Analysing Amino Acids in Galanin

Graph Theoretical Approach

Suresh Singh G.
Department of Mathematics
University of Kerala, Kariavattom,
Thiruvananthapuram – 695581, Kerala, India,
E-mail: *sureshsinghg@yahoo.co.in*

Akhil C. K.
Department of Mathematics
University of Kerala, Kariavattom,
Thiruvananthapuram – 695581, Kerala, India,
E-mail: *redakhil87@gmail.com*

Abstract— Graph theoretical analysis is an important area of the research in biological networks. Here first we introduce Pt-graph of peptide/protein based on physicochemical properties and adjacency of amino acids in the corresponding peptide/protein. Based on the Pt-graph, we introduce the graph of species which containing the peptide/protein named as SPt-graph. Finally, we analyze graph theoretically Pt-graphs of fourteen species of animals containing Galanin, a neuropeptide and their SPt-graph. From the graph theoretical analyses of Pt-graph and SPt-graph we get some observations about the relations among the amino acids, physicochemical properties of amino acids, peptide/protein and species containing peptide/protein and it may help in the field of evolution of peptide/protein and drug design in future.

Keywords*- Amino acid, Galanin, Graph, Pt-graph, SPt-graph, biological networks.*

*****

I. INTRODUCTION

Proteins [3] are the most abundant biological macromolecules, occurring in all cells and all parts of cells. Their building blocks are called amino acids and they are organic compounds which contain at least one amino group (-NH_2) and a carboxyl (-COOH) group. They have some physicochemical properties [1] (Hydrophobicity, Hydrophilicity, Polarity, Non-polarity, Aliphaticity, Aromaticity and Charge (Positive and Negative)). The sequence of amino acids [3] in a protein is characteristic of that protein and is called its primary structure. Peptides/proteins are the compounds of amino acids in which a carboxyl group of one is united with an amino group of another. Neuropeptides are peptides formed and released by neurons.They are often localized in axon terminals at synapses and are classified as putative neurotransmitters, although some are also hormones.

Galanin [6] is a peptide widely distributed in the central and peripheral nervous systems and the endocrine system. It was first identified from porcine intestinal extracts in 1978 by Professor Viktor Mutt and colleagues at the Karolinska Institute, Sweden. It is a peptide consisting of a chain of 29 amino acids (30 amino acids in humans). Galanin is so-called because it contains an N-terminal glycine residue and a C-terminal alanine.

Graph theory [5] has so many applications in biology. Amino acid network [4] with in protein was studied by S. Kundu. Graph theoretic approach to analyze amino acid network [1] was studied by Adil Akthar and Nisha Gohan. Centralities in amino acid networks [2] were used by Adil Akhtar and Tazid Ali. By using the concept of amino acid network we define new graphs called Pt-graph and SPt-graph of peptide/protein. Here we analyze graph theoretically Pt-graphs of fourteen species of animals containing Galanin, a neuropeptide and their SPt-graph. Finally, we get some observations about the relations among the amino acids, physicochemical properties of amino acids, peptide/protein and species containing peptide/protein and it may help in the field of evolution of peptide/proteins and drug design in future.

II. BASIC CONCEPTS OF GRAPH THEORY

A Graph [5] G is a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a finite set V of vertices and a set $\mathcal{E}$ of 2-element subsets of $\mathcal{V}$. The elements of $\mathcal{V}$ are called vertices [5] and elements of $\mathcal{E}$ are called edges [5]. The set $\mathcal{V}$ is known as the vertex set of $\mathcal{G}$ and $\mathcal{E}$ as the edge set of $\mathcal{G}$. Two vertices u and v of $\mathcal{G}$ are said to be adjacent [5], if an edge join u and v, and two edges are adjacent if they have common vertex.

2.1. Centralities in Graphs [2]. In graph theory, centrality measure of a vertex represents its relative importance within the graph. A centrality is a real-valued function on the vertices of a graph. More formally a centrality is a function f which assigns every vertex $v \in V$ of a given graph G a value $f(v) \in \mathbb{R}$. In the following we have discussed four most commonly used centrality measures.

