Article

Application of signed graph in amino acid network

Adil Akhtar[1], **Nasrin Irshad Hussain**[2]

[1]Department of Mathematics, Golaghat Engineering College, Assam 785621, India

[2]Department of Computer Application, Sibsagar University, Assam 785665, India

E-mail: adil.akhtar19@gmail.com, hussainnasrin531@gmail

Received 4 September 2024; *Accepted 15 October 2024*; *Published online 20 October 2024*; *Published 1 March 2025*

Abstract

This article is aimed to construct a model related to protein structure and amino acids viz., Brandstein and Tooze's condition. To study Brandstein and Tooze's condition we have considered signed graph in amino acids. For this we have constructed a network structure, where the link is based on the one point mutation of codons of corresponding 20 different essential amino acids. To study this we have considered signed graph in amino acids and consider two cases. In the first case positive sign is assigned to an edge if both the amino acids are either hydrophobic or hydrophilic, otherwise negative sign is assigned. And in the second case positive sign is assigned to an edge if one amino acid is hydrophobic and the other is hydrophilic, otherwise negative sign is assigned.

Keywords amino acid; graph; chemical graph; signed graph.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Amino acids are biologically important organic compounds composed of amine (-NH_2) and carboxylic acid (-COOH) functional groups with a side-chain specific to each amino acid. The key elements of an amino acid are carbon, hydrogen, oxygen and nitrogen. Each amino acid have different properties. The different properties of the amino acids changes with the change in R groups. The R group is often referred to as the amino acid side chain. Amino acids are the building blocks of proteins. Each protein is formed by a linear chain of amino acids. There are 20 different amino acids being found till now that occurs in proteins. Each amino acid is a triplet code of four possible bases. A sequence of three bases forms a unit called codon. A codon specifies one amino acid. The genetic code is a series of codons that specify which amino acids are required to make up a specific protein. As there are four bases, (Adenine (A), Cytosine(C), Guanine (G) or Thymine (T/U)) this gives us 64 codons. A codon can change in several ways which is known as mutation.

In this paper we have considered one point mutation of all possible bases. The compatibility relation of the graph is defined based on the mutation of the codon. For example the amino acid M (Methionine) is

connected with K (Lysine), T (Threonine), R (Arginine), I (Isoleucine), V (Valine), L (Leucine), because all possible mutations of the base of the codon AUG (M) codes for the amino acids K, T, R, I, V and L.

The main focus in our manuscript is to give mathematical interpretation of Brandstein and Tooze's (Branden and Tooze, 2012) condition. To study this we have considered signed graph in amino acids. We have considered two cases. In the first case positive sign is assigned to an edge if both the amino acids are either hydrophobic or hydrophilic, otherwise negative sign is assigned. And in the second case positive sign is assigned to an edge if one amino acid is hydrophobic and the other is hydrophilic, otherwise negative sign is assigned.

As evolution of amino acids is mediated by mutation of corresponding codons, we argue that the network also gives a picture of evolutionary pattern. Therefore two amino acids connected by an edge can be interpreted as having affinity towards each other in the sense that one may evolve from the other.

