Article

A study of the total graph in genetic code algebra

Birinchi Kumar Boruah, **Tazid Ali**
Department of Mathematics, Dibrugarh University, Assam 786004, India
E-mail: birinchikumar49@gmail.com, tazid@dibru.ac.in

Received 20 September 2021; *Accepted 30 October 2021; Published 1 March 2022*

Abstract
Suppose R be a commutative ring and $Z(R)$ its set of zero-divisors. Total graph is the (undirected) graph where set of all elements of R is taken as the vertex set and two vertices say x and y ($x \neq y$) in R are adjacent if and only if their sum is zero-divisor. Genetic code is the blueprint for protein synthesis. In this paper we discuss total graph in the genetic code algebra.

Keywords genetic code; amino acid; mutation; total 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

Genetic code is the rule defining the sequence of nucleotides in DNA (Deoxyribonucleic acid) or mRNA (Messenger ribonucleic acid) that determines the specific amino acid sequence in the synthesis of protein. The DNA consists of two complementary long chains of nucleotides, viz. Adenine (A), Cytosine (C), Guanine (G) or Thymine (T) (Uracil (U) in case of mRNA). Deoxyribonucleic acid (DNA) stores genetic information about how to construct or synthesize proteins. Amino acids are the building blocks of proteins. Each amino acid is a triplet code of four possible bases (nucleotides). Three consecutive DNA nucleotides form a codon. Each codon specifies a particular amino acid. Mathematically DNA can be considered as a sequence of four letters that is A, G, C and U (or T). As there are four bases, this gives us 64 codons. In the evolutionary importance of genetic code, the second base is considered as biologically most significant base, where as third base is least significant base in a codon. Different kinds of mutations are possible in codons namely, point mutation, deletion, insertion, inversion. In this paper we will consider only the case of point mutation. In case of a point mutation, there is a simple change in one base of the gene sequence. It replaces a single base nucleotide with another nucleotide of the genetic material, DNA or mRNA. 64 codons make up the genetic code, though there are only 20 amino acids. This means that there are some overlap i.e., more than one codon code for the same amino acid. The codons that code for the same amino acids are known as synonymous codons. We can consider this as a function of many to one carrying codons to amino acids.

Various authors (Bashford et al., 1998; Bashford and Jarvis, 2000; Beland and Allen, 1994; Siemion et al.,

1995; Antoneli et al., 2003 and others) worked on this field and tried to give some algebraic formulation of the structure of the genetic code. Sanchez et al. (2005) brought a new idea for describing the quantitative relationship between DNA genomic sequences through algebraic structures.

Various types of graph structures can be introduced corresponding to a given algebraic structure. Cameron and Ghosh (2011) introduced the power graph of a finite group. The power graph of a group is the graph whose vertex set is the group and two elements being adjacent if one is a power of the other. Bertholf et al. (1978) introduced graphs of finite abelian groups whose vertices are in one-to-one correspondence with the non-identity subgroups of G and two vertices are joined by an edge if and only if the corresponding subgroups intersect. In Anderson and Badawi, 2008 introduced the total graph of a commutative ring and denoted by T(Γ(R)), where R is the commutative ring. It is the (undirected) graph with all elements of R as vertices and for distinct $x, y \in R$, the vertices x and y are adjacent if $x + y \in Z(R)$, where $Z(R)$ is the set of zero-divisors of R. They also discussed the three (induced) subgraphs namely, *Nil*(Γ(R)), Z(Γ(R)), and *Reg* (Γ(R)) of T(Γ(R)), with vertices *Nil*(R), Z(R) and *Reg*(R), respectively, where *Nil*(R) is the ideal of nilpotent elements of R and *Reg*(R) is the set of regular elements of R. Beck (1988) introduced the concept of the graph of the zero divisors of R, where he was mainly interested in colorings. The investigation of colorings of a commutative ring was then continued by Anderson and Naseer (1993). Anderson and Livingston (1999) associate a graph, Γ(R), to R with vertices $Z(R)\backslash\{0\}$, where Z (R) is the set of zero divisors of R and for distinct $x, y \in Z(R)\backslash\{0\}$, the vertices x and y are adjacent if and only if $xy = 0$. Akbari et al. (2009) proved that the total graph is a Hamiltonian graph if it is connected.

The vital interest of this work is for analyzing mathematical structures viz., graph structures that may naturally occur in genetic code.

