On Continuity of Evolution of Amino Acids

Nisha Gohain

Abstract*:** ***Continuity plays a crucial role in any discussion on evolution. A nearness relation in the phenotype space allows to distinguish continuous evolution from discontinuous one. The genotype-phenotype map plays a crucial role in evolutionary biology. In this paper we have investigated continuity in evolution of amino acids. Nearness relations in the space of amino acids as well as in the set of codons are introduced. Here any codon is a genotype and the corresponding amino acid is the phenotype. We examined the continuity of the genotype-phenotype map with respect different neighbourhood structures of amino acids.

Keywords* : *Amino acids, Continuity, Genotype space, Phenotype space, vicinity.

I.   INTRODUCTION

Evolutionary biology is concerned with the determination of the organism's fitness, changes, adaptability in the nature. Out of all essential features of living organisms one is variation which leads to evolution. The course of biological evolution cannot be understood without the dichotomy of genotype and phenotype. The genotype is the genetic makeup of an individual organism. And the phenotype is the actual appearance of an organism which describes the processes of development of the organism that is the phenotype is the observed characteristic of the organism. Understanding which phenotype is accessible from which genotype is fundamental for understanding the evolutionary process [1].

Natural selection is the fundamental mechanism through which evolution occurs. Since selection usually acts to reduce variability, there must be a source of new genetic variation. This is introduced at the molecular level, in DNA of individuals. Depending on the nature of these changes in the DNA, off springs may be more, less, or equally viable than the parents. The DNA within a particular gene may continue to mutate from generation to generation, gradually accumulating more differences from its ancestral form.

In 1991, Alberch[2] introduced the concept of GP-map in genes. Several researchers [3]-[9] study the sturucture and patterns of   RNA folding through GP-map. The GP-map reflects the ancestral situation in the RNA world hypothesized within the context of the origin of life problem. The connections between genotype and phenotype on one hand and between phenotype and fitness on the other hand are

explicit, relatively simple and biologically meaningful [10]. Fontana found the RNA folding is an empirically tractable model of GP- mapping[3]. Cowperthwaite & Meyers [4] discussed that the connectivity in genotypic space is asymmetrical, so evolution depends to some extent on contingent events .Also, Sumedha et al.[6] states that genotypic one-step neighbours are often connected to phenotypes that are not structurally similar. Stich et al. [8] pointed that GP mapping is characterized by high degrees of 'degeneracy' or genetic redundancy, where the same phenotype is generated by a variety of not necessarily similar genotypes. Stadler states that patterns of phenotypic evolution such as punctuation, irreversibility, and modularity result naturally from the properties of the genotype-phenotype map, which, given the genetic accessibility structure, defines accessibility in the phenotype space [11].

Evolution is change of phenotypes with respect to time. Mathematically expressed, it as a function

$$\phi : R \to P$$

Where R is the time axis, say real line with a suitable topology and P is the phenotype set. This function is called evolutionary trajectory or evolutionary map.

The change in phenotype is actually caused by change in genotype. Thus with respect to time genotype changes and that change in genotype produced changes in phenotype which is observable. Hence the evolutionary map can be expressed as composition of two maps:

$$g{:}R \to G \text{ and } f{:}G \to P \text{ that is } f{\cdot}g = \phi .$$

Here G stands for the genotype set.

Now to study whether the evolution is continuous or not, we need to examine whether the changes from one phenotype to another is gradual or not. For that we need some notion of nearness or accessibility among phenotypes. As the evolution in phenotypes is mediated by change in genotypes, we also require a nearness relation in the set of genotypes. Fontana et al. considered evolution of RNA shapes where various nearness relations were discussed [1][3][11][12][13]. In case of genotypes, there was no ambiguity regarding neighbourhood. Given a RNA sequence, all one point mutations were considered as its neighbour. In case of phenotypes the case is not that straight forward. Fontana and Schuster defined two nearness relations[14]:

Accessibility pre-topology: Given a shape α, another shape β is accessible from α; written as β ← α if ∃ a sequence folding into α which admits a one-error mutant folding into β.

Shadow pre-topology: β ← α if every sequence folding

into α must admit at least one one-error mutant folding into β.

These two are very extreme cases of nearness. In the former case evolution is always continuous where as in case of the later it is hardly ever.

So it was suggested to define more meaningful neighbourhood structure depending on the likelihood of phenotypic change β ← α, which is basically frequency based[14].

In this regard two frequencies, namely, neighbourhood frequency and occurrence frequency were defined respectively as:

$$v(\beta, \alpha) = \frac{|S(\alpha) \cap \partial S(\beta)|}{|S(\alpha)|} = \frac{|D(\alpha \leftarrow \beta)|}{|S(\alpha)|} \quad (1)$$

$$\vartheta(\beta, \alpha) = \frac{|S(\beta) \cap \partial S(\alpha)|}{3n|S(\alpha)|} = \frac{|D(\beta \leftarrow \alpha)|}{3n|S(\alpha)|} \quad (2)$$

where, |X| stands for the number of elements of set X.
S(α)=The set of all sequences folding into α.
∂S(α)=Set of all sequences that are obtained as point mutations of elements from the neutral set S(α). It is called the boundary of S(α).

