del Sol *et al*., 2006)

Determination of center of mass of a protein: To estimate the physical location of the residues, the centre of mass needs to be determined first. The center of mass of the protein whose position coordinates (x_{CM}, y_{CM}, z_{CM}) were calculated by:

$$x_{CM} = \frac{\sum_{i=1}^{N} m_i x_i}{\sum_{i=1}^{N} m_i}$$

$$y_{CM} = \frac{\sum_{i=1}^{N} m_i y_i}{\sum_{i=1}^{N} m_i}$$

$$Z_{CM} = \frac{\sum_{i=1}^{N} m_i z_i}{\sum_{i=1}^{N} m_i}$$

Where:
x_i,y_i,z_i = Cartesian coordinates of the i-th atom
m_i = Atomic mass

The distance of a particular residue j (whose position coordinates are assumed to be same as that of the C-α atom within it) from the protein center of mass is given by:

$$D_{CM} = \sqrt{\left(\left(x_j + x_{CM}\right)^2 + \left(y_j - y_{CM}\right)^2 + \left(Z_j - z_{CM}\right)^2\right)}$$

Functionally important residues: The conservation score for each amino acid residue in a protein is obtained using the *Consurf* server (http://consurf.tau.ac.il), which is a relative measure of the evolutionary conservation at each sequence site of the target protein with the lowest score representing the most conserved position. It uses ClustalW Multiple Sequence Alignment for calculating the scores of all residues and then performs a normalization to make the mean score = 0 with standard deviation = 1 (Glaser *et al*., 2003; Landau *et al*., 2005). From the PDBsum database (Laskowski *et al*., 1997) the ligand binding residues were obtained.

RESULTS AND DISCUSSION

Visualization of the networks: The constructed protein contact network consists of 348 nodes and 1930 edges. Figure 1a represents a circular layout in which the amino acids are arranged sequentially and the interactions between the residues are represented as lines. The circular layout places graph nodes in such a manner that they connect to many other nodes thus increasing the graph density. This helps you find out in an intuitive way, the critical nodes of your graph. Figure 1b represents the Kamada-Kawai graph layout (Kamada and Kawai, 1989), which is an algorithm for drawing undirected graph. The layout attempts to position nodes on the space so, that the geometric (Euclidean) distance between them is similar to that of the theoretical distance.

Out of 100 control random network generated, we have represented one of them. These networks are found to have the same number of nodes and the edges but the connections between nodes are randomized while maintaining the degree distribution in the protein contact network. Figure 2a represents the circular layout and Fig. 2b represents the kamada-Kawai graph layout of the control random network.

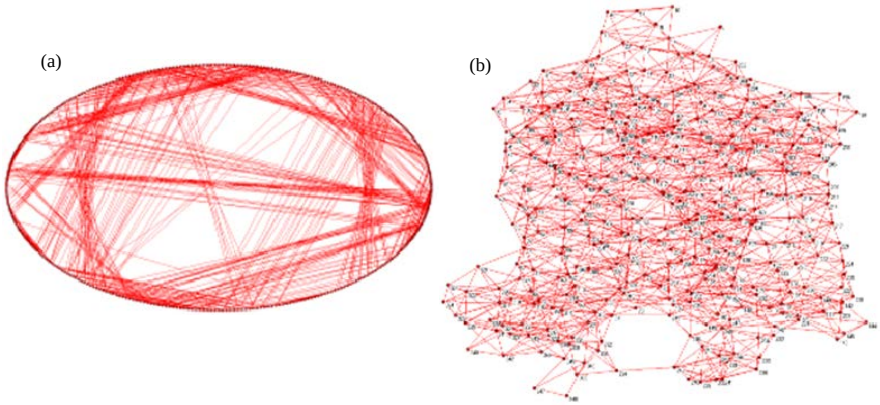

Fig. 1(a-b): Protein contact network (a) Circular layout and (b) Kamada-Kawai graph layout