2.1.1. *Degree of Centrality [2]*. The most simple centrality measure is degree of centrality, $c_d(u)$. It is defined as the number of vertices to which the vertex u is directly connected. The vertices directly connected to a given vertex u are also called first neighbours of the given vertex. Degree centrality shows that an important vertex is involved in a large number of interactions. The interaction gives the immediate importance or risk of the vertex in the corresponding network. Mathematically it is defined as

$$c_d(u) = \deg(u). \qquad (1)$$

However in real world problem the degree of centrality is not an actual measurement for finding important vertex may be connected indirectly with other vertices.

2.1.2. *Eigenvector Centrality [2]*. Another important measure of centrality is eigenvector centrality. An eigenvalue of a square matrix A is a value λ for which $det(A - \lambda I) = 0$, where I is the identity matrix of the same order as A. Eigenvector centrality is defined as the principal eigenvector of the adjacency matrix of corresponding graph. In matrix-vector notation we can write

$$\lambda X = AX, \tag{2}$$

where A is the adjacency matrix of the graph, λ is a constant (the eigenvalue), and X is the eigenvector. In general, there will be different eigenvalues λ for which an eigenvector solution exists. However eigenvector of the greatest eigenvalue is the eigenvector centrality. Eigenvector centrality gives the direct as well as indirect importance of a vertex in a network.

2.1.3. *Closeness Centrality [2]*. The closeness centrality is the idea how a vertex is close to all other vertices not only to the first neighbor but also in global scale. Generally a vertex is central; then it is close to all other vertices. If a vertex is close to other vertices, then it can quickly interact with all other vertices. In general closeness centrality is defined as the inverse of the sum of the shortest path distances between each vertex and every other vertex in the network. The closeness centrality of a vertex depicts an important vertex that can easily reach or communicate with other vertex of the network. Mathematically it is defined as

$$C_{cl}(u) = \frac{(n-1)}{\sum_{v\in V} d(u,v)}, \tag{3}$$

where n is the number of vertices of the network and $d(u, v)$ is the shortest path distance between the pair of vertices u and v. From the above definition it is clear that if a vertex has minimum cumulative shortest path distance, then that vertex has maximum closeness centrality. And maximum closeness centrality vertex is very well connected to all other verties.

2.1.4. *Betweenness Centrality [2]*. Another well-known centrality measure is the betweenness centrality. Betweenness centrality interactions between two nonadjacent vertices depend on the other vertex, generally on those on the paths between the two. The betweenness centrality of a vertex u is the number of shortest paths going through u. Mathematically it is defined as

$$C_{btw}(u) = \sum_{s\neq u\in V} \sum_{t\neq u\in V} \frac{\sigma_{st}(u)}{\sigma_{st}}, \tag{4}$$

where σ_{st} is the number of shortest paths from vertex s to t and $\sigma_{st}(u)$ is the number of shortest paths from s to t that pass through u. Betweenness centrality depicts identifying vertices that make most information flow of the network. An important vertex will lie on a large number of paths between other vertices in the network. From this vertex we can control the information of the network. Without these vertices, there would be no way for two neighbors to communicate with each other. In general the high degree vertex has high betweenness centrality because many of the shortest paths may pass through that vertex. However a high betweenness centrality vertex need not always be high degree vertex.

III. Pt-GRAPH AND SPt-GRAPH OF PEPTIDES/ PROTEINS

In the amino acid network [1], a vertex set was defined as collection of all twenty natural amino acids and the edge set is defined as the connection between two amino acids having at least one (two in another case) common properties. Using the concepts of amino acid networks we define a new graph called Pt-graph of peptide/protein.

Definition 3.1: A Pt-graph is defined as a graph $\mathcal{G} = (V, E)$ of a peptide/protein in which the vertex set, V is the collection of all different amino acids presented in the peptide/protein and weight of a vertex in $\mathcal{G}$ is the number of times it appears in the sequence of the peptide/protein. Two vertices are said to be adjacent in $\mathcal{G}$ if they are consecutive elements in the sequence and also have at least one common physicochemical property.

Remark 3.2: Weight of a vertex implies the frequency of occurrence of a specific amino acid in a sequence. Obviously greater the weight of a vertex of a Pt-graph implies greater the characteristics of those particular amino acid can be attributed to the peptide/protein.

Remark 3.3: Centrality Measures of a Pt-graph help us to identify the quantity of related amino acids by direct or indirect with neighbor amino acids in the sequence of corresponding peptide/protein which shares at least one common physicochemical property.