Researchers have made significant contributions in the field of biological network. Ali and Borah (2021a) constructed a distance matrix of the amino acids. The distance is defined based on the relative evolutionary importance of the base position of the corresponding codons. From this distance matrix a network of the amino acids is obtained. They have argued that this network depicts the evolutionary pattern of the amino acids. Further they have discussed the relative importance of the amino acids with respect to this network. In Gohain et al. (2015), a partial ordering is equipped on the genetic code and a lattice structure has been developed from it. The codon-anticodon interaction, hydrogen bond number and the chemical types of bases play an important role in the partial ordering in the base set and on the structure of the genetic code. Some relations between the lattice structure of genetic code and physico-chemical properties of amino acids had established. Manisekhar and Lalitha (2013) used chemical graph theory to calculate the energies during the formation of dipeptide from amino acids. They have shown that the energy during the formation of dipeptide is conserved. They calculated correlation between the eigen values and the atomic mass for twenty amino acids in chemical graphs. Aftabuddin and Kundu (2007) discussed hydrophobic, hydrophilic and charged network within the protein. Hydrophobic amino acids are considered as vertices in the hydrophobic network, whereas hydrophilic and charged amino acids are considered as vertices of hydrophilic and charged networks, respectively. If any two atoms from different amino acids are within a cutoff distance of $5A^{o}$, the amino acids are considered to be connected. Further they showed that the average degree of the hydrophobic networks has a significantly larger value than that of hydrophilic and charged networks. The average degree of the hydrophilic networks is slightly higher than that of the charged networks. The average strength of the nodes of hydrophobic networks is nearly equal to that of the charged network, whereas that of hydrophilic networks has a smaller value than that of hydrophobic and charged networks. The average strength for each of the three types of networks varies with its degree. The average strength of a node in a charged network increases more sharply than that of the hydrophobic and hydrophilic networks. Each of the three types of networks exhibits the ''small-world'' property. Jiao et al. (2007) discussed the weighted amino acids network based on the contact energy. They have shown that weighted amino acid network satisfy "small-world" property. Schreiber and Koschutzki (Koschützki and Schreiber, 2004) compared centralities for biological networks, namely PPI network and transcriptional network. As a result of their study, it was observed that in the analysis of biological networks various centrality measures should be considered. Strub et al. (Strub et al., 2004) discussed effect of the mutation on stability. They have also studied protein stability by mutation of hydrophobic amino acids to arginine (hydrophilic). Wuchty and Stadler (2003) discussed various centrality measures in biological network. They concluded that the degree of vertex centrality alone is not sufficient to distinguish lethal protein from viable ones. Estraade et al. (Estrada et al., 2003)gave a review on the use of topological indices in drug design and discovery. Fell and Wagner (2000) examined whether metabolites with highest degree may belong to the

oldest part of the metabolism. Balaban (1985) used chemical graph theory to mathematically model the molecules in order to gain insight into their physical and chemical properties. Chemical graph theory is an important tool for designing drug. In (Boruah, 2022) construct identity sub-graph on the genetic code and they have studied different centrality measures, clustering coefficient and degree of distribution. Also, they construct a network based on distance matrix. The distance matrix is obtained by transition and transversion mutation of codons. Finally they proposed that this network reflects the evolutionary patterns of amino acids. Ali and Borah (2021b) constructed the amino acid networks based on properties that exhibits by amino acids. Correlating some physico-chemical and physical properties of amino acid, they analyzed the relative importance of different amino acids using centrality measures.

2 Graph and Signed Graph

2.1 Concepts of graph and signed graph

An undirected graph $G = (V, E)$ consists of a finite set V of vertices and a finite set $E \subseteq V \times V$ of edges. If an edge $e = (u, v)$ connects two vertices u and v then vertices u and v are said to be incident with the edge e and adjacent to each other. In general, the vertices of a graph represent objects or entities while the edges represent relationship between the vertices (Zhang, 2016, 2018). The complete graph is a graph in which each vertices connects to each other. A directed graph or digraph G consists of a set V of vertices and a set E of edges such that $e \in E$, if each edge of the graph G has a direction. An adjacency matrix A of a graph $G = (V, E)$ is a $(n \times n)$ matrix (a_{ij}), where the entries are either 1 or 0. $a_{ij} = 1$ if $(v_i, v_j) \in E$, $v_i, v_j \in V$, $a_{ij} = 0$ otherwise.

The adjacency matrix of any undirected graph is symmetric.

Signed graph and balanced theory are two important tools of graph theory. Harary (1953) in his first article on signed graph gave a simple characterization of those in which the product of signs around every cycle is positive. Such graph is called balanced or stable. The concept of balanced graph proposed by Heider's balance theory discusses the relations among individual groups. Using signed graph and balanced theory, we can say whether the graph is stable (balanced) or unstable (unbalanced). A signed graph is simply a graph where each edge between the nodes is labeled as either positive (+) or negative (–). With the help of signed graph we may represent the interpersonal relationships between individuals or groups of individuals.

The following theorem will be used in the sequel.

Theorem 2.1: A signed graph G is called balanced (Harary, 1953) if and only if its vertex set V can be separated into two disjoint subsets namely V_1 and V_2, in such a way that each positive edge of G joins two vertices of the same subset and each negative edge joins two vertices of different subsets.

In the following we have given example of a balanced and an unbalanced graph.