D(α← β)=S(β)∩ ∂S(α) describes all those sequences folding into β which are neighbours of sequences folding into α.

Cupal et al. [15] defined a topology on RNA space as β is accessible from α only if the probability is high that a population is located on the neutral network S(α) will find a sequence folding into β.

Accessibility may now be based on size of D(β ← α) relative to S(α), S(β) or their boundaries. Probably the most natural quantity is the ratio probability

$$\frac{|D(\beta \leftarrow \alpha)|}{|\partial S(\alpha)|}$$

that a non-neutral mutant of an α-sequence folds into the shape β.

Our study will be concerned with amino acid evolution. Genetic code lies at the heart of this evolution. The genetic code consists of 64 triplets of nucleotides (bases). These triplets are called codons. Each codon encodes for one of the 20 amino acids used in the synthesis of proteins where most of the amino acids being encoded by more than one codon. There are 64 codons. Out of these 64, the three triplets UAA, UAG and UGA are known as stop codons or nonsense codons and their role is to stop the biosynthesis. The codon AUG codes for the initiation of the translation process and is therefore also known as start codon.

II. TOPOLOGICAL ASPECTS AND CONTINUITY

In this section we discuss the basic concepts of topology and the notion of continuity.

A. Topology

A structured set is called a space. A set can be structured in different ways and one such method is by equipping it with a topology. Topology on a given set is a structure that enables us to discuss notions like convergence and continuity. The notion is an abstraction of the concrete idea of distance (metric). A topology on a set can be defined in many equivalent ways, namely in terms of open (closed) sets, interior operator, closure operator, neighbourhood system, cluster points, boundary operator. The most frequently used definition of topology is in terms of open (closed) sets.

Definition: Let X be set. A topology on a set X is a non-empty collection T of subsets of X such that

- The empty set and whole set X belongs to T.
- Arbitrary union of members of T is in T (respectively finite intersection of members of T is in T).
- Finite intersection of members of T is in T (respectively arbitrary union of members of T is in T).

The members of T are called open sets (resp. closed sets).

This definition induces the following notions:

Neighbourhood of apoint: A subset N of X is a neighborhood of a point x if for every element of N there is an open set containing x and contained in N. The collection of all the neighbourhoods of a point x is called the neighbourhood system of x.

Limit point of a set: A point x in X is a limit point of a set G if every open set containing x contains a point of G other that x itself. The set of all the limit points of G is called the derived set of G. The union of G and derived set of G is called closure of G.

Interior point of a set. An element xof a set G is called an interior point of G if there exists an open set containing x and contained in G. The set of all the interior points of G is called the interior of G.

Boundary point of a set : An element x in X is said to be a boundary point of a set G if it is a limit point of both G and G^c. The set of all the boundary points of G is called the boundary of G.

Other equivalent definitions of topology are given below
Definition
Let X be a set and $C_l : P(X) \rightarrow P(X)$, a map such that

$(1) C_l(\phi) = \phi$

$(2) A \subseteq C_l(A)$

$(3) C_l(A \cup B) = C_l(A) \cup C_l(B)$

$(4) C_l(C_l(A)) = C_l(A)$

Then the collection $\mathcal{T} = \{G \subseteq X | G = C_l(G)\}$ is a topology on X.

Each member of $\mathcal{T}$ is called a closed set and complement of a closed set is called open set.

Definition
Let X be a set and $B_d : P(X) \rightarrow P(X)$ be a map such that

(1) $B_d(\phi) = \phi$

(2) $B_d(A) = B_d(A^c)$

(3) $A \cap B \cap B_d(A \cap B) = A \cap B \cap (B_d(A) \bigcup B_d(B))$

(4) $B_d(B_d(A)) \subseteq B_d(A)$

Then the collection $\mathcal{T} = \{G \subseteq X \mid B_d(A) \subseteq G\}$ is a topology on X. Each member of $\mathcal{T}$ is called a closed set.

Definition

Let X be a non-empty set. A map $N: X \to P(P(X))$ such that

(1) $x \in N, \qquad \forall N \in N(x)$

(2) $N \in N(x)$ and $N \subseteq N_1, N_1 \in N(x)$

(3) $N_1, N_2 \in N(x), N_1 \cap N_2 \in N(x)$

(4) For each $N \in N(x)$, $\exists\ N' \in N(x)$ such that $N \in N(y) \forall y \in N'$.

Each member of $N(x)$ is called a neighbourhood of x. Then the collection
$\mathcal{T} = \{G \subseteq X \mid G \in N(x), \forall x \in G\}$ is a topology on X. Each member of $\mathcal{T}$ is called an open set.

Definition

Let X be a non-empty set and $\gamma: P(X) \to P(X)$, a map satisfying

(1) $\gamma(\emptyset) = \emptyset$

(2) $\gamma(\gamma(A)) \subseteq A \cup \gamma(A)$

(3) $\gamma(A \cup B) = \gamma(A) \cup \gamma(B)$

(4) For each $x \in X, x \in \gamma(\{x\})$

Then the collection $\{A \cup \gamma(A) \mid A \in P(X)\}$ is a topology on X. Here every member of the collection is a closed set.