2 Graphical Depiction of Genetic Code Algebra & Mapping Among Their Subgraphs

Sanchez et al. (2005) observed that the four RNA (or DNA) bases can be arranged or ordered considering the codon- anticodon interactions between them. The hydrogen bond number and the chemical type (purine and pyrimidine) of bases play an important role on this. From which two orders of the base sets: {A, C, G, U} and {U, G, C, A} are obtained. They discussed sum operation of the bases obtained from the above two possible orders ({A, C, G, U} and {U, G, C, A}) which makes the two sets isomorphic to Z_4. Here, Tables 1(A) and (B) represents the sum operation of the bases obtained from the two possible orders. Further they defined sum and product operation on the set of codons. It was observed that the group obtained on the set of codons is isomorphic to the group of integer module 64, (Z_{64}, +). These two sum & product operation on the set of whole codons represents a ring structure isomorphic to the ring of (Z_{64}, +, .). In 2015, Akhtar et al. discussed the total graph of this ring structure Z_{64}. By taking the ordered base set {A, C, G, U} isomorphic Z_4 from Sanchez et al., 2005, we observe that $Z_4 \times Z_4 \times Z_4$ forms a ring structure isomorphic to the ring of $(Z_4 \times Z_4 \times Z_4, +, .)$.

In this paper it is being found that some special types of mutation in codons partitioned the whole set of codons into eight disjoint sets which gives eight disjoint graphs, all are individually connected.

In the ring $Z_4 \times Z_4 \times Z_4$, $Z(Z_4 \times Z_4 \times Z_4) = \{AAG, AGG, GAA, GAG, AGA, GGG, GGA\}$ is the set of zero divisors. Now we construct the total graph of the codon set. We use base set {A, C, G, U} corresponding to {0, 1, 2, 3}. Since $Z_4 = \{0, 1, 2, 3\}$ and with respect to $Z_4 = \{0, 1, 2, 3\}$ we have

$$Z_4 \times Z_4 \times Z_4 = \{A, C, G, U\} \times \{A, C, G, U\} \times \{A, C, G, U\}$$
$$= \{GGC, AGC, AAC, GAC, AGU, AAU, GAU, GGU, GGG, AAG, GAA, GAG, AGG, GGA, AGA, AAA,$$
$$UAA, UGG, CGA, CGG, CAA, UAG, CAG, UGA, ACC, GCC, AUC, GUU, AUU, ACU, GCU, GUC,$$
$$CUC, UUU, UCC, CUU, CCC, UUC, CCU, UCU, GUA, ACA, AUG, GUG, ACG, GCA, AUA, GCG,$$

$CAU, UGC, CAC, CGC, UAU, CGU, UAC, UGU, CCA, CUG, UUG, CUA, UCA, UUA, CCG, UCG\}$

And

$AAA + AAC = AAC$ & $AAG + UCG = UCA$ and so on.

And the base set $\{A, C, G, U\}$ analogous to {0, 1, 2, 3} we can obtain the total graph of $Z_4 \times Z_4 \times Z_4$ as shown in figure -1. In the total graph of $Z_4 \times Z_4 \times Z_4$ (fig-1), we observe that there are eight disjoint graphs. So,

$$G_1 = \{GGC, AGC, AAC, GAC, AGU, AAU, GAU, GGU\}$$
$$G_2 = \{GGG, AAG, GAA, GAG, AGG, GGA, AGA, AAA\}$$
$$G_3 = \{UAA, UGG, CGA, CGG, CAA, UAG, CAG, UGA\}$$
$$G_4 = \{CAU, UGC, CAC, CGC, UAU, CGU, UAC, UGU\}$$
$$G_5 = \{ACC, GCC, AUC, GUU, AUU, ACU, GCU, GUC\}$$
$$G_6 = \{GUA, ACA, AUG, GUG, ACG, GCA, AUA, GCG\}$$
$$G_7 = \{CUC, UUU, UCC, CUU, CCC, UUC, CCU, UCU\}$$
$$G_8 = \{CCA, CUG, UUG, CUA, UCA, UUA, CCG, UCG\}$$

The codons from 000 to 303 (codons XAZ) codes the most hydrophilic amino acids and the codons from 030 to 333 (codons XUZ) codes the most hydrophobic amino acids (Table 2).

Table 1 Sum operation tables defined in the set of four bases of the DNA.