Consider the following graph (Fig. 1), where a, b, c, d, e, f represent six persons working in an office. Where positive edge between two vertices (persons) means they wish to work together and negative edge means they do not wish to work together.

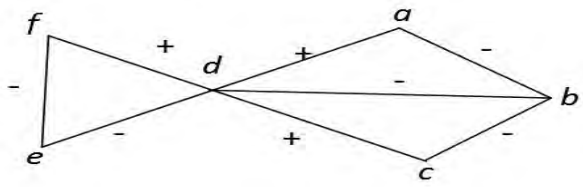

Fig. 1 Balanced graph.

Here the vertices can be partitioned into two disjoint sets say, $V_1 = \{a, c, d, f\}$, and $V_2 = \{b, e\}$. We observed that between vertices of V_1 only positive edges exist and similarly between vertices of V_2 only positive edges exist whereas only negative edges connect vertices of V_1 with vertices of and V_2. Hence, by Harary's theorem the graph represented by Fig. 1 is balanced.

Again we consider the same graph as above by changing some signs. The corresponding graph is shown in Fig. 2.

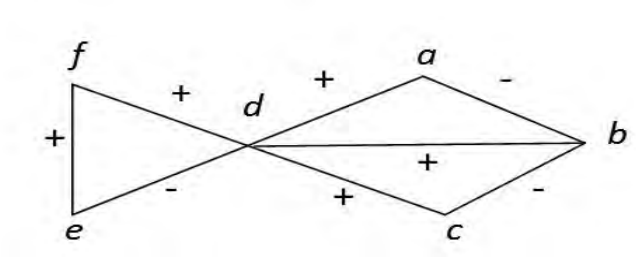

Fig. 2 Unbalanced graph.

In the signed graph of Fig. 2 the set of vertices cannot be partitioned into two disjoint subsets, satisfying the conditions of Harary's theorem. Hence the graph is unbalanced.

2.2 Signed graph in amino acids network

Every codon codes a unique amino acid. A one point mutation of a codon may or may not change the corresponding coded amino acid. All one point mutations of a codon gives nine more codons. Some of these nine codons will code for some amino acid(s) other than the original one. Since any mutation has its reverse mutation, an amino acid transforming to another may revert back to its original state. This process of mutation/transformation may be looked upon as a bidirectional relation. Thus in the set of codons (amino acids) we have a binary relation which generates an undirected graph.

In our amino acids graph the vertex set is the set of amino acids and two amino acids α and β are linked/connected by an edge if one point mutant of any codon coding α codes for β. For example the codon AUG codes for the amino acid M (Methionine). All possible one point mutations(first, second and third base) of the codon AUG are CUG, GUG, UUG (obtained by first base mutation); AAG, ACG, AGG (obtained by second base mutation); AUA, AUC, AUU(obtained by third base mutation). The mutated codons codes for the amino acids K (Lysine, coded by AAG), T (Threonine, coded by ACG), R (Arginine, coded by AGG), I (Isoleucine, coded by AUA, AUC, AUU), V (Valine, coded by GUG), L (Leucine, coded by CUG, UUG). So in our amino acids graph the amino acid M is connected to the amino acids K, T, R, I, V and L. Again the amino acid K (coded by AAG and AAA) is not connected to the amino acid P (coded by CCU, CCC, CCA CCG) as neither AAG nor AAA mutates to any of the codons CCU, CCC, CCA, CCG. Similarly other links between amino acids are obtained.

The corresponding graph is depicted in Fig. 3.