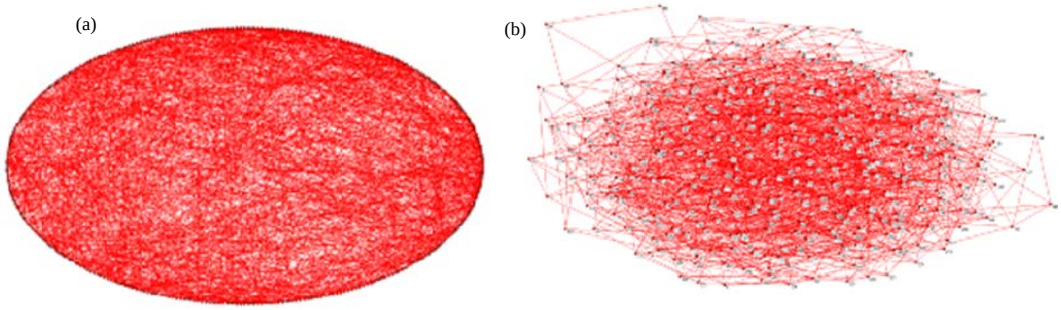

Fig. 2(a-b): Control random network, (a) Circular layout and (b) Kamada-Kawai graph layout

Degree distribution: From Fig. 3, the average degree of the protein contact network was found to be 11.09 and the mean value of the 100 control random network's average degree was also found to be 11.09. This indicates that the degree distribution of the protein contact network was maintained in their random counter parts. A peaked degree distribution is seen in control random networks indicating that the system has a characteristic degree and that there are no highly connected nodes (which are also known as hubs). By contrast the shape of the protein network distribution obtained is also found to be Gaussian (Greene and Higman, 2003).

Shortest path distribution: Figure 4a represents the shortest path distribution for protein contact network and Fig. 4b represents the distribution of the mean shortest path lengths of the control random networks. The shortest path distribution is the probability distribution of the short path distances in the network. The average shortest path length for the PCN is found to be 4.78 and the mean value of 100 shortest path lengths for the control random networks is found to be networks. The standard deviation for the 100 control random networks is calculated to be 2.71±0.003. Atilgan *et al.* (2004) have analyzed numerous globular proteins and they shown that

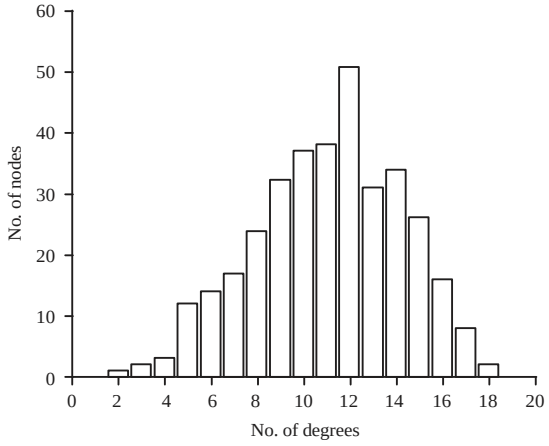

Fig. 3: Represents the degree distribution for both PCN and the mean degree distribution of control random networks

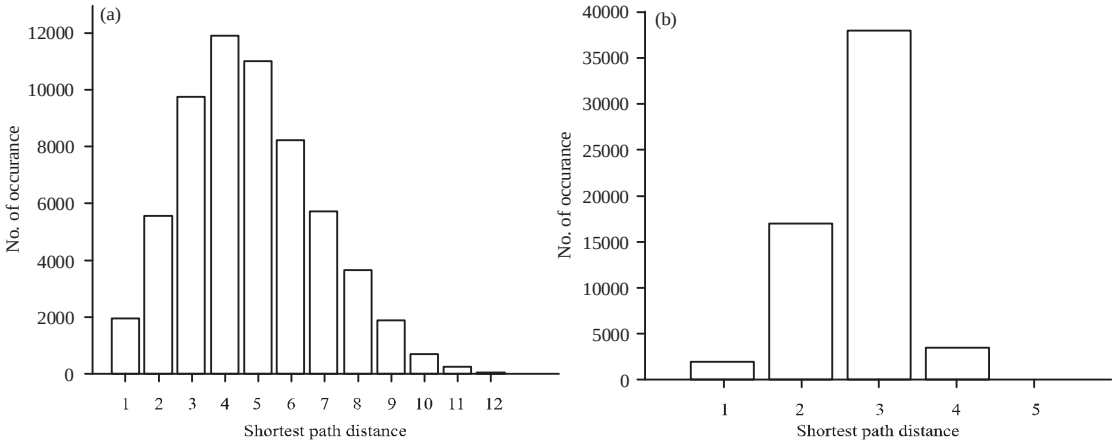

Fig. 4(a-b): (a) Shortest path length distribution for PCN and (b) Mean shortest path length distribution for control random networks

residue fluctuations are highly correlated with the shortest path lengths. The highly connected residues, which are in turn connected to the rest of the molecule, on average, in a shorter number of steps (Atilgan *et al*., 2004). Similarly, the membrane protein rhodopsin also have a shorter number of steps in path lengths.