A: Primal algebra

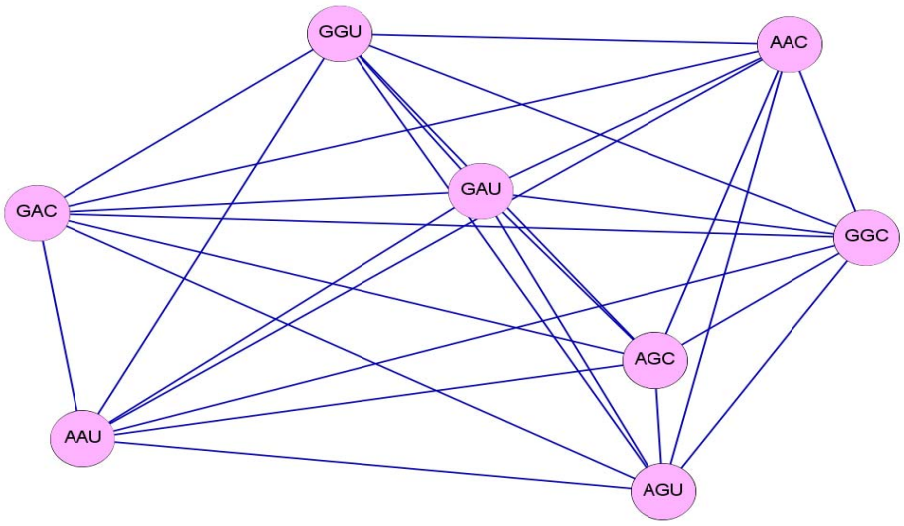
+	A	C	G	U
A	A	C	G	U
C	C	G	U	A
G	G	U	A	C
U	U	A	C	G

B: Dual algebra

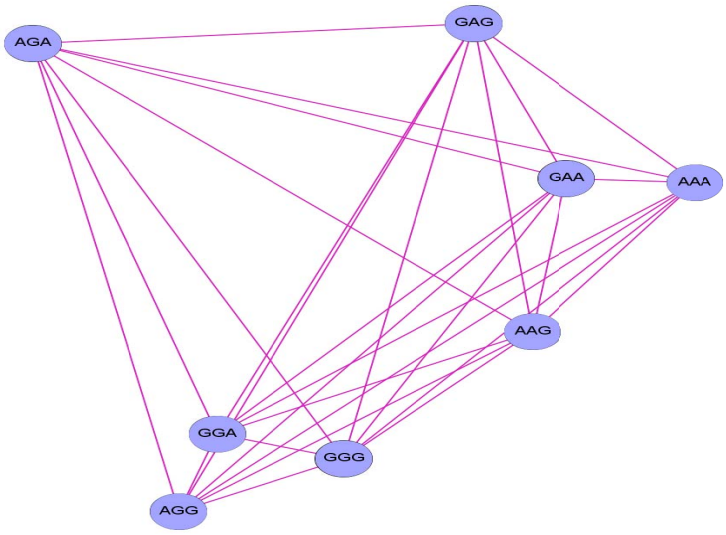
+	U	G	C	A
U	U	G	C	A
G	G	C	A	U
C	C	A	U	G
A	A	U	G	C

Table 2 The genetic code table induced by the order {A, C, G, U}.