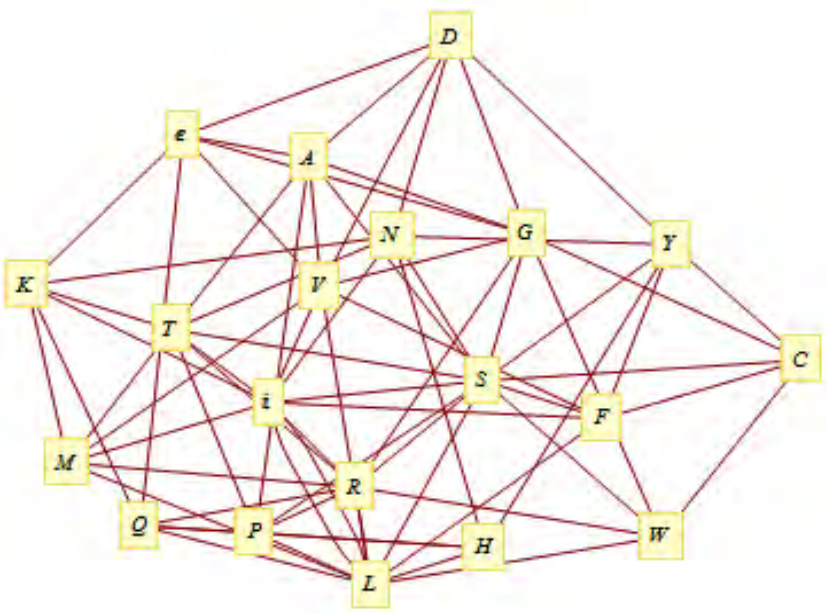

Fig. 3 Amino acids graph.

Corresponding adjacency matrix of the graph is given below

$$M=\begin{bmatrix}
0&1&1&0&1&0&0&0&0&0&1&0&1&1&0&0&1&0&1&0\\
1&0&1&0&0&0&0&0&0&1&1&1&1&0&0&0&1&0&0&0\\
1&1&0&1&0&1&1&1&0&0&0&0&1&0&0&0&1&0&0&0\\
0&0&1&0&0&1&1&0&0&0&0&1&0&0&0&0&0&1&1&0\\
1&0&0&0&0&1&0&0&0&0&1&0&0&1&0&0&0&0&1&0\\
0&0&1&1&1&0&1&1&0&1&1&0&0&0&0&1&0&0&1&1\\
0&0&1&1&0&1&0&1&0&0&1&1&0&0&1&0&0&1&1&0\\
0&0&1&0&0&1&1&0&1&0&1&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&1&0&0&1&0&0&1&1&0&1&0&0&1\\
0&1&0&0&0&1&0&0&0&0&1&1&0&0&0&1&0&0&1&1\\
1&1&0&0&1&1&1&1&1&1&0&1&0&1&1&0&0&0&1&0\\
0&1&0&1&0&0&1&0&0&1&1&0&0&0&1&0&0&1&1&0\\
1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&0&0\\
1&0&0&0&1&0&0&1&1&0&1&0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&1&0&1&0&1&1&0&0&0&0&1&1&0&1\\
0&0&0&0&0&1&0&0&0&1&0&0&1&0&0&0&0&1&1&1\\
1&1&1&0&0&0&0&0&1&0&0&0&1&0&1&0&0&0&0&1\\
0&0&0&1&0&0&1&0&0&0&0&1&1&0&1&1&0&0&1&0\\
1&0&0&1&1&1&1&0&0&1&1&1&0&1&0&1&0&1&0&1\\
0&0&0&0&0&1&0&0&1&1&0&0&0&0&1&1&1&0&1&0
\end{bmatrix}$$

From the above adjacency matrix it is clear that the graph (Fig. 3) is connected because no row or column of the upper (or lower) triangular matrix is zero.

We next constructed a signed graph in this amino acids network. We consider two cases here. In the first case we assign positive sign to an edge between two amino acids if both the amino acids are either hydrophobic or hydrophilic, otherwise negative sign is assigned. The positive edges are marked **E+**. The corresponding signed graph is shown in Fig. 4.