Definition 3.3: SPt-graph is a graph derived from the Pt-graphs of a peptide/protein. It is defined as a graph $\mathcal{G} = (V, E)$ of a peptide/protein in which the vertex set, V is the collection of species containing the given peptide/protein. Two vertices S_1 and S_2 in SPt-graph are said to be adjacent if in both Pt-graphs of species, the highest value of at least one centrality measures must be received by the same vertex.

IV. GRAPH THEORETICAL ANALYSIS OF Pt-GRAPH AND SPt-GRAPH OF GALANIN

Here we construct and analyze graph theoretically the Pt-graph and SPt-graph of Galanin. For the construction, we consider fourteen species of animals. They are American alligator (*Alligator mississipiensis*), Bowfin (*Amia calva*), Brown rat (*Rattus norvegicus*), House mouse (*Mus musculus*), Cattle (*Bos Taurus*), Ciona intestinalis (*Ciona intestinalis*), Human (*Homo sapiens*), Japanese quail (*Coturnix japonica*), Red junglefowl (*Gallus gallus*), Marsh frog (*Pelophylax ridibundus*), Rainbow trout (*Oncorhynchus mykiss*), Sheep (*Ovis aries*), Small-spotted catshark (*Scyliorhinus canicula*) and Wild boar (*Sus scrofa*). Table 1 gives the amino acid sequences of Galanin in fourteen species of animals.

ISSN: 2321-8169
342 – 346

Table 1: Amino acid sequences of Galanin contained in fourteen species of animals

	Different species	Amino acid sequences
1	American alligator	GWTLN SAGYL LGPHA IDNHR SFNEK HGIA
2	Bowfin	GWTLN SAGYL LGPHA VDNHR SLNDK HGLA
3	Cattle	GWTLN SAGYL LGPHA LDSHR SFQDK HGLA
4	Ciona intestinalis	GWTLN SAGYL LGPHA IDSHR SLGDK RGVA
5	Japanese quail	GWTLN SAGYL LGPHA VDNHR SFNDK HGFT
6	Red junglefowl	GWTLN SAGYL LGPHA VDNHR SFNDK HGFT
7	Human	GWTLN SAGYL LGPHA VGNHR SFSDK NGLTS
8	House mouse	GWTLN SAGYL LGPHA IDNHR SFSDK HGLT
9	Rainbow trout	GWTLN SAGYL LGPHG IDGHR TLSDK HGLA
10	Sheep	GWTLN SAGYL LGPHA IDNHR SFHDK HGLA
11	Marsh frog	GWTLN SAGYL LGPHA IDNHR SFNDK HGLA
12	Brown rat	GWTLN SAGYL LGPHA IDNHR SFSDK HGLT
13	Small-spotted catshark	GWTNL SAGYL LGPHA VDNHR SLNDK HGLA
14	Wild boar	GWTLN SAGYL LGPHA IDNHR SFHDK YGLA

By using the sequences of Galanin we construct fourteen Pt-graphs of species of animals. For the construction we consider the Pt-graph of Galanin of Human as an example. Here the amino acids $G_5, A_2, V_1, W_1, L_4, F_1, P_1, Y_1, S_4, T_2, N_3, D_1, K_1, H_2$ and R_1 are the vertices. Table 2 represents the amino acids and corresponding centrality measures of Pt-graph of Galanin of Human.

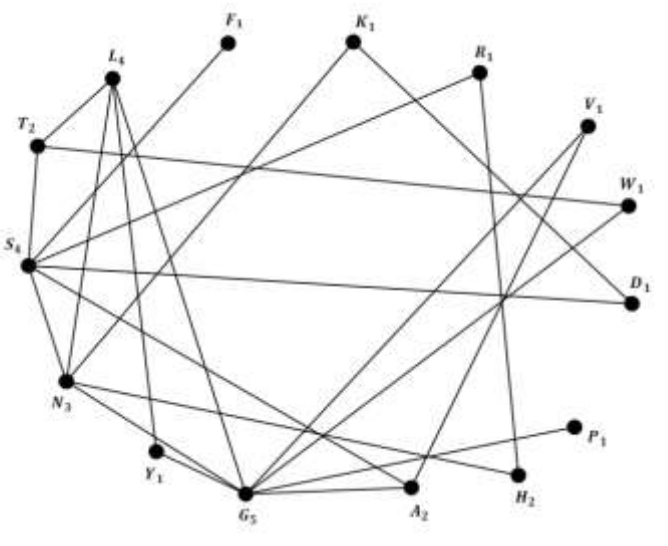

Fig 1: Pt-graph of Galanin of Human

Table 2: Amino acids and corresponding centrality measures of Pt-graph of Galanin of Human