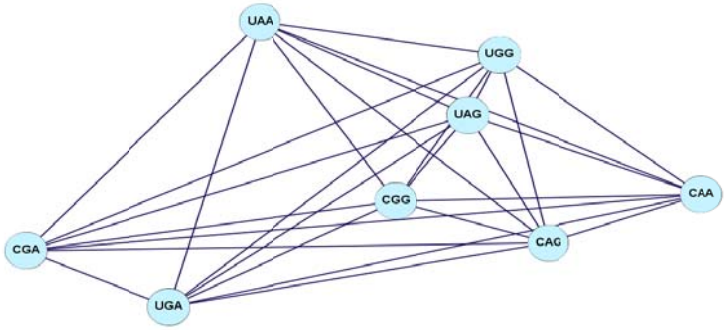
	A			C			G			U			
	No	Codon	Amino Acid	No	Codon	Amino Acid	No	Codon	Amino Acid	No	Codon	Amino Acid	
A	000	AAA	K	010	ACA	T	020	AGA	R	030	AUA	I	A
	001	AAC	N	011	ACC	T	021	AGC	S	031	AUC	I	C
	002	AAG	K	012	ACG	T	022	AGG	R	032	AUG	M	G
	003	AAU	N	013	ACU	T	023	AGU	S	033	AUU	I	U
C	100	CAA	Q	110	CCA	P	120	CGA	R	130	CUA	L	A
	101	CAC	H	111	CCC	P	121	CGC	R	131	CUC	L	C
	102	CAG	Q	112	CCG	P	122	CGG	R	132	CUG	L	G
	103	CAU	H	113	CCU	P	123	CGU	R	133	CUU	L	U
G	200	GAA	E	210	GCA	A	220	GGA	G	230	GUA	V	A
	201	GAC	D	211	GCC	A	221	GGC	G	231	GUC	V	C
	202	GAG	E	212	GCG	A	222	GGG	G	232	GUG	V	G
	203	GAU	D	213	GCU	A	223	GGU	G	233	GUU	V	U
U	300	UAA	-	310	UCA	S	320	UGA	-	330	UUA	L	A
	301	UAC	Y	311	UCC	S	321	UGC	C	331	UUC	F	C
	302	UAG	-	312	UCG	S	322	UGG	W	332	UUG	L	G
	303	UAU	Y	313	UCU	S	323	UGU	C	333	UUU	F	U