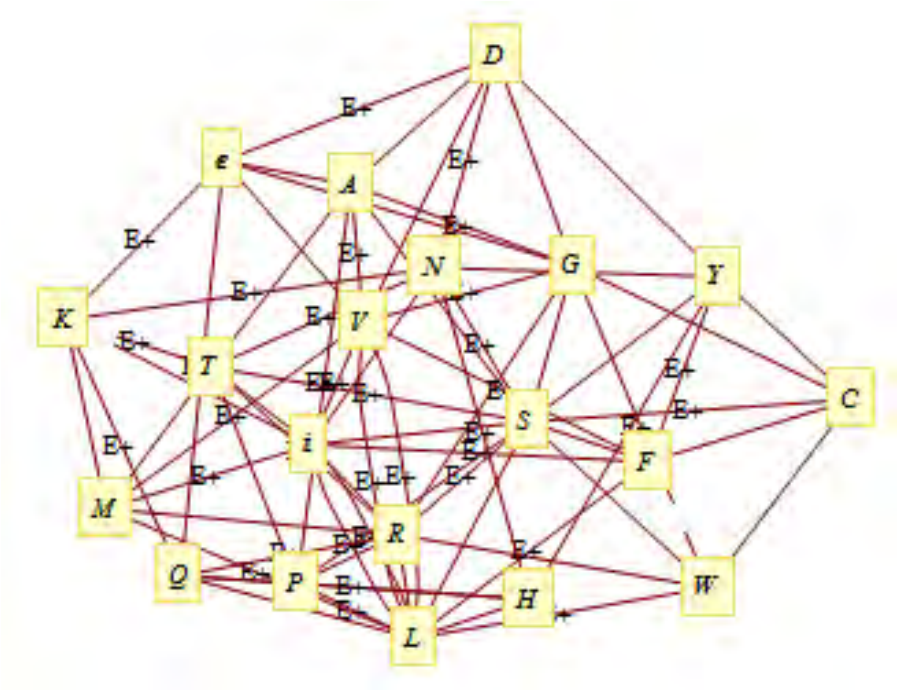

Fig.4 Balanced signed graph.

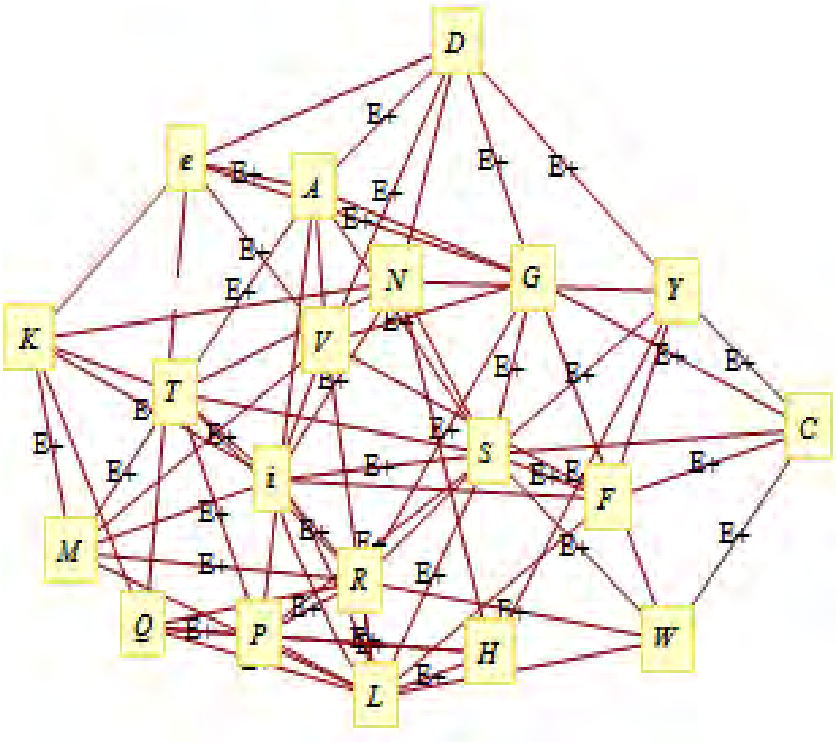

Fig. 5 Unbalanced signed Graph.

Following pairs of vertices represent positive edges.

V={{G, A}, {G, V},{G, W},{A, P},{A, V},{V, I},{V, L}, {V, F}, {V, M}, {M, L}, {M, I}, {W, L},{L, I},{L, V}, {L, F}, {L, P}, {I, L}, {I, F}, {F, Y}, {S, R}, {S, C}, {S, N}, {S, T}, {T, N}, {T, K}, {T, R}, {E, K}, {E, Q}, {E, D}, {C, R}, {N, H}, {N, D}, {N, K}, {Q, K}, {Q, R}, {Q, H}, {D, H}, {R, Q}}.