Amino Acids	Centrality measures			
	Degree	Closeness Centrality	Betweeness Centrality	Eigenvector Centrality
G_5	7	0.58	31.83	1
A_2	3	0.52	8.5	0.58
V_1	2	0.44	0	0.42
W_1	2	0.45	1.33	0.40
L_4	4	0.52	7	0.75
F_1	1	0.37	0	0.20
P_1	1	0.37	0	0.27
Y_1	2	0.4	0	0.47
S_4	6	0.58	34.67	0.74
T_2	3	0.48	7	0.51
N_3	5	0.61	28.33	0.84
D_1	2	0.40	2	0.29
K_1	2	0.41	2.17	0.31
H_2	2	0.41	2.17	0.31
R_1	2	0.40	2	0.30

By proceeding like this we can construct all Pt-graphs of Galanin in fourteen species of animals. Table 3 represents the vertices which receive the highest values of centrality measures of Pt-graphs of Galanin.

Table 3: Vertices which receive the highest value of centrality measure for Pt-graphs

	Species	Centrality Measures			
		Degree	Closeness	Betweeness	Eigenvector
1	American alligator	G_4,N_3	N_3	N_3	G_4
2	Bowfin	L_5	L_5	L_5	L_5
3	Brown rat	G_4, L_4	N_2, S_3	S_3	L_4
4	House mouse	G_4, L_4	N_2, S_3	S_3	L_4
5	Cattle	S_3	S_3	S_3	S_3
6	Ciona intestinalis	G_5	G_5	G_5	G_5
7	Human	G_5	N_3	S_4	G_5
8	Japanese quail	G_4, N_3	N_3, F_2	N_3	N_3
9	Red junglefowl	G_4, N_3	N_3, F_2	N_3	N_3
10	Marsh frog	G_4	L_4	N_3	N_3
11	Rainbow trout	G_6, L_5	G_6	G_6	L_5
12	Sheep	G_4	L_4	L_4	L_4
13	Small-spotted catshark	G_4, L_5	L_5	L_5	L_5
14	Wild boar	G_4, L_4	L_4	L_4	L_4

From the graph theoretical analyses of Pt-graphs of Galanin we get some observations.

Observation 4.1: The amino acids which receive the highest value of degree centrality for American alligator, Japanease quail and Red junglefowl are same having the same weight. (that is G_4 and N_3).

Observation 4.2: The amino acids which receive the highest value of degree centrality for Brown rat, House mouse and wild boar are same having the same weight. (that is G_4 and L_4).

Observation 4.3: The amino acids which receive the highest value of degree centrality for Ciona intestinalis and Human are same having the same weight. (that is G_5).

Observation 4.4: The amino acids which receive the highest value of degree centrality for Marsh frog and Sheep are same having the same weight. (that is G_4).

Observation 4.5: The amino acids which receive the highest value of closeness centrality for American alligator and Human are same having the same weight. (that is N_3).

Observation 4.6: The amino acids which receive the highest value of closeness centrality for Bowfin, Marsh frog, Sheep, Small-spotted catshark and Wild boar are same (that is L).

Observation 4.7: The amino acids which receive the highest value of closeness centrality for Ciona intestinalis and Rainbow trout are same (that is G).

Observation 4.8: The amino acids which receive the highest value of betweenness centrality for American alligator, Japanese quail, Red junglefowl and Marsh frog are same having the same weight (that is N_3).

Observation 4.9: The amino acids which receive the highest value of betweenness centrality for Bowfin, Sheep, Small-spotted catshark and Wild boar are same (that is L).

Observation 4.10: The amino acids which receive the highest value of betweenness centrality for Brown rat, House mouse and Cattle are same having same weight (that is S_3).

Observation 4.11: The amino acids which receive the highest value of Eigenvector centrality for American alligator, Ciona intestinalis and Human are same (that is G).

Observation 4.12: The amino acid which receive the highest value of all centrality measures for Bowfin is L and that for Cattle is S.

Next we construct and analyze graph theoretically the SPt-graph of Galanin from fourteen Pt-graphs of Galanin. Here the vertices of SPt-graphs are the fourteen species of animals containing Galanin. Table 4 represents the species of animals and corresponding centrality measures of SPt-graph of Galanin.

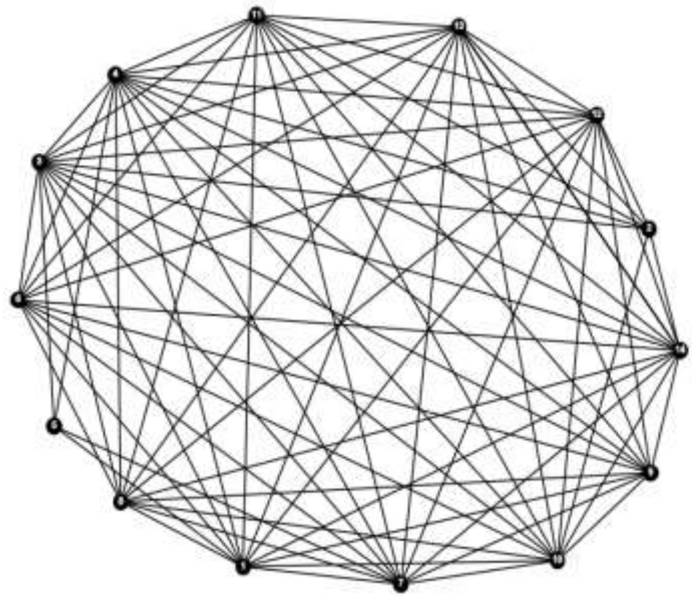

Fig 2: SPt-graph of Galanin of fourteen species of animals

Table 4: Species and corresponding centrality measures of SPt-graph of Galanin

Label	Species	Centrality Measures			
		Degree	Closeness	Betweeness	Eigenvector
1	American alligator	12	0.93	2.09	0.95
2	Bowfin	7	0.68	0.09	0.59
3	Brown rat	13	1	3.31	1
4	House mouse	13	1	3.31	1
5	Cattle	4	0.59	0	0.34
6	Ciona intestinalis	11	0.87	0.09	0.92
7	Human	12	0.93	2.09	0.95
8	Japanese quail	11	0.87	0.09	0.92
9	Red junglefowl	11	0.87	0.09	0.92
10	Marsh frog	12	0.93	0.81	0.97
11	Rainbow trout	12	0.93	0.81	0.97
12	Sheep	12	0.93	0.81	0.97
13	Small-spotted catshark	11	0.87	0.71	0.90
14	Wild boar	12	0.87	0.71	0.90

From the graph theoretical analyses of SPt-graph of Galanin we get some observations.

Observation 4.13: Brown rat and House mouse receive the same value of amino acids having highest among all the centrality measures in SPt-graph of Galanin.

Observation 4.14: Marsh frog, Rainbow trout and Sheep receive the same amino acids which have the same values for all centrality measures in SPt-graph of Galanin.

ISSN: 2321-8169
342 – 346

References

[1] Adil Akhtar and Nisha Gohain, Graph theoretic approach to analyze amino acid network, Int. J. Adv. Appl. Math. And Mech. 2(3) (2015) 31-37.

[2] Adil Akhtar and Tazid Ali, Analysis of Unweighted Amino Acids Network, Hindawi Publishing Corporation, Vol. 2014, Article ID 350276 (2014) 6 pages.

[3] David L. Nelson and Michael M. Cox, Lehninger Principles of Biochemistry, W. H Freeman and Company (2008).

[4] S. Kundu, Amino acid network with in protein, Physica A, 346 (2005) 104-109.

[5] G. Suresh Singh, Graph Theory, PHI Learning Private Limited (2010).

[6] Yang Wang, Mingnia Wang, Sanwen Yin, Richard Jang, Jian Wang, Zhidong Xue and Tao Xu, Neoropep: a comprehensive Resource of Neuropeptides, Database (Oxford) 2015 Apr 29. doi:10.1093/ database/ bar038.