Clustering coefficient distribution: The clustering coefficient is an important domain for each graph vertex, it ranges from 0 and 1. Figure 5 represents the clustering coefficient distributions for PCN and random control networks. The protein contact network has an average clustering coefficient value of 0.51 and the 100 control random networks have an average of 0.03. The standard deviation for the 100 control random networks is found to be 0.03±0.002. Our findings show that the protein contact network is highly clustered than the control random network. Similarly, study done by Yan *et al*. (2014) shown that the network clustering coefficient had a significantly positive and negative correlation with the protein secondary structure density (Yan *et al*., 2014). Thus higher the coefficient value, better the secondary structure density in the protein structures. In this case, the rhodopsin protein contact network is found to have a higher value of clustering coefficient in comparison with the random network.

L-C plot: The mean shortest path length distribution L and the mean clustering coefficient distribution C for the protein contact networks is found to be 4.78 and 0.51. For the control random network the L and C values are found to be 2.71 and 0.03. The L-C values for PCN is highly significant than the random control network. The standard deviation for the L and C for the 100 random control networks is found to be 2.71±0.003 and 0.03±0.002. Watts and Strogatz (1998) characterized the small world networks with two parameters namely the average shortest path length L and the clustering coefficient C. Regular networks have large L and large C values than random networks. Small-world networks have small L and large C values (Watts and Strogatz, 1998). In this study, the L and C values for the rhodopsin PCN were found to be higher in comparison with the random control networks. Suggesting that the structures of rhodopsin proteins can be conveniently analyzed by using the small-world networks approach.

Closeness centrality: The closeness centrality for each residue in the protein contact network was calculated and then the statistical significance of the central residues was determined using the

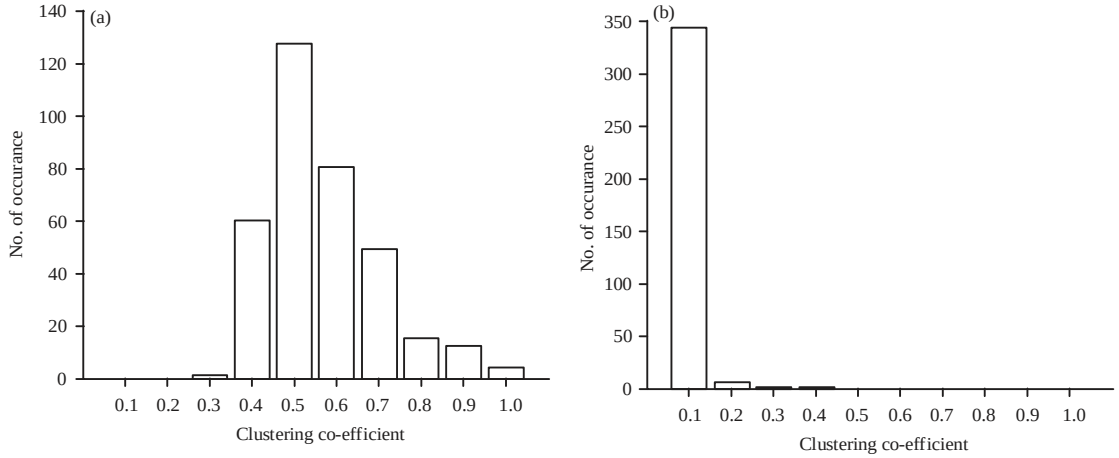

Fig. 5(a-b): (a) Clustering coefficient distribution for the PCN and (b) Mean distribution of the control random networks

z-score values. Figure 6 shows the plot of z-score values with the distance from the Center Of Mass (COM) of the protein structure of 348 amino acids. The correlation coefficient between the z-score values and distance from COM, is -0.9644 ($p<0.0001$). This suggests that z-score values have a negative correlation with the distances from the COM. Our findings indicate that statistically significant central residues are closer to the COM of the protein structures. Study done by Amitai *et al.* (2004) have shown that the residues with high closeness centrality value are present in the active site, ligand-binding and evolutionary conserved (Amitai *et al*., 2004).

To understand the significance of the residues belonging to the high z-score values, we analyze their degree of conservation (that measures its rate of evolutionary change) as a function of the z-score intervals (from lower to higher). As the changes in different positions in a protein are not homogeneous but rather differ significantly, with some residues mutating rapidly (called "variable" positions) relative to others (termed as "conserved" positions), we are thus interested to understand if residues belonging to the high Z-score values are more conserved than those belonging to the low z-score values.

As shown in Fig. 7, conservation plot for each Z-score interval are plotted against the percentage of amino acids. Highly conserved residues (conservation score 9) are more numerous in the high Z-score intervals (1-2). On the contrary, highly variable residues (conservation score 1) are less in

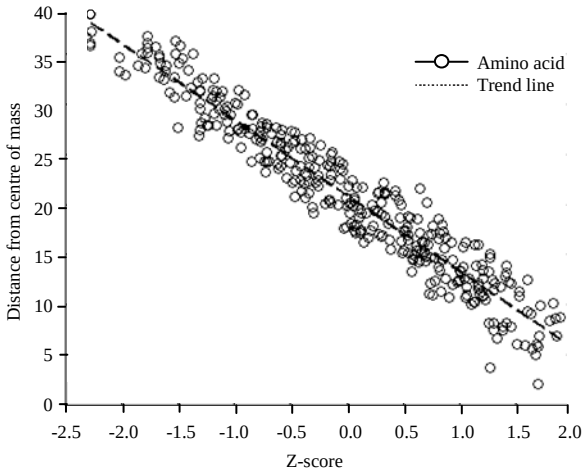

Fig. 6: Plot of Z-score values with distance from COM (Center of mass)

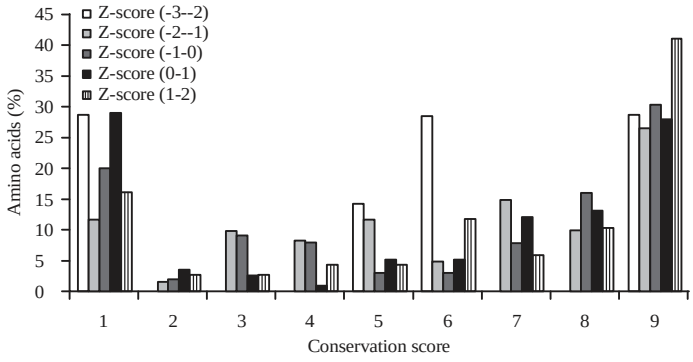

Fig. 7: Histogram representation of conservation scores in z-score intervals

Table 1: List of amino acids having z-score>1.5 with conservation score and distance from the center of mass of the protein structure

Amino acid names	Chain	Amino acid no.	Z-score (>1.5)	Conservation score (1-9)	Distance from center of mass (Å)
Leu	A	79	1.63	8	9.07
Ala	A	117*	1.64	9	5.51
Thr	A	118*	1.51	9	6.01
Gly	A	121*	1.69	9	2.09
Glu	A	122*	1.69	2	5.83
Ile	A	123	1.70	5	6.83
Ala	A	124	1.58	1	5.91
Ser	A	127	1.53	7	10.83
Met	A	257	1.60	9	12.48
Ala	A	260	1.53	1	11.30
Phe	A	261*	1.85	3	8.55
Cys	A	264	1.89	6	8.77
Trp	A	265*	1.80	9	8.41
Lys	A	296*	1.86	9	6.83
Ser	A	298	1.67	6	5.02
Ala	A	299	1.68	1	6.06
Val	A	300	1.60	1	9.30
Tyr	A	301	1.83	4	10.17
Asn	A	302	1.72	9	9.91

*Amino acids are in contact with ligand molecule

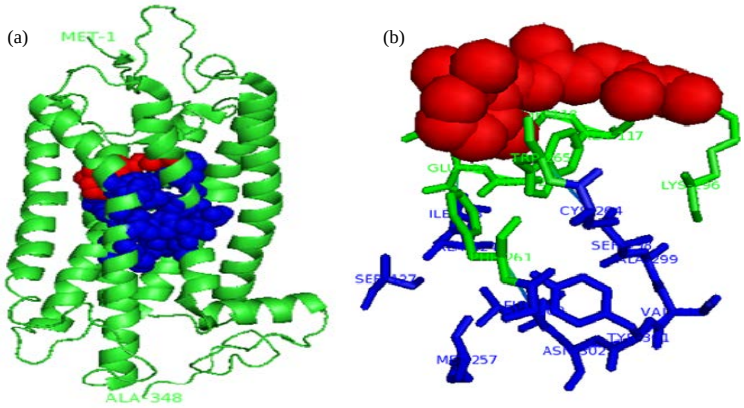

Fig. 8(a-b): (a) Structure of the bovine rhodopsin (PDB code: 1u19A), predicted central residues are represented in blue color and the ligand in red and (b) Amino acids are in contact with the ligand (RETINAL) and are depicted in green color

number in the low z-score interval (-3 to -1). Our findings confirm that the residues falling in the range of high z-score values are highly conserved.

From Table 1, it is seen that centrally conserved residues tend to be buried in the structure. Statically significant residues are closer to the center of the mass and are also highly conserved through evolution. Amino acids with high z scores are also seen to be in close proximity with the ligand molecule: Different approaches have been proposed based on the structural features for identifying the active sites in various proteins (Lichtarge *et al*., 1996; Aloy *et al*., 2001; Landgraf *et al*., 2001; Ondrechen *et al*., 2001). It was further shown that active site residues tend to have high centrality values for most of the enzymes (Amitai *et al*. 2004). Thus, we may understand that closeness centrality is a global topological characteristic and it provides more information than just a local analysis of residue centrality (Vendruscolo *et al*., 2001). High closeness residues tend to be clustered around those cavities containing functionally important residues.

The bovine rhodopsin structure is represented as a ribbon model using PyMOL (0.99rc6) software as show in Fig. 8a. The extracellular n-terminal amino acid methionine (1) and the

cytoplasmic or intracellular amino acid alanine (348) are represented. Residues with high z-score values (>1.5) are depicted in blue and the retinal ligand molecule in red. It is found that residues with high statistical significance are buried in the structure and are also closer to the center of mass of the protein. In Fig. 8b, the predicted central residues are represented as a stick model with the retinal ligand represented as a spherical shape. The ligand interacting residues are shown in green color. From this we understand that, predicted central residues are located in the binding pockets of the retinal ligand molecule and are also seen to be highly conserved through evolution.

CONCLUSION

In this study, we have applied network principles to the native protein structure of the Bovine rhodopsin (PDB ID: 1U19) belonging to the family of G-protein coupled receptor. Our results show that the seven trans-membrane helices of this protein can be visualized with signature patterns in the contact map of the network. The rhodopsin protein contact network has a clustering coefficient value C = 0.51, which is significantly higher than the average of 100 control random network and the average shortest path length L = 4.78 for the PCN and this is again seen to be higher in comparison to the random control networks. This shows that the protein contact network possesses the "Small World Network" property. We also report that the residues with high closeness centrality values tend to be clustered together with functionally important amino acids situated in the protein cavities. The results shows that the residues with high z-score values are closer to the center of mass of the protein are closer to the ligand molecule and also contain more percentages of conserved amino acids.

ACKNOWLEDGMENT

The authors express their sincere thanks to the VIT University for providing the necessary computational facilities needed for the study without which much would have been impossible.

REFERENCES

Aloy, P., E. Querol, F.X. Aviles and M.J.E. Sternberg, 2001. Automated structure-based prediction of functional sites in proteins: Applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. J. Mol. Biol., 311: 395-408.

Amitai, G., A. Shemesh, E. Sitbon, M. Shklar, D. Netanely, I. Venger and S. Pietrokovski, 2004. Network analysis of protein structures identifies functional residues. J. Mol. Biol., 344: 1135-1146.

Arlinghaus, S.L., W.C. Arlinghaus and J.D. Nystuen, 1990. The hedetniemi matrix sum-an algorithm for shortest-path and shortest distance. Geograph. Anal., 22: 351-360.

Atilgan, A.R., P. Akan and C. Baysal, 2004. Small-world communication of residues and significance for protein dynamics. Biophys. J., 86: 85-91.

Bagler, G. and S. Sinha, 2005. Network properties of protein structures. Phys. A: Stat. Mech. Applic., 346: 27-33.

Bollobas, B., 2001. Random Graphs. Academic Press, London, UK.

Del Sol, A. and P. O'Meara, 2005. Small-world network approach to identify key residues in protein-protein interaction. Proteins: Struct. Funct. Bioinform., 58: 672-682.

Del Sol, A., H. Fujihashi, D. Amoros and R. Nussinov, 2006. Residue centrality, functionally important residues and active site shape: Analysis of enzyme and non-enzyme families. Protein Sci., 15: 2120-2128.

Glaser, F., T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz and N. Ben-Tal, 2003. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics, 19: 163-164.

Greene, L.H. and V.A. Higman, 2003. Uncovering network systems within protein structures. J. Mol. Biol., 334: 781-791.

Greene, L.H., 2012. Protein structure networks. Briefings Funct. Genomics, 11: 469-478.

Heringa, J. and P. Argos, 1991. Side-chain clusters in protein structures and their role in protein folding. J. Mol. Biol., 220: 151-171.

Holme, P., M. Huss and H. Jeong, 2003. Subnetwork hierarchies of biochemical pathways. Bioinformatics, 19: 532-538.

Hu, G., J. Zhou, W. Yan, J. Chen and B. Shen, 2013. The topology and dynamics of protein complexes: Insights from intra-molecular network theory. Curr. Protein Peptide Sci., 14: 121-132.

Jeong, H., S.P. Mason, A.L. Barabasi and Z.N. Oltvai, 2001. Lethality and centrality in protein networks. Nature, 411: 41-42.

Jordan, F., W.C. Liu and A.J. Davis, 2006. Topological keystone species: Measures of positional importance in food webs. Oikos, 112: 535-546.

Kamada, T. and S. Kawai, 1989. An algorithm for drawing general undirected graphs. Inform. Process. Lett., 31: 7-15.

Khor, S., 2014. Protein Residue Networks from a Local Search Perspective. Cornell University Library, Ithaca, NY., USA., Pages: 69.

Klabunde, T. and G. Hessler, 2002. Drug design strategies for targeting G-protein-coupled receptors. Chembiochem, 3: 928-944.

Landau, M., I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, 2005. ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res., 33: W299-W302.

Landgraf, R., I. Xenarios and D. Eisenberg, 2001. Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. J. Mol. Biol., 37: 1487-1502.

Laskowski, R.A., E.G. Hutchinson, A.D. Michie, A.C. Wallace, M.L. Jones and J.M. Thornton, 1997. PDBsum: A web-based database of summaries and analyses of all PDB structures. Trends Biochem. Sci., 22: 488-490.

Lichtarge, O., H.R. Bourne and F.E. Cohen, 1996. An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol., 257: 342-358.

Okada, T., M. Sugihara, A.N. Bondar, M. Elstner, P. Entel and V. Buss, 2004. The retinal conformation and its environment in rhodopsin in light of a new 2.2 A crystal structure. J. Mol. Biol., 342: 571-583.

Ondrechen, M.J., J.G. Clifton and D. Ringe, 2001. THEMATICS: A simple computational predictor of enzyme function from structure. Proc. Natl. Acad. Sci., 98: 12473-12478.

Ravasz, E., A.L. Somera, D.A. Mongru, Z.N. Oltvai and A.L. Barabasi, 2002. Hierarchical organization of modularity in metabolic networks. Science, 297: 1551-1555.

Sakmar, T.P., 2002. Structure of rhodopsin and the superfamily of seven-helical receptors: The same and not the same. Curr. Opin. Cell Biol., 14: 189-195.

Shen-Orr, S.S., R. Milo, S. Mangan and U. Alon, 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nature Genet., 31: 64-68.

Tinoco, I., K. Sauer and J.C. Wang, 1995. Physical Chemistry: Principles and Applications in Biological Sciences. Prentice-Hall, London, UK., ISBN-13: 978-0131865457, Pages: 736.

Vendruscolo, M., N.V. Dokholyan, E. Paci and M. Karplus, 2002. Small-world view of the amino acids that play a key role in protein folding. Phys. Rev. E, Vol. 65.

Wasserman, S. and K. Faust, 1994. Social Network Analysis: Methods and applications. Cambridge University Press, Cambridge, UK., ISBN-13: 9780521387071, Pages: 857.

Watts, D.J. and S.H. Strogatz, 1998. Collective dynamics of 'small-world' networks. Nature, 393: 440-442.

Wilson, S. and D. Bergsma, 1999. Orphan G-protein coupled receptors: Novel drug targets for the pharmaceutical industry. Drug Design Discov., 17: 105-114.

Wolf, Y.I., G. Karev and E.V. Koonin, 2002. Scale-free networks in biology: New insights into the fundamentals of evolution? Bioessays, 24: 105-109.

Yan, W., M. Sun, G. Hu, J. Zhou and W. Zhang *et al*., 2014. Amino acid contact energy networks impact protein structure and evolution. J. Theoret. Biol., 355: 95-104.