We want to show that the graph is balanced (stable). Let the vertex set be separated into two disjoint subset, say V_1 and V_2 where, V_1 = {G, A, V, W, P, L, I, F, M, Y} and V_2 ={S, R, C, D, E, T, K, N, Q, H}. Here we observed that any amino acid of V_1 and any other amino acid of V_2 are linked by a negative edge, but within the same set each pair of amino acids are linked by positive edges. Hence the pair (V_1, V_2) satisfies the Harary's theorem. Consequently signed graph of amino acids network based on mutation is stable or balanced.
Again in the second case we assigned positive sign to an edge between amino acids if one is hydrophobic and another is hydrophilic, otherwise negative sign is assigned. The corresponding signed graph is shown in the Fig. 5.

Following pairs of vertices represent positive edges.
V={{G, S}, {G, R},{G, C},{G, D},{G, E},{A, T},{A, S}, {A, D}, {A, E}, {V, D}, {V, E}, {M, K},{M, T},{M, R}, {W, R}, {W, S}, {W, C}, {L, S}, {L, H}, {L, R}, {L, Q}, {I, N}, {I, T}, {I, S}, {I, K}, {I, R}, {F, S}, {F, C}, {Y, N}, {Y, H}, {Y, D}, {Y, S}, {Y, C}, {P, T}, {P, S}, {P, H}, {P, R}, {P, Q}}.

Here the vertex set cannot be separated in two disjoint subsets such that which satisfies Harary's theorem (Harary, 1953). Consequently the amino acids network given by Fig. 5, is unbalanced.

According to the first case the network is stable or balanced where there is positive sign in between similar amino acids (both hydrophobic or both hydrophilic). And for the second case the network is unbalanced or unstable where there is positive sign between amino acids of different properties (hydrophobic and hydrophilic). Therefore from the first case it is clear that if any one of the hydrophobic amino acid is replaced by another hydrophobic amino acid or any one of the hydrophilic amino acid is replaced by another hydrophilic amino acid then the network remains stable or balanced.

Again from the second case it is clear that if any one of the hydrophobic amino acid is replaced by other hydrophilic amino acid or vice-versa then the network is unbalanced.

3 Brandstein and Tooze's Condition

Brandstein and Tooze (2012) observed that the physico-chemical nature of the amino acids (whether hydrophobic or hydrophilic) of a protein plays a major role in determining its 3D structure. We know that the hydrophobic amino acids are water hating and the hydrophilic amino acids are water loving. Since a protein within the cell is in aqueous environment, hydrophobic amino acids come close each other and are located in the core region of the folding structure, whereas the hydrophilic amino acids go to the surface region. For a mathematical interpretation of Brandstein and Tooze's condition in graph theoretic sense we have considered signed graph in amino acids. When one of the hydrophobic amino acid is replaced by another hydrophilic amino acid or vice-versa then the protein structure is deformed, that is it becomes unbalanced or unstable. Again when hydrophobic amino acid is replaced with other hydrophobic amino acid or hydrophilic amino acid is replaced by other hydrophilic amino acid then it remains un-deformed or deformation probability is less in comparison to previous case.

We know that mRNA contains base sequences. A segment is formed from the base sequences to produce a protein. A protein is a sequence of amino acids. Each amino acid is coded by a codon from the mRNA segment. If there is a mutation in any of the codons, there may be a change in the corresponding amino acid. If there is any change in the amino acid, the nature may change (hydrophobic changed to hydrophilic or vice-versa) or may not change (hydrophobic remaining hydrophobic, hydrophilic remaining to hydrophilic). If there is a change in any amino acid of the protein, the 3D structure of the protein may change. This fact can be interpreted in terms of the signed graphs discussed above. The first signed graph represents the situation where due to mutation of codon the amino acid nature (hydrophobic/hydrophilic) remains intact. The graph is balanced can be interpreted as no change in 3D structure of the protein structure whereas the second graph

represents the opposite situation. There mutation of codon changes the nature of the amino acid and consequently a change in the 3D structure of the protein. This is represented by the unbalanced signed graph. From the above discussion it is clear that in amino acids network at an unbalanced or unstable situation the folding of proteins is easier as compared to the balanced or stable situation.