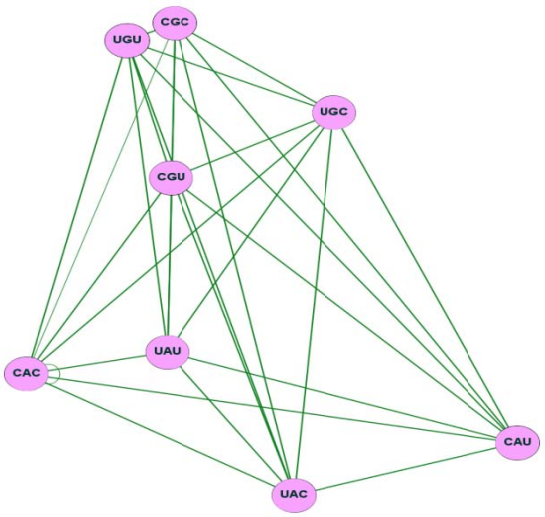
G_1


GAG
AGA
GAA
AAA
AAG
GGA
GGG
AGG

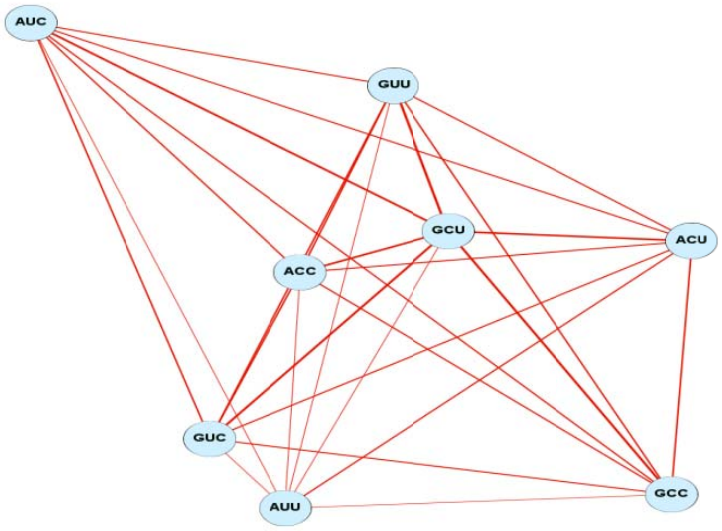
G_2


UAA
UGG
UAG
CGG
CAA
CGA
CAG
UGA

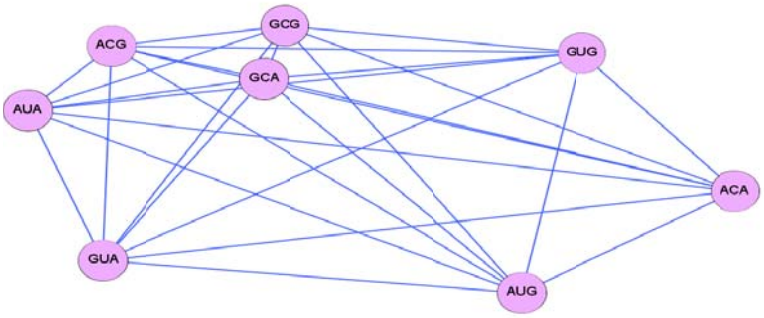
G_3


CGC
UGU
UGC
CGU
UAU
CAC
CAU
UAC

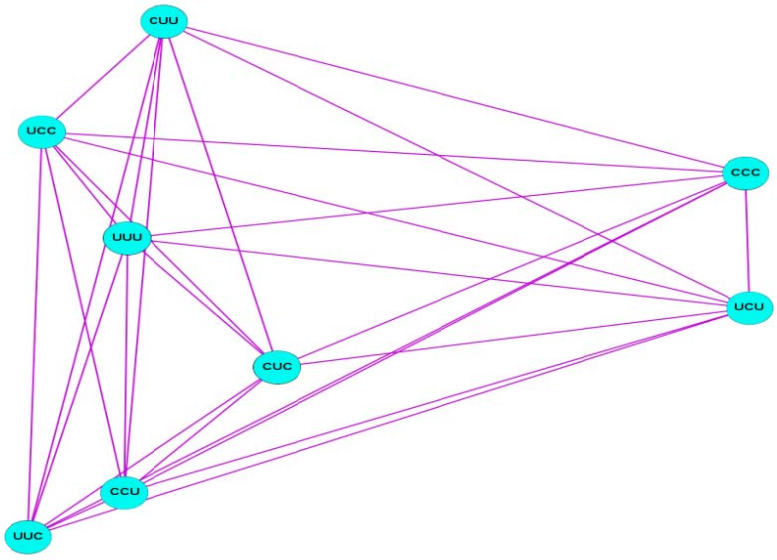
G_4


AUC
GUU
GCU
ACU
ACC
GUC
AUU
GCC

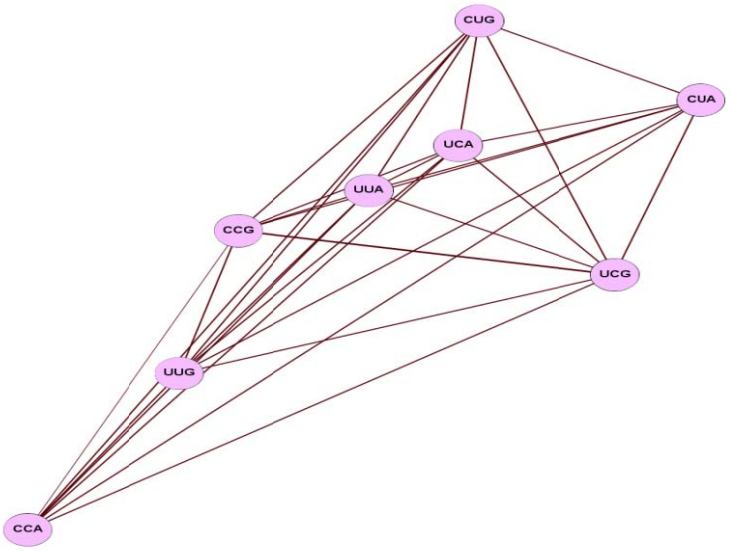
G_5


GCG
ACG
GUG
GCA
AUA
ACA
GUA
AUG

G_6


CUU
UCC
CCC
UUU
UCU
CUC
CCU
UUC

G_7



G_8

Fig. 1 Total graph of $Z_4 \times Z_4 \times Z_4$.

We may note in passing that the vertices of G_2 form a group under addition.

We can define a function p_1 from the set of vertices of G_1 to the set of vertices to G_2 given by $p_1 : G_1 \to G_2$ such that

$$p_1(LMN) = \begin{cases} LMG & if\ N = C \\ LMA & if\ N = U \end{cases} \quad \forall\ LMN \in G_1$$

Again, we can define a function $p_1^{/}$ from the set of vertices of G_2 to the set of vertices to G_1 given by $p_1^{/} : G_2 \to G_1$ such that

$$p_1^{/}(LMN) = \begin{cases} LMC & if\ N = G \\ LMU & if\ N = A \end{cases} \quad \forall\ LMN \in G_2$$

Yet again, we can define a function p_2 from the set of vertices of G_3 to the set of vertices to G_4 given by $p_2 : G_3 \to G_4$ such that

$$p_2(LMN) = \begin{cases} LMU & if\ N = A \\ LMC & if\ N = G \end{cases} \quad \forall\ LMN \in G_3$$

Furthermore, we can define a function $p_2^{/}$ from the set of vertices of G_4 to the set of vertices to G_3 given by $p_2^{/} : G_4 \to G_3$ such that

$$p_2^{/}(LMN) = \begin{cases} LMA & if\ N = U \\ LMG & if\ N = C \end{cases} \quad \forall\ LMN \in G_4$$

We can also define a function p_3 from the set of vertices of G_5 to the set of vertices to G_6 given by $p_3 : G_5 \to G_6$ such that

$$p_3(LMN) = \begin{cases} LMA & if\ N = C \\ LMG & if\ N = U \end{cases} \quad \forall\ LMN \in G_5$$

Also, the function $p_3^{/}$ from the set of vertices of G_6 to the set of vertices to G_5 given by $p_3^{/} : G_6 \to G_5$ and defined as

$$p_3^{/}(LMN) = \begin{cases} LMC & if\ N = A \\ LMU & if\ N = G \end{cases} \quad \forall\ LMN \in G_6$$

We can define a function p_4 from the set of vertices of G_7 to the set of vertices to G_8 given by $p_4 : G_7 \to G_8$ such that

$$p_4(LMN) = \begin{cases} LMG & if\ N = C \\ LMA & if\ N = U \end{cases} \quad \forall\ LMN \in G_7$$

We can define a function $p_4^{/}$ from the set of vertices of G_8 to the set of vertices to G_7 given by $p_4^{/} : G_8 \to G_7$ such that

$$p_4^{/}(LMN) = \begin{cases} LMC & if\ N = G \\ LMU & if\ N = A \end{cases} \quad \forall\ LMN \in G_8$$

This function p_1 can also be represented as $p_1(XYZ) = XYZ + GGU \quad \forall\ XYZ \in G_1$

Similarly, the functions p_2, p_3, p_4 can be defined as

$$p_2(XYZ) = XYZ + GGU \quad \forall\ XYZ \in G_3$$
$$p_3(XYZ) = XYZ + GGU \quad \forall\ XYZ \in G_5$$
$$p_4(XYZ) = XYZ + GGU \quad \forall\ XYZ \in G_7$$

There are seven other codons which can be used in case of *GGU* and they are *GGC*, *AGC*, *AAC*, *GAC*, *AGU*, *AAU* and *GAU*. Interestingly these eight codons constitute the set G_1.

Here it is clear that all the functions are bijective mappings. Also, the mappings p_i are inverse of $p_i^{/}$ (where i = 1, 2, 3, 4). From a biological point of view, it is observed that these functions represent the transversion mutation of the third base of a codon. Hence, we can conclude that the transversion of the third base of codons can be represented in terms of total graph of the genetic code.

Also, from our defined function it is clear that the transversion of the third base of all codons gives a bijective map.

Also, From the function $p_1 : G_1 \to G_2$, we observed that for any codon whose third base is pyrimidine, then under the function it changes to purine.

In $p_1^{/} : G_2 \to G_1$, we observed that for any codon whose third base is purine, then under the function it changes to pyrimidine.

In the function $p_2 : G_3 \to G_4$, we observed that for any codon whose third base is purine, then under the

function it changes to pyrimidine.

From the function $p_2^{/} : G_4 \to G_3$, we observed that for any codon whose third base is pyrimidine, then under the function it changes to purine.

From the function $p_3 : G_5 \to G_6$, we observed that for any codon whose third base is pyrimidine, then under the function it changes to purine.

From the function $p_3^{/} : G_6 \to G_5$, we observed that for any codon whose third base is purine, then under the function it changes to pyrimidine.

From the function $p_4 : G_7 \to G_8$, we observed that for any codon whose third base is pyrimidine, then under the function it changes to purine.

From the function $p_4^{/} : G_8 \to G_7$, we observed that for any codon whose third base is purine, then under the function it changes to pyrimidine.

And from the functions $p_1 : G_1 \to G_2$, $p_1^{/} : G_2 \to G_1$, $p_2 : G_3 \to G_4$, $p_2^{/} : G_4 \to G_3$, $p_4 : G_7 \to G_8$ and $p_4^{/} : G_8 \to G_7$, we observed that for any codon whose third base is strong (weak), then under the function it changes to strong (weak).

From the functions $p_3 : G_5 \to G_6$ and $p_3^{/} : G_6 \to G_5$, we observed that for any codon whose third base is strong (weak), then under the function it changes to weak (strong).

And from the functions $p_1 : G_1 \to G_2$, $p_1^{/} : G_2 \to G_1$, $p_2 : G_3 \to G_4$, $p_2^{/} : G_4 \to G_3$, $p_4 : G_7 \to G_8$ and $p_4^{/} : G_8 \to G_7$, we observed that for any codon whose third base is Amino (keto), then under the function it changes to Keto (Amino).

From the functions $p_3 : G_5 \to G_6$ and $p_3^{/} : G_6 \to G_5$, we observed that for any codon whose third base is Amino (keto) , then under the function it changes to Amino (keto).

Also, we observed that in the functions $p_1 : G_1 \to G_2$, $p_1^{/} : G_2 \to G_1$, $p_3 : G_5 \to G_6$, $p_3^{/} : G_6 \to G_5$, $p_4 : G_7 \to G_8$ and $p_4^{/} : G_8 \to G_7$, Non-polar amino acids are mapped to non-polar amino acids.

In the functions $p_1 : G_1 \to G_2$, $p_1^{/} : G_2 \to G_1$, $p_2 : G_3 \to G_4$ and $p_2^{/} : G_4 \to G_3$, charge amino acids are mapped to charge amino acids.

In the functions $p_2 : G_3 \to G_4$, $p_2^{/} : G_4 \to G_3$, $p_3 : G_5 \to G_6$, $p_3^{/} : G_6 \to G_5$, $p_4 : G_7 \to G_8$ and $p_4^{/} : G_8 \to G_7$, polar amino acids are mapped to polar amino acids. The stop codons belong to the graph G_3 and Start codons belong to the graph G_6.

3 Conclusion

Given an algebraic structure, various graph structures can be obtained from it. Here we discussed the total graph structure of the genetic code. We have noticed that the total graph structure partitions the set of codons into eight disjoint graphs. Also, Bijective mappings between each pair of graphs have been obtained. It is observed that each mapping & its inverse can be represented as translation mappings of eight types. We have observed that the property of purine/ pyrimidine, keto/ amino and strong/ weak of the third base of the codons play an important role in these mappings.

References

Akhtar A, Ali T, Gohain N. 2015. Graph in genetic code algebra. Journal of Bioinformatics and Intelligent Control, 4(1): 1-4

Anderson DF, Badawi A. 2008. The total graph of a commutative ring. Journal of Algebra, 320: 2706-2719

Antoneli F, Braggion L, Forger M, Hornos JEM. 2003. Extending the search for symmetries in the genetic code. International Journal of Modern Physics B, 17: 3135-3204

Anderson DD, Naseer M. 1993. Coloring of a commutative ring. Journal of Algebra, 159: 500-514

Anderson DF, Livingston PS. 1999. The zero-divisor graph of a commutative ring. Journal of Algebra, 217: 434-447

Akbari S, Kiani D, Mohammadi F, Moradi S. 2009. The total graph and regular graph of a commutative ring. Journal of Pure and Applied Algebra, 213: 2224-2228

Bashford JD, Jarvis PD. 2000. The genetic code as a periodic table: Algebraic aspects. Biosystems, 57: 147-161

Bashford JD, Tsohantjis I, Jarvis PD. 1998. A supersymmetric model for the evolution of the genetic code. Proceedings of the National Academy of Sciences USA, 95: 987-992

Beck I. 1988. Coloring of commutative rings. Journal of Algebra, 116: 208-226

Bertholf D, Walls G. 1978. Graph of finite abelian groups. Czechoslovak Mathematical Journal, 28: 365-368

Beland P, Allen T. 1994. The origin and evolution of the genetic code. Journal of Theoretical Biology, 170: 359-365

Cameron JP, Ghosh S. 2011. The power graph of a finite group. Discrete Mathematics, 311: 1220-1222

Gohain N, Ali T, Akhtar A. 2015. Reducing redundancy of codons through total graph. American Journal of Bioinformatics, 4: 1-6

Sanchez R, Morgado E, Grau R. 2005. Gene algebra from a genetic code algebraic structure. Journal of Mathematical Biology, 51: 431-457

Siemion IZ, Siemion PJ, Krajewski K. 1995. Chou-Fasman conformational amino acid parameters and the genetic code. Biosystems, 36: 231-238