Definition

Let X be a non-empty set and $Int: P(X) \to P(X)$ be a map satisfying

(1) $Int(\emptyset) = \emptyset$

(2) $Int(A) \subseteq A$

(3) $Int(A \cap B) = Int(A) \cap Int(B)$

(4) $Int(Int(A)) = Int(A)$

Then the collection $\{Int(A): A \in P(X)\}$ is a topology on X. Here each member of the collection is an open set.
Spaces that we encounter in practical situation may not satisfy all axioms. If the fourth condition is not satisfied in the definitions, then we get a weaker structure on X, called pre-topology. Relation between closure operator, boundary operator and neighbourhood system is given below:

$C_l(A) = \{x \mid \forall N \in N(x), A \cap N = \emptyset\} = A \cup B_d(A)$

$B_d(A) = \{x \mid \forall N \in N(x), A \cap N = \emptyset, (X - A) \cap N = \emptyset\} = C_l(A) \cap C_l(A^c)$

$N(x) = \{A \mid x \notin C_l(A^c)\} = \{A \mid x \in A - B_d(A)\}$

Though arbitrary intersection of neighbourhoods need not be a neighbourhood, finite intersection is always a neighbourhood. Hence in a finite pre-topological space there always exists a smallest neighbourhood for each point, namely, the intersection of all neighbourhood of x. The smallest neighbourhood of x is called the *vicinity* of x.
Conversely if we know the vicinity of x, then we can construct the neighbourhood system of x simply by taking all the possible supersets of the vicinity. In our study the neighbourhood structure will be based on vicinity of each point.

B. Continuity

The notion of continuity which lies at the core of topology can be defined in a variety of equivalent ways. Roughly speaking, a map *f* from a space X to a space Y is continuous if small (gradual) change in X produces small (gradual) change in Y. The small (gradual) change will be determined by the topology on X. What may be gradual with respect to one topology may not be so with respect to another topology.

Let X and Y be two pre-topological (topological) spaces and N denote neighbourhood structure for both spaces.

A map *f*: X→Y is said to be continuous at $x \in X$, if $B \in N(f(x)) \Rightarrow f^{-1}(B) \in N(x)$.

If *f* is continuous at each point of X, then we say $f: X \to Y$ is continuous.

Below we consider other equivalent definitions of continuity.

(1) $Cl(f^{-1}(B)) \subseteq f^{-1}(Cl(B)), \quad \forall B \in P(Y)$

(2) $f(Cl(A)) \subseteq Cl(f(A)), \qquad \forall A \in P(X)$

(3) For every neighbourhood M of f(x), $\exists$ neighbourhood N of $x: f(N) \subseteq M$.

(4) $M(f(x)) \subseteq f(N(x))\ \forall x \in X$, where N is the neighbourhood system of the domain X and M is the neighbourhood system of the co-domain Y.

In case of mappings between finite pre-topological spaces, continuity can be conveniently formulated in terms of vicinity.

Mapping is continuous if vicinity of the image point contains image of the vicinity of the pre-image,

i.e., $f: X \to Y$ is continuous at x,

if $f(N(x)) \subseteq M(f(x))$,

where $N(x)$ is the vicinity of x and $M(f(x))$ is the vicinity of $f(x)$. (3)

Since both our spaces, genotype and phenotype are finite; we shall examine continuity in terms of vicinity.

III. TOPOLOGY IN CODON SET AND AMINO ACID SET

We shall equip the codon set and amino acid set with neighbourhood structure which will be formulated in terms of vicinity.

It is already mentioned in the previous section that once we know the vicinity of each points, the neighbourhood structure on the set can be obtained. Now the first question is how to define the vicinity of a codon, in other words which codons can be termed to be in vicinity of a given codon. Perhaps the most obvious and natural approach would be to consider one-point mutation.

We will accept this.So by vicinity of a given codon say C_1, we will mean all one-point mutations of C_1 including C_1 itself. For example, vicinity of $C_1 = GCU$ will be

$$N(G) = \{GCU, ACU, CCU, UCU, GAU, GGU, GUU, GCA, GCC, GCG\}$$

Next question is: what would be the most natural vicinity of an amino acid? As mentioned in the introduction section, several approaches have been suggested for shaping RNA space. Similar approaches can be adapted here.

As the 64 codons code for the 20 amino acids, we consider the 64codons as genotype space and the 20 amino acids as the phenotype space. We have already mentioned that the understanding of which phenotype is accessible from which genotype is fundamental for understanding the evolutionary process. The notion of accessibility is studied using different topological structures [11]-[15]. Accessibility at the phenotypic level is the crucial determinant for evolution since it is the phenotype that is subject to selection. The structure of physical process, namely mutation, acting on genes, determines genotype space. The phenotype, on the other hand, is therefore not modified directly, but indirectly through the modification of the genome from which it arises. The accessibility relation at phenotypic level therefore has to be understood as a consequence of the interplay of genotypic accessibility and the Genotype-Phenotype (GP) map. [11]

To define vicinity in amino acid set we employ three different approaches. In the first approach we express the accessibility relationship between pair of amino acids α and β in terms of the ratio probability given by the formula

$$d(\alpha, \beta) = \frac{|S(\beta) \cap \partial S(\alpha)|}{|\partial S(\alpha)|} = \frac{|D(\beta \leftarrow \alpha)|}{|\partial S(\alpha)|} \quad (4)$$

For example, the ratio probability between the amino acids T and I can be obtained as follows:
S(I) (the codons that codes I) = AUU, AUC, AUA.
S(T) (the codons that codes T) = ACU, ACC, ACA, ACG.

$\partial S(I)$ = AUG(3), CUA, CUC, GUA, UUA, AAA, ACA, AGA, GUC, UUC, AAC, ACC, AGC, CUU, GUU, UUU, AAU, ACU, AGU.

Thus $|\partial S(I)| = 21$ and $|S(T) \cap \partial S(I)| = 3$.
Therefore,

$$d(I, T) = \frac{|S(T) \cap \partial S(I)|}{|\partial S(I)|} = \frac{3}{21} = 0.14$$

In this way the ratio probability between each pair of amino acids are obtained as shown in Table 1.

Now for defining the accessibility of β from α we fix some threshold value $\varepsilon > 0$.
Hence, $\beta \leftarrow \alpha$ iff $|D(\beta \leftarrow \alpha)| < \varepsilon |\partial S(\alpha)|$.

For example, let us choose ε= 0.11 and the codon AUG. The codon AUG code for the amino acid Methionine (M). The vicinity of the codon AUG is {AUG, CUG, GUG, UUG, AAG, ACG, AGG, AUA, AUC, AUU}. The vicinity of the amino acid M is {V, T, K, R} and the codons within the vicinity of AUG codes for the amino acids L, V, K, T, R, I. So by applying definition (3) we observe that the GP-map is continuous at AUG.

Next, we consider another codon say GGA. The codon GGA code for the amino acid Glycine (G). The amino acids which lie within the proximity of the amino acid G are W, C, S, E, D. Here we noticed that the amino acids G and W are close but their coded codons GGA and UGG are not related. Since the vicinity of G is{GGA,GAA, GCA, GUA, GGC, GGG, GGU, UGA, CGA, GAA}. From table 1, it is observed that when ε = 0.1, the GP map is continuous at all codons coding for the amino acids F, W, M, C, Y, H, E, N, Q, D whereas it is discontinuous at all codons coding for I, L, A, G, P, T, S, K, V, R.

Also, when ε = 0.12, , the GP map is continuous at all codons coding for the amino acids F, C, Y, H, E, N, Q, D whereas it is discontinuous at all codons coding for I, L, W, M, A, G, P, T, S, K, V, R. And for ε≥0.13, the map is discontinuous at all points. From this we can conclude that the continuity of the GP-map depends upon the threshold chosen. What value of the threshold is biologically realistic/natural is a matter of debate and discussion.

	I	F	V	L	W	M	A	G	C	Y	P	T	S	H	E	N	Q	D	K	R
I	0.00	0.10	0.14	0.19	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.14	0.10	0.00	0.00	0.10	0.00	0.00	0.05	0.05
F	0.13	0.00	0.13	0.38	0.00	0.00	0.00	0.00	0.13	0.13	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V	0.13	0.08	0.00	0.25	0.00	0.04	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.08	0.00	0.08	0.00	0.00
L	0.11	0.17	0.17	0.00	0.03	0.06	0.00	0.00	0.00	0.11	0.00	0.06	0.06	0.00	0.00	0.06	0.00	0.00	0.00	0.11
W	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.11	0.22	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.22
M	0.33	0.00	0.11	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11
A	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G	0.00	0.00	0.17	0.00	0.04	0.00	0.17	0.00	0.08	0.00	0.00	0.00	0.08	0.00	0.08	0.00	0.00	0.08	0.00	0.25
C	0.00	0.13	0.00	0.00	0.13	0.00	0.00	0.13	0.00	0.13	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.13	0.00	0.13
Y	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.13	0.13	0.00	0.13	0.00	0.13	0.00	0.13	0.00	0.00
P	0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.17	0.17	0.08	0.00	0.00	0.08	0.00	0.00	0.17
T	0.13	0.00	0.00	0.00	0.00	0.04	0.17	0.00	0.00	0.00	0.17	0.00	0.25	0.00	0.00	0.08	0.00	0.00	0.08	0.08
S	0.05	0.05	0.00	0.05	0.03	0.00	0.10	0.05	0.10	0.05	0.10	0.15	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.15
H	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.13	0.13	0.00	0.00	0.00	0.13	0.25	0.13	0.00	0.00	0.13
E	0.00	0.00	0.13	0.00	0.00	0.00	0.13	0.13	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.13	0.25	0.13	0.00	0.00
N	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.13	0.13	0.13	0.00	0.00	0.13	0.25	0.00	0.00
Q	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.25	0.13	0.00	0.00	0.00	0.13	0.13
D	0.00	0.00	0.13	0.00	0.00	0.00	0.13	0.13	0.00	0.13	0.00	0.00	0.13	0.00	0.25	0.13	0.00	0.00	0.00	0.00
K	0.06	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.13	0.25	0.13	0.00	0.00	0.13
R	0.03	0.00	0.00	0.11	0.06	0.03	0.00	0.17	0.06	0.00	0.11	0.06	0.17	0.06	0.00	0.00	0.06	0.00	0.05	0.00

Table 1: The ratio probability between each pair of amino acids obtained from equation (4).

Next by using the two frequencies, namely, neighbourhood frequency (equation (1)) and occurrence frequency (equation (2)), the accessibility measure between each pair of amino acids is obtained in table 2 and table 3 respectively.

From Table 2 and Table 3, it is observed that when ε = 0.07(for structure 2) and ε = 0.67 (for structure 3), the GP map is continuous at all codons coding for the amino acids F, W, M, C, Y, H, E, N, Q, D whereas it is discontinuous at all codons coding for I, L, A, G, P, T, S, K, V, R.

Also, when ε = 0.11 (for structure 2) and ε = 1 (for structure 3) or more, the GP map is discontinuous at all points.

It is observed that in all the three structures, if the average of all non-zero frequencies between the amino acids is considered as threshold εthen the GP map becomes discontinuous at all codons.

	I	F	V	L	W	M	A	G	C	Y	P	T	S	H	E	N	Q	D	K	R
I	0.22	0.07	0.11	0.15	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.07	0.00	0.00	0.07	0.00	0.00	0.04	0.04
F	0.11	0.11	0.11	0.33	0.00	0.00	0.00	0.00	0.11	0.11	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V	0.08	0.06	0.33	0.17	0.00	0.03	0.11	0.11	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.06	0.00	0.00	0.00
L	0.07	0.11	0.11	0.33	0.02	0.04	0.00	0.00	0.00	0.00	0.07	0.00	0.04	0.04	0.00	0.04	0.00	0.00	0.00	0.07
W	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.11	0.22	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22
M	0.33	0.00	0.11	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11
A	0.00	0.00	0.11	0.00	0.00	0.03	0.00	0.33	0.11	0.00	0.00	0.11	0.11	0.11	0.00	0.00	0.06	0.00	0.00	0.11
G	0.00	0.00	0.11	0.00	0.00	0.03	0.11	0.33	0.06	0.00	0.00	0.00	0.06	0.00	0.06	0.00	0.00	0.06	0.00	0.17
C	0.00	0.11	0.00	0.00	0.11	0.00	0.00	0.00	0.11	0.11	0.11	0.00	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.11
Y	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.00	0.00	0.11	0.00	0.11	0.00	0.11	0.00	0.00	0.00
P	0.00	0.00	0.00	0.11	0.00	0.00	0.11	0.00	0.00	0.00	0.33	0.11	0.11	0.00	0.00	0.00	0.06	0.00	0.00	0.11
T	0.08	0.00	0.00	0.00	0.00	0.00	0.03	0.11	0.00	0.00	0.00	0.11	0.33	0.17	0.00	0.00	0.06	0.00	0.00	0.06
S	0.04	0.04	0.00	0.04	0.02	0.00	0.07	0.04	0.07	0.04	0.07	0.11	0.26	0.00	0.00	0.04	0.00	0.00	0.00	0.11
H	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.11	0.11	0.00	0.00	0.11	0.00	0.11	0.00	0.11	0.22	0.00	0.11
E	0.00	0.00	0.11	0.00	0.00	0.00	0.11	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.11	0.22	0.11	0.00	0.00
N	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.11	0.11	0.11	0.00	0.11	0.00	0.22	0.00
Q	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.22	0.11	0.00	0.11	0.00	0.11	0.11
D	0.00	0.00	0.11	0.00	0.00	0.00	0.11	0.11	0.00	0.00	0.11	0.00	0.00	0.11	0.22	0.11	0.00	0.13	0.00	0.00
K	0.06	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.11	0.22	0.11	0.00	0.11
R	0.02	0.00	0.00	0.07	0.04	0.02	0.00	0.11	0.04	0.00	0.07	0.04	0.11	0.04	0.00	0.00	0.04	0.00	0.04	0.33

Table 2: The neighbourhood frequency between each pair of amino acids obtained from equation (1).

	I	F	V	L	W	M	A	G	C	Y	P	T	S	H	E	N	Q	D	K	R
I	2.00	0.67	1.00	1.33	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.67	0.00	0.00	0.67	0.00	0.00	0.33	0.33
F	1.00	1.00	1.00	3.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V	0.75	0.50	3.00	1.50	0.00	0.25	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.00	0.00
L	0.67	1.00	1.00	3.00	0.17	0.33	0.00	0.00	0.00	0.00	0.67	0.00	0.33	0.33	0.00	0.33	0.00	0.00	0.00	0.67
W	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	2.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00
M	3.00	0.00	1.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
A	0.00	0.00	1.00	0.00	0.00	0.00	0.00	3.00	1.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.50	0.00	0.00	1.00
G	0.00	0.00	1.00	0.00	0.00	0.25	1.00	3.00	0.50	0.00	0.00	0.00	0.50	0.00	0.50	0.00	0.00	0.50	0.00	1.50
C	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Y	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00
P	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	3.00	1.00	1.00	0.50	0.00	0.00	0.50	0.00	0.00	1.00
T	0.75	0.00	0.00	0.00	0.00	0.00	0.25	1.00	0.00	0.00	0.00	1.00	3.00	1.50	0.00	0.00	0.50	0.00	0.00	0.50
S	0.33	0.33	0.00	0.33	0.17	0.00	0.67	0.33	0.67	0.33	0.67	1.00	2.33	0.00	0.00	0.33	0.00	0.00	0.00	1.00
H	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	2.00	0.00	1.00
E	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	2.00	1.00	0.00	0.00
N	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00	2.00	0.00
Q	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	2.00	1.00	0.00	1.00	0.00	1.00	1.00
D	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	2.00	1.00	0.00	1.00	0.00	0.00
K	0.50	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	2.00	1.00	0.00	1.00
R	0.17	0.00	0.00	0.67	0.33	0.17	0.00	1.00	0.33	0.00	0.67	0.33	1.00	0.33	0.00	0.00	0.33	0.00	0.33	3.00

Table 3: The occurrence frequency between each pair of amino acids obtained from equation (2).

The ratio probability, neighbourhood frequency and occurrence frequency varies over the intervals [0, 0.38], [0, 0.33], and [0, 3] respectively. For a comparative analysis we transformed each of the above intervals to [0, 1]. The corresponding transformed tables are given below.

	I	F	V	L	W	M	A	G	C	Y	P	T	S	H	E	N	Q	D	K	R
I	0.00	0.25	0.38	0.51	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.38	0.25	0.00	0.00	0.25	0.00	0.00	0.13	0.13
F	0.33	0.00	0.33	1.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V	0.33	0.22	0.00	0.67	0.00	0.11	0.44	0.44	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.22	0.00	0.00	0.00
L	0.30	0.44	0.44	0.00	0.07	0.15	0.00	0.00	0.00	0.00	0.30	0.00	0.15	0.15	0.00	0.15	0.00	0.00	0.00	0.30
W	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.30	0.59	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.59
M	0.89	0.00	0.30	0.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.30
A	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.44	0.44	0.44	0.00	0.00	0.22	0.00	0.00	0.44
G	0.00	0.00	0.44	0.00	0.11	0.00	0.44	0.00	0.22	0.00	0.00	0.00	0.22	0.00	0.22	0.00	0.00	0.22	0.00	0.67
C	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.33	0.00	0.33	0.33	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.33
Y	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.33	0.33	0.00	0.33	0.00	0.33	0.00	0.00	0.00
P	0.00	0.00	0.00	0.44	0.00	0.00	0.44	0.00	0.00	0.00	0.00	0.44	0.44	0.22	0.00	0.00	0.22	0.00	0.00	0.44
T	0.33	0.00	0.00	0.00	0.00	0.00	0.11	0.44	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.22	0.00	0.00	0.22	0.22
S	0.13	0.13	0.00	0.13	0.07	0.00	0.27	0.13	0.27	0.13	0.27	0.40	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.40
H	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.00	0.33
E	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.00
N	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.33	0.33	0.00	0.33	0.00	0.33	0.67	0.00
Q	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.67	0.33	0.00	0.33	0.33	0.33
D	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.33	0.67	0.33	0.00	0.33	0.00
K	0.17	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.11	0.33
R	0.07	0.00	0.00	0.30	0.15	0.07	0.00	0.44	0.15	0.00	0.30	0.15	0.44	0.15	0.00	0.00	0.15	0.00	0.15	0.00

Table 4: The ratio probability between each pair of amino acids in the interval [0, 1].

	I	F	V	L	W	M	A	G	C	Y	P	T	S	H	E	N	Q	D	K	R
I	0.67	0.22	0.33	0.44	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.22	0.00	0.00	0.22	0.00	0.00	0.11	0.11
F	0.33	0.33	0.33	1.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V	0.25	0.17	1.00	0.50	0.00	0.08	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00
L	0.22	0.33	0.33	1.00	0.06	0.11	0.00	0.00	0.00	0.00	0.22	0.00	0.11	0.11	0.00	0.11	0.00	0.00	0.00	0.22
W	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.67	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67
M	1.00	0.00	0.33	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.33
A	0.00	0.00	0.33	0.00	0.00	0.00	1.00	0.33	0.00	0.00	0.33	0.33	0.33	0.00	0.17	0.00	0.00	0.17	0.00	0.00
G	0.00	0.00	0.33	0.00	0.08	0.00	0.33	1.00	0.17	0.00	0.00	0.00	0.17	0.00	0.17	0.00	0.00	0.17	0.00	0.50
C	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.33	0.33	0.33	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.33
Y	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.33	0.00	0.33	0.00	0.33	0.00	0.00	0.00
P	0.00	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	1.00	0.33	0.33	0.17	0.00	0.00	0.17	0.00	0.00	0.33
T	0.25	0.00	0.00	0.00	0.00	0.08	0.33	0.00	0.00	0.00	0.33	1.00	0.50	0.00	0.00	0.17	0.00	0.00	0.17	0.17
S	0.11	0.11	0.00	0.11	0.06	0.00	0.22	0.11	0.22	0.11	0.22	0.33	0.78	0.00	0.00	0.11	0.00	0.00	0.00	0.33
H	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.00	0.33
E	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.00
N	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.33	0.33	0.00	0.33	0.00	0.33	0.67	0.00
Q	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.67	0.33	0.00	0.33	0.00	0.33	0.33
D	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.33	0.67	0.33	0.00	0.33	0.00
K	0.17	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.33	0.67	0.33	0.00	0.33	0.33
R	0.06	0.00	0.00	0.22	0.11	0.06	0.00	0.33	0.11	0.00	0.22	0.11	0.33	0.11	0.00	0.00	0.11	0.00	0.11	1.00

Table 5: The neighbourhood frequency between each pair of amino acids in the interval [0,1].

	I	F	V	L	W	M	A	G	C	Y	P	T	S	H	E	N	Q	D	K	R
I	0.67	0.22	0.33	0.44	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.22	0.00	0.00	0.22	0.00	0.00	0.11	0.11
F	0.33	0.33	0.33	1.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
V	0.25	0.17	1.00	0.50	0.00	0.08	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00
L	0.22	0.33	0.33	1.00	0.06	0.11	0.00	0.00	0.00	0.00	0.22	0.00	0.11	0.11	0.00	0.00	0.11	0.00	0.00	0.22
W	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.67	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67
M	1.00	0.00	0.33	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.33
A	0.00	0.00	0.33	0.00	0.00	0.00	1.00	0.33	0.00	0.00	0.33	0.33	0.33	0.00	0.17	0.00	0.00	0.17	0.00	0.50
G	0.00	0.00	0.33	0.00	0.08	0.00	0.33	1.00	0.17	0.00	0.00	0.00	0.17	0.00	0.17	0.00	0.00	0.17	0.00	0.50
C	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.33	0.33	0.33	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.33
Y	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.33	0.00	0.33	0.00	0.33	0.00	0.00	0.00
P	0.00	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	1.00	0.33	0.33	0.17	0.00	0.00	0.17	0.00	0.00	0.33
T	0.25	0.00	0.00	0.00	0.00	0.08	0.33	0.00	0.00	0.00	0.33	1.00	0.50	0.00	0.00	0.17	0.00	0.00	0.17	0.17
S	0.11	0.11	0.00	0.11	0.06	0.00	0.22	0.11	0.22	0.11	0.22	0.33	0.78	0.00	0.00	0.11	0.00	0.00	0.00	0.33
H	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.00	0.33
E	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.67	0.33	0.00
N	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.33	0.33	0.00	0.33	0.00	0.33	0.67	0.00
Q	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.67	0.33	0.00	0.33	0.00	0.33	0.33
D	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.33	0.00	0.00	0.33	0.00	0.00	0.33	0.67	0.33	0.00	0.33	0.00
K	0.17	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.33	0.67	0.33	0.00	0.33	0.33
R	0.06	0.00	0.00	0.22	0.11	0.06	0.00	0.33	0.11	0.00	0.22	0.11	0.33	0.11	0.00	0.00	0.11	0.00	0.11	1.00

Table 6: The occurrence frequency between each pair of amino acids in the interval [0, 1].

From Table 5 and Table 6, it is being observed that the neighbourhood frequency and the occurrence frequency are same between each pair of amino acids. Also, the ratio probability (Table 4) is highly correlated (0.94) with the neighbourhood frequency as well as the occurrence frequency. Thus we conclude that all the three accessibility structures are equivalent.

Fig 1: percentage of continuous codons against the threshold value.

From Table 4 (equivalently from table 5 and table 6) we examine continuity of codons for different threshold values (ε). The graph of percentage codons at which GP-map is continuous against threshold values is shown in figure 1.

For a fixed threshold (ε), we get a binary relation in the amino acids set which induces a graph of the amino acids. With respect to this ε, the graph gives a pictorial representation of the continuous evolution of the amino acids. That is to say, evolution of one amino acid from another is continuous if there is an edge connecting them. Below we consider such a graph with $\varepsilon = \dfrac{1}{4}$
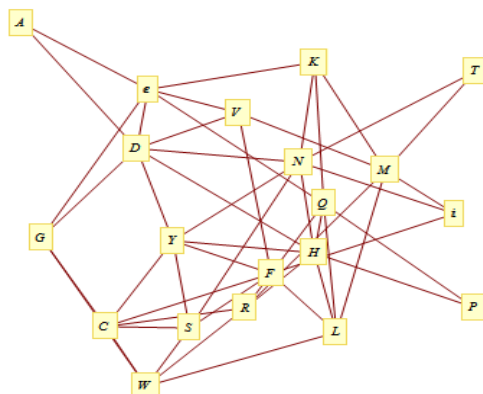

Fig 2: Graph of the evolution of amino acids w.r.t. the threshold $\varepsilon = \dfrac{1}{4}$

The above graph shows that evolution of amino acid V from T is discontinuous as they are not connected to each other by an edge, whereas evolution of F from C or vice versa is a continuous one as there exists an edge connecting them. If the value ε is changed it is obvious that the scenario of continuity of the codons will also change. So we can say that this notion of continuity is not an absolute concept but varies depending upon the neighbourhood structure.

IV. CONCLUSION

In this paper we have examined the continuity of the GP-map from codons to amino acids. Three approaches were employed and we observed that they are all equivalent. Neighbourhood structure is defined in the amino acid set in terms of vicinity which in turn depends upon the threshold value chosen; the choice of threshold value is open for debates and discussion. However once a threshold value is agreed upon, we obtain a graph of amino acids which depicts their continuous evolution.

According to evolutionary importance in genetic code, the second base of a codon is considered as biologically most significant base, whereas third base is least significant base in a codon. This evolutionary importance of base positions is connected with the accessibility measure between the amino acids. The accessibility relation gives us that how much an amino acid is accessible from the other. From Table 4 (Table 5 or Table 6) it is observed that the pair of amino acids which are accessible from one other with high frequency (more or equal to 0.5) are coded by codons having same base in the second base position. Also, for the amino acids coded by only two codons accessible from one another with high frequency have codons with same base in the first and second base position and transition occurs in the third base. Also, the amino acids, accessible from one other with high frequency (more or equal to 0.5) have same hydrophobic-hydrophilic property.

REFERENCES

1. W. Fontana, P. Schuster, Continuity in Evolution: On the Nature of Transitions, Science 280, 1998, pp1451-1455.
2. P. Alberch, From genes to phenotype: dynamical systems and evolvability. Genetica 84, 1991, pp 5–11.
3. W. Fontana, Modeling evo-devo with RNA. BioEssays 24, 2004, pp1164–1177.
4. M.C. Cowperthwaite & L.A. Meyers, How mutational networks shape evolution: lessons from RNA models. Ann. Rev. Ecol. Evol. Syst. 38, 2007, pp203–230.
5. P. Ferna´ndez & R.V. Sole, Neutral fitness landscapes insignalling networks. J. R. Soc. Interface 4, 2007, pp41–47.
6. Sumedha, O.C. Martin & A. Wagner, New structural variation in evolutionary searches of RNA neutral networks. BioSystems 90, 2007, pp 475–485.
7. R. Wroe, H.S. Chan & E. Bornberg-Bauer, A structural model of latent evolutionary potentials underlying neutral networks in proteins. HFSP J. 1, 2007, pp 79–87.
8. M. Stich, C. Briones & S.C. Manrubia, On the structural repertoire of pools of short, random RNA sequences. J. Theor. Biol. 252, 2008, pp750–763.
9. N. Takeuchi & P. Hogeweg, Evolution of complexity in RNA-like replicator systems. Biol. Direct. 3, 2008, pp11.
10. M. Pigliucci, G.B. Muller(eds), Evolution: the extended synthesis. Cambridge, MA: MIT Press. 2010.
11. P. Stadler, B. Stadler, Genotype phenotype maps. Biol Theory 3, 2006, pp268–279.
12. W. Fontana, T. Griesmacher, W. Schnabl, P.F. Stadler, P. Schuster, Statistics of landscapes based on free energies, replication and degradation rate constants of RNA secondary structures. Monatsh. Chem., 122, 1991, pp795-819.
13. W. Fontana, D.A.M Konings, P. Schuster, Statistics of RNA secondary structures. Biopolymers, 33, 1993, pp 1389-1404.
14. W.Fontana, P. Schuster, Shaping Space: The possible and the Attainable in RNA Genotype-Phenotype Mappings, J. Theor. Biol. 194, 1998, pp 491-515.
15. J. Cupal, S. Kopp and P.F. Stadler: RNA shape space topology, Alife 6, 2000, pp 3-23.

AUTHOR PROFILE

Dr. Nisha Gohain
Assistant Professor
Directorate of Open and Distance Learning
Dibrugarh University
Assam, India.
Email: gohainnisha@gmail.com

Reasearch Area: Mathematical Biology, Algebra, Topology.
Publications:

(i) Analysis of amino acids network based on distance matrix: Physica A, 452, 69–78, 2016.**[ISSN 0378-4371]**

(ii) Lattice Structure and Distance Matrix of Genetic Code, Journal of Biological Systems, 23(3), 485-504, 2015. **[ISSN 1793-6470]**

(iii) Reducing Redundancy of Codons Through Total Graph: American Journal of Bioinformatics (Science Publications), 4, doi : 10.3844/ajbsp.2015.1.6, 2015. **[ISSN 1948-9870]**

(iv) Graph Theoretic Approach to Analyze Amino Acid Network: International Journal of Advances in Applied Mathematics and Mechanics, 2(3), 31-37, 2015. **[ISSN 2347-2529]**

(v) Centrality Measures in Graph of Amino Acids: Mathematical Forum, XVI, 74-87, 2014. **[ISSN 0972-9852]**

(vi) Graph in Genetic Code Algebra: Journal of Bioinformatics and Intelligent Control (American Scientific Publisher), 4(1), 1-4, 2015. **[ISSN 2326-750X]**