4 Conclusion

In this manuscript we have attempted a modeled of Brandstein and Tooze's phenomenon by signed graph. Based on mutations of codons, link between corresponding amino acids were obtained to give a graph structure to the amino acids set. When a codon mutates, the corresponding amino acid may change and consequently the hydrophobicity/hydrophilicity nature of the amino acid may change. Depending on the change of hydrophobicity/hydrophilicity nature, positive /negative sign may be assigned to the edges, which in turn gives a signed graph. We have considered two cases to define the positive and negative sign of the edges between the amino acids. Using Harary's theorem we observed that in the first case the amino acids network is stable or balanced but in the second case the network is unbalanced or unstable. Balanced graph may be interpreted as no change in the 3D structure of the protein molecule, whereas unbalanced graph indicates a change in the 3D structure. Thus sign graph gives a mathematical interpretation or modeling of the Brandstein and Tooze’s phenomenon. In other words, Brandstein and Tooze’s phenomenon is represented in mathematical terms by the unstable signed graph discussed above.

References

Aftabuddin M, Kundu S. 2007. Hydrophobic, hydrophilic, and charged amino acid networks within protein. Biophys Journal, 93: 225-231. https://doi.org/10.1529/biophysj.106.098004

Ali T, Borah C. 2021a. Analysis of amino acids network based on mutation and base positions. Gene Reports, 24: 101291

Ali T, Borah C. 2021b. Analysis of amino acids network based on transition and transversion mutation of codons. Network Biology, 11(3): 125-136

Application of Algebraic Graph Theory in Genetic Code [WWW Document]. https://www.ijaresm.com/application-of-algebraic-graph-theory-in-genetic-code (Accessed 10.17.24).

Balaban AT. 1985. Applications of graph theory in chemistry. Journal of Chemical Information and Computer Sciences, 25: 334-343. https://doi.org/10.1021/ci00047a033

Borah C. 2023. Algebraic structures and distance based analysis of genetic code. Network Biology, 13(1): 17-36

Boruah BK. 2022. Special identity subgraph in genetic code. Network Biology, 12(2): 45-63

Boruah BK, Ali T. 2022. A study of the total graph in genetic code algebra. Network Biology, 12(1): 1-10

Branden CI, Tooze J. 2012. Introduction to Protein Structure. Garland Science, USA

Estrada E, Patlewicz G, Uriarte E. 2003. From molecular graphs to drugs. A review on the use of topological indices in drug design and discovery. Indian Journal of Chemistry.

Fell DA, Wagner A. 2000. The small world of metabolism. Nature Biotechnology, 18: 1121-1122

Gohain N, Ali T, Akhtar A. 2015. LATTICE STRUCTURE AND DISTANCE MATRIX OF GENETIC CODE. Journal of Biological Systems, 23: 485-504. https://doi.org/10.1142/S0218339015500254

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. Science, 185: 862-864. https://doi.org/10.1126/science.185.4154.862

Harary F. 1953. On the notion of balance of a signed graph. Michigan Mathematical Journal, 2: 143-146

Hussain NI, Boruah K. 2024. Analysis of amino acids network based on graph mining. Network Biology, 14: 242-253

Jiao X, Chang S, Li C, Chen W, Wang C. 2007. Construction and application of the weighted amino acid network based on energy. Physics Review E, 75: 051903. https://doi.org/10.1103/PhysRevE.75.051903

Koschützki D, Schreiber F. 2004. Comparison of Centralities For Biological Networks. Germany

Kumar P. Mina U. 2014. Life Sciences Fundamentals and Practice (Part-II). Pathifnder Publication, India

Manisekar M, Lalitha S. 2013. A study of matrix energy during peptide formation through chemical graphs. International Journal of Mathematics and Soft Computing, 3(2): 11-15

Sanchez R, Morgado E, Grau R., 2005. Gene algebra from a genetic code algebraic structure. Journal of Mathematical. Biology, 51: 431-457. https://doi.org/10.1007/s00285-005-0332-8

Sengupta D, Kundu S. 2012. Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization. BMC Bioinformatics, 13: 142. https://doi.org/10.1186/1471-2105-13-142

Strub C, Alies C, Lougarre A, Ladurantie C, Czaplicki J, Fournier D. 2004. Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. BMC Biochemistry, 5: 9. https://doi.org/10.1186/1471-2091-5-9

Wuchty S, Stadler PF. 2003. Centers of complex networks. Journal of Theoretical Biology. 223: 45-53. https://doi.org/10.1016/S0022-5193(03)00071-7

Zhang WJ. 2016. Selforganizology: The Science of Self-Organization. World Scientific, Singapore

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK