Research Article

IJCRR
Section: General Science
Sci. Journal
Impact Factor 4.016
ICV: 71.54

Identifying Structurally Similar/ Dissimilar Proteins Through Graph Similarity

R. Mageswari[1], B. Yaminipriya[2]

[1]Assistant Professor, Department of Mathematics, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University (SCSVMV University), Kanchipuram - 631 561, Tamilnadu, India; [2]Scholar, Department of Mathematics, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University (SCSVMV University), Kanchipuram - 631 561, Tamilnadu, India.

ABSTRACT

Proteins having similar three dimensional structures have similar functions. It is proposed to find the structurally similar proteins through graph similarity. Graph theoretical concepts are used to analyze protein similarity using contact maps. Protein graph is derived from its contact map. In this paper, clustering coefficient of nodes is used to find the structurally similar/dissimilar proteins. High positive correlation is observed between the clustering coefficient of similar proteins 2RM2 and 2RR1 and no correlation between dissimilar proteins 2RM2, 5JXI and 2RR1, 5JXL.

Key Words: Contact map, Carbon alpha, Clustering Coefficient

INTRODUCTION

An amino acid is a simple organic compound containing a carboxyl (-COOH) functional group and an amino (-NH2) group along with a side chain (R group) specific to each amino acid. The elements of an amino acid are carbon, hydrogen, oxygen and nitrogen though other elements are found in the side-chains of certain amino acids. There are twenty different amino acids which are the building blocks of protein. The chain itself represents the protein molecule. Proteins are lengthy chains of amino acids. Protein chains are then warped and pleated together in specific ways to create certain molecules. Structure of a protein determines its biological function. Graph of the protein is obtained from its contact map. It is proposed to identify structurally similar/ dissimilar proteins through graph similarity from the clustering coefficient of each node. Similarity and dissimilarity of protein sequences is found from the geometric properties of secondary structure elements [1], using the spatial medians between the Euclidean distances between the proteins under study [2], by the compression ratio of the concatenated image using the image and audio compression – based approach [3], from the matching pairs of secondary structure elements by the interaction of each pair between their axial line segments using a fast bipartite graph matching algorithm [6], by comparing the centroids of all secondary structures using the largest common sub graph detection algorithm [7], from the maximum common edge sub graph of the labeled graph of the protein [8], Protein structure is identified using homology modeling [9]. Similarity between contact maps can also be studied through content Based Image Retrieval (CBIR) and image registration techniques [10]. Similarity has been measured by the Contact map overlap based on the pair wise distances of the Cα – atoms of each protein [11], using the maximum clique algorithm PLS (Phased Local Search) [12], using the bipartite graph matching algorithm [13], by aligning the proteins to maximize the number of shared contacts in their corresponding contact maps [14], A Clustering Coefficient [15] is a measure of the degree to which nodes in a graph tend to cluster together. In this paper, similarity & dissimilarity of the protein structure is studied through the Clustering coefficient of each node in the protein graph.

PRELIMINARIES

Alpha carbon

The alpha carbon refers to the first carbon that attaches to a functional group (the carbon is attached at the first, or alpha, position), the second carbon is the beta carbon, and so on.

Corresponding Author:
R. Mageswari, Assistant Professor, Department of Mathematics, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University (SCSVMV University), Kanchipuram - 631 561, Tamilnadu, India.

ISSN: 2231-2196 (Print) **ISSN:** 0975-5241 (Online)

Received: 18.03.2017 **Revised:** 14.04.2017 **Accepted:** 11.05.2017

Contact Maps

The contact map of a protein is given in the form of a matrix of order n for a protein with n atoms. ie, T = $(t_{ij})1 \leq i,j \leq n$,

where $t_{ij} = 1$ if $i \neq j$ & $d_{ij} \leq 6$Å

= 0 otherwise.

A protein can be considered as a graph with residues as vertices and edges between the vertices i and j if $t_{ij} = 1$.

Clustering Coefficient

Clustering coefficient of the node v is computed as follows:

$$C_v = \frac{2\,N_v}{k_v\,(k_v - 1)}$$

Where, k_v is denotes the degree of the v-th node, N_v denotes number of links between neighbors of v

Protein Graph from its Contact Map

A part of the three dimensional structures of the proteins 2RM2, 2RR1 and 5JXL and its ball and stick representation of alpha carbons are shown in the figure 1 and 2 respectively. The contact maps of the proteins 2RM2, 2RR1 and 5JXL are found by using the definition stated above and the graphs of the respective proteins are obtained from the contact maps as shown in figure 3,4 and 5.

From the contact map of 2RM2, it is clear that v_1 is connected to the vertices 2 and 3. Therefore, the degree of vertex v_1 is 2 and it is denoted by k_{v_1}. Total number of links between the neighboring vertices of v_1 is 1.Therefore, Clustering coefficient is calculated as 1.The remaining Clustering coefficients are calculated in the similar way and tabulated in Table 1.

CONCLUSION

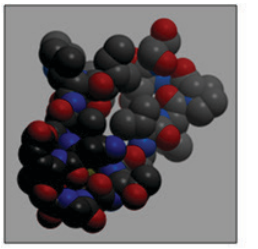
The two proteins 2RM2 and 2RR1 are exactly similar in structure since the coefficient of correlation between the clustering coefficients of 2RM2 and 2RR1 is calculated as1 and the two proteins 2RR1 and 5JXL are dissimilar in structure since the correlation between the clustering coefficients of 5JXL and 2RR1 is calculated is 0.37.

REFERENCES

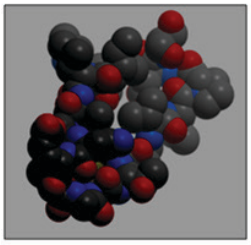
1. Pooja Jain and Jonathan D.Hirst, Study of Protein Structural Descriptors Towards Similarity & Classification, Computational Biophysics to Systems Biology, 2007, Vol. 36, pp:165-167.
2. Abo-Elkhier, Mervat M, Similarity /dissimilarity analysis of protein sequences using the spatial median as a descriptor, Journal of Biophysical Chemistry, 2012, Vol. 3, pp:142-148.
3. Morihiro Hayashida and Tatsuya Akutsu, Image compression – Based approach to measuring the similarity of protein structures, WSPC- Proceedings October 2007 pp:17.
4. Pankaj Barah & Som Data Sinha, Analysis of Protein folds using protein contact networks, Pramana Journal of physics, August 2008, Vol. 71, pp: 2.
5. Mahnaz Habibi, ChangizEslahchi, Mehdi Sadeghi , Hamid Pezashk, The interpretation of protein structures based on graph theory and contact map, Open Access Bioinformatics 2010, pp: 127-137.
6. William R. Taylor, Protein structure comparison using Bipartite Graph Matching and its application to protein structure classification, Molecular & Cellular Proteomics, 2002, 1(4) pp. 334-339.
7. Stoicho Stoichev, Debrinka Petrova, Protein Structure Models for Determining Protein Structure Similarity, Comp. Sys. Tech'06.
8. Cheng – Hsien Hsu, Sheng-Lung Peng and Yu-Wei Tsay, An improved Algorithm for Protein Structural Comparison Based on Graph Theoretical Approach, Journal of Science, 2011, pp: 71-81.
9. Yan Yan, Shenggui Zhang, Fang-Xiang Wu, Applications of graph theory in protein structure identification, Proteome Science 2011, 9.
10. Fernando Fernandes JR, Carlos Eduardo Lopes, Raquel Melo, Marcelo Santoro, Rodrigo Carceroni, Wagner Meira JR, Arnaldo Araujo, An Image –Matching Approach to Protein Similarity Analysis, Computer Graphics and Image Processing, 2004. Proceedings pp: 17-24.
11. Pankajk Agarwal, NabicH Mustafa and Yusu Wang, Fast Molecular Shape Matching Using Contact Maps, Journal of Computational Biology, Vol. 14, 2007, pp: 131-143.
12. Wayne Pullan, Protein Structure Alignment Using Maximum Cliques and Local Search, M.A Orgun and J. Thornton(Eds): AI 2007, 4830, pp: 776-780.
13. Rosni Abdullah, Nu' Aihi Abdul Rashid & Fazilah Othman, Graph Theory in Protein Sequence Clustering and Tertiary Structural Matching, AIP Conference Proceedings, Vol. 971, pp:19.
14. B. Carr, W.Hart, E. Burke J. Smith, Alignment of protein Structures with a Memetic Evolutionary Algorithm, Proceedings of the Genetic and Evolutionary Computation Conference, 2002.
15. W. Gruszczynski and P. Arabas, Application of social network to improve effectiveness of classifiers churnmodelling'', in proc. 3rdInt. Conf. Comput. Aspect of social netw.CASoN'11, Salamanca Spain, 201

Table 1:

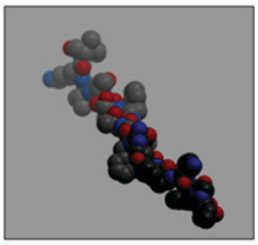
S.NO	C_V	2RM2	2RR1	5JXL
1	C_{V_1}	1	1	1
2	C_{V_2}	0.67	0.67	1
3	C_{V_3}	0.67	0.67	0.33
4	C_{V_4}	0.33	0.33	0
5	C_{V_5}	0	0	0
6	C_{V_6}	0.33	0.33	0
7	C_{V_7}	0.67	0.67	0
8	C_{V_8}	0.67	0.67	0
9	C_{V_9}	0.67	0.67	0
10	$C_{V_{10}}$	0.67	0.67	0
11	$C_{V_{11}}$	0.33	0.33	0
12	$C_{V_{12}}$	1	1	0
13	$C_{V_{13}}$	0.33	0.33	0
14	$C_{V_{14}}$	1	1	0
15	$C_{V_{15}}$	0.33	0.33	0
16	$C_{V_{16}}$	0.5	0.5	0
17	$C_{V_{17}}$	0.83	0.83	0.33
18	$C_{V_{18}}$	0.83	0.83	1
19	$C_{V_{19}}$	0.83	0.83	0.33
20	$C_{V_{20}}$	1	1	0



2RM2

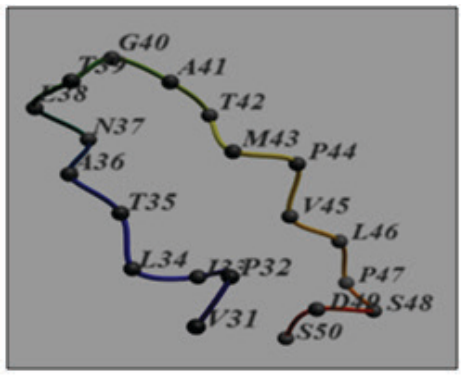


2RR1

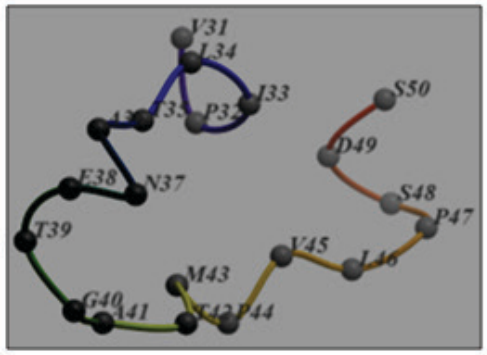


5JXL

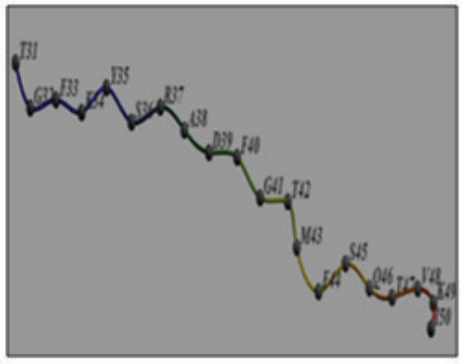
Figure 1: A Part (residue 31 – residue 50) of the 3D structure of the proteins.



(a)



(b)



(c)

Figure 2: Ball and stick representation of carbon-alpha (C_α) of (a) 2RM2 (b) 2RR1 and (c) 5JXL.

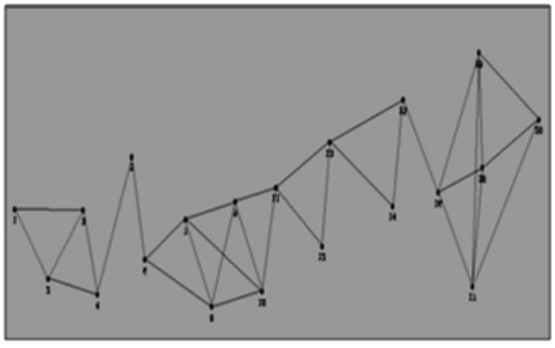


Figure 3: Graph of the protein 2RM2.

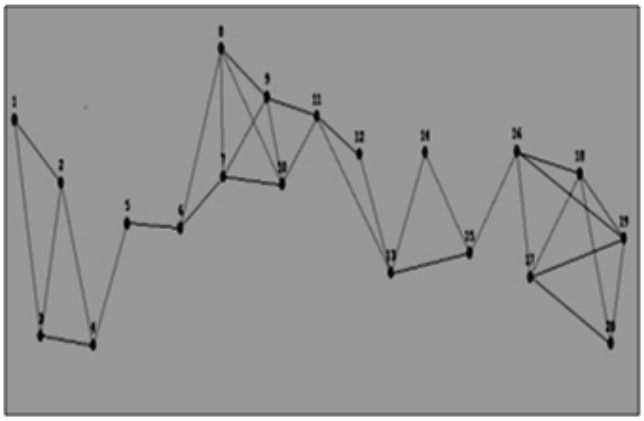


Figure 4: Graph of the protein 2RR1.

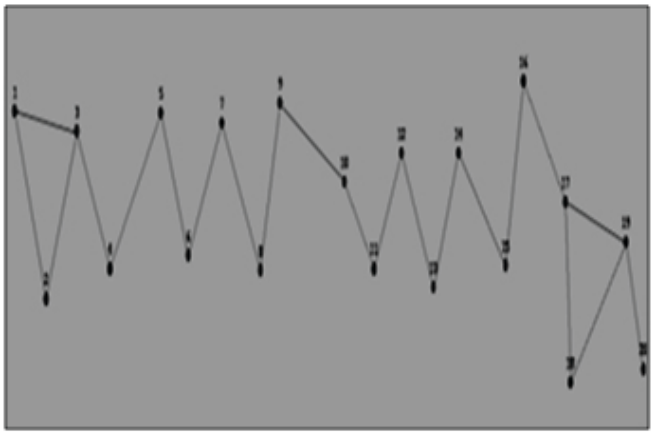


Figure 5: Graph of the protein 5JXL.