Representation mutations from standard genetic codes

I Aisah*, M Suyudi, E Carnia, Suhendi, A K Supriatna

Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Universitas Padjadjaran Bandung, Indonesia

*Corresponding author: isah.aisah@unpad.ac.id

Abstract. Graph is widely used in everyday life especially to describe model problem and describe it concretely and clearly. In addition graph is also used to facilitate solve various kinds of problems that are difficult to be solved by calculation. In Biology, graph can be used to describe the process of protein synthesis in *DNA*. Protein has an important role for *DNA* (*deoxyribonucleic acid*) or *RNA (ribonucleic acid*). Proteins are composed of amino acids. In this study, amino acids are related to genetics, especially the genetic code. The genetic code is also known as the triplet or codon code which is a three-letter arrangement of *DNA* nitrogen base. The bases are adenine (*A*), thymine (*T*), guanine (*G*) and cytosine (*C*). While on *RNA* thymine (*T*) is replaced with Urasil *(U*). The set of all Nitrogen bases in *RNA* is denoted by *N* = {*C U, A, G*}. This codon works at the time of protein synthesis inside the cell. This codon also encodes the stop signal as a sign of the stop of protein synthesis process. This paper will examine the process of protein synthesis through mathematical studies and present it in three-dimensional space or graph. The study begins by analysing the set of all codons denoted by *NNN* such that to obtain geometric representations. At this stage there is a matching between the sets of all nitrogen bases *N* with $Z_2 \times Z_2$; $C = (\bar{0}, \bar{0}), \mathrm{U} = (\bar{0}, \bar{1}), \mathrm{A} = (\bar{1}, \bar{0}), \mathrm{G} = (\bar{1}, \bar{1}\ )$ .By matching the algebraic structure will be obtained such as group, group Klein-4,Quotien group etc. With the help of Geogebra software, the set of all codons denoted by *NNN* can be presented in a three-dimensional space as a multicube *NNN,* and also can be represented as a graph, so that can easily see relationship between the codon.

Keyword: *Codon, Graph , DNA, RNA*

1. Introduction

Genes are made up of nucleic acids called deoxyribonucleic acids (*DNA*). *DNA* forms *RNA* (ribonucleic acid) through the transcription process. *DNA*-forming molecules are pentose sugars (deoxyribose), phosphates, and nitrogenous bases comprising purines (guanine (*G*) and adenine (*A*)) and pyrimidine (thymine (*T*) and cytosine (*C*)). One of the differences between *DNA* and *RNA* lies in its pyrimidine base. *RNA* contains uracid pyrimidine (*U*) and cytosine (*C*).

The genetic code is the composition of three nitrogenous bases of nucleic acids (*DNA* and *RNA*) called codons, or standard genetic codes. The number of codons there are 64 pieces. Of these 64 codons, 61 codons encode amino acids while the remaining three (*UAA*, *UAG*, and *UGA*) encode a stop signal or as a signal indicating the end of the protein formation process.

The nitrogen base of *RNA* can be represented as the set *N = {C, U, A, G}.* The set of all these nitrogen bases can be matched with $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(\bar{0}, \bar{0}), (\bar{0}, \bar{1}), (\bar{1}, \bar{0}), (\bar{1}, \bar{1})\}$ [6]. In addition, the set of all nitrogen bases can also be matched with a polynomial ring over $\mathbb{Z}_2$. The matching of *N* with the polynomial ring over $\mathbb{Z}_2$causes *N* to form a four-element galois field denoted by *GF*(4) and *NNN* which is the set of all codons can be presented into a three-dimensional space as multicube *NNN* . Based on the corresponding transformation it is found that 24 representations of *NNN* [6]. Through this presentation it can be seen mutations that occur in the standard genetic code. Besides through multicube, its representation can be seen more easily through graph.

2. Theoretical Model

2.1. Graph

Definition 2.1: A graph *G* is defined as *G* = (*V, E*) Where *V* is a set of all vertices and *E* is a set of all edges in the graph.

Each side of the graph is associated with one or two vertices. These vertexes are called end vertices. The side associated with only one point is called a loop. Two different sides connecting the same vertices are called parallel sides. Two vertices are said to be connected or adjacent if there is a side connecting the two. If all the sides in a directed graph then the graph is called directed graph (directed graph, abbreviated digraph). Conversely, if all the sides in the graph are not directional, then the graph is called undirected graph.

2.2. Genetic

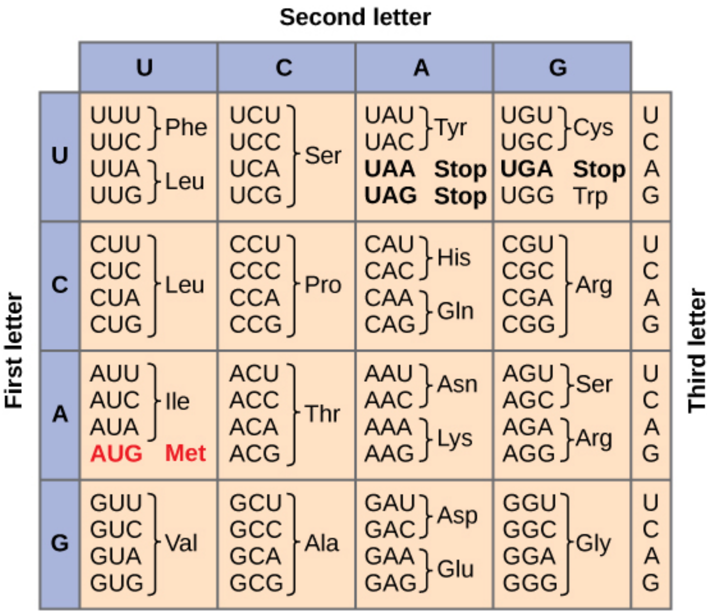
Deoxyribonucleic acids (*DNA*) are very important chemical compounds for living things. *DNA* consists of three kinds of molecules, namely:
1) The pentose sugar, known as deoxyribose
2) Posterior acid
3) Nitrogen base, distinguished from primidin consisting of cytosine (*C*) and thymine (*T*) and purine consisting of adenine (*A*) and guanine (*G*)

Beside *DNA*, there are other important nucleic acids, namely ribonucleic acid (*RNA*). Like *DNA*, *RNA* consists of three kinds of molecules, namely pentose sugars, posropicacid, and nitrogenous bases. Differences of *DNA* and *RNA* one of them lies in its primidin base. The pyrimidine content in *DNA* consists of cytosine and thymine while the pyrimidine base in *RNA* consists of cytosine and uracil (*U*). Based on location and function, *RNA* is divided into three kinds, namely:

1) *RNA* ambassadors (*RNA-d*)
RNA ambassadors or *RNA-d* are located within the nucleus and are made (printed) by *DNA* in a process called transcription. *RNA-d* serves to carry the genetic information it receives from *DNA*, which is information to form proteins.
2) *RNA* transfer (*RNA-t*)
Transfer R*NA* or *RNA*-t is printed in the nucleus, before placing itself in the cytoplasm*. RNA-t* has the task of binding amino acids in the cytoplasm.
3) *RNA* ribosome (*RNA-r*)
RNA-r is made in the nucleus and located within the ribosome. *RNA-r* is the meeting place between *RNA-d* and *RNA-t* at the time of protein synthesis [10].

2.3. Codon

Codon is nucleotide sequences in nucleic acids (*DNA* and *RNA*) that encodes the amino acids in the protein chain. *DNA* and *RNA* bases' built by the nucleotides that will specify the 20 amino acids. Thus, if each nucleotide is translated into amino acids, then there would be only 4 of the 20 amino acids that will be specified. If the nucleotide sorted by 2 pieces (example: *AG*, *GT*), then there would be 16 amino acids to be specified. This amount is not enough to specify the 20 amino acids. Therefore, there must be combination of at least three nucleotides (a triplet of nucleotides) to determine each particular amino acid. This nucleotide triplet code shall be referred to the standard genetic code or codon. Codons will provide $4^3 = 64$ amino acids to be specified, this amount is more than enough so that it will no amino acids are specified by more than one codon [1]. Nucleotida at codon basic component is written in the set of nucleotide bases found in *RNA* is {*U, G, C, A*} List Codon arranged at the *RNA* can be seen in Figure 1 as follows:



	Second letter								
First letter	U		C		A		G		Third letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		**UAA**	**Stop**	**UGA**	**Stop**	A
	UUG		UCG		**UAG**	**Stop**	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	**Met**	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 1. Code, Amino Acid and Encoded Stop Signal
(Source: [11])

2.4. Genetic Mutation

Mutation of genes is a change that occurs in nitrogen bases [3]. Based on the size of the number of altered bases, gene mutations are divided into two types namely point mutations and large (gross) mutations. Point mutations are mutations that involve only one nitrogen change while large mutations are mutations involving the change of more than one nitrogen base. An example of a point mutation is a *CCC* mutation into *CAC* and an example of a large mutation is *CCC* mutation into an *ACA*.

Gene mutations are also grouped based on their effects on the amino acids and *stopsignal* formed. Mutation of genes is generally differentiated into three types of mutations [3], i.e.,

1. Mute mutation (silent) are gene mutations that do not alter the amino acids and stopsignal formed. An example of this type of mutation is a mutation from CCC to CCA which both produce proline amino acids.
2. Nonsense mutation is a mutation of the gene that causes the amino acids to be formed to become stopsignal so that the protein synthesis process stops prematurely and the resulting amino acid chain becomes shorter than it should be .
3. Mutations of mis-meaning (missense) is a mutation that causes amino acids formed different from amino acids that should be formed.

2.5. Algebra Structure

Definition 2 (Group) [4]

Let G a nonempty set together with a binary operation that assigns to each ordered pair (a, b) of elements of G an element in G denoted by ab. We say G is a group under this operation if the following three properties are satisfied.

1. Associativity. The operation is associative;that is ($(ab)c = a(bc)$ for all a,b,$c$in$G$.
2. Identity. There is an element e (called the identity) in G such that $ae = ea = a$for all a in G.

3. Inverse. For each element a in G,there is an element b in G (called an inverse of a) such that $ab = ba = e$

If a group has $ab = ba$ for every a and b in G, then the group is called abelian group.

Definition 3 (order of a group) [4]
The number of elements of a group (finite or infinite) is called is *orde*r. We will use |*G*| to denote the order of *G*.

Ring Theory
Definition 4 (Ring) [4]
A ring R is a nonempty set with two binary operation, addition (denoted by $a + b$) and multiplication(denoted by ab), such that for all *a,b, c* in R :

1. $a + b = b + a$
2. $(a + b) + c = a + (b + c)$
3. There is an additive identity 0. That is , there is an element 0 in R such that $a + 0 = a$ for all $a \in R$.
4. There is an element $-a \in R$ such that $a + (-a) = 0$
5. $a(bc) = (ab)c$
6. $(b + c) = a \cdot b + ac$and $(b + c)a = ba + ca$.

Vector Spaces
Definition 5 : A set *V* is said to be a vector spaces over a field *F* if *V* is an abelian group under addition (denoted by +) and, if for each $\alpha \ \in F$ and $v \ \in V$, there is an element αv in *V* such that the following conditions hold for all α, β in *F* and all u, v in *V*.

1. $\alpha(u + v) = \alpha u + \alpha v$
2. $(\alpha + \beta)\, u = \alpha u + \beta u$
3. $\alpha(\beta u) = (\alpha\beta)u$
4. 1 *u* = *u*

Finite Field (Galois Field)
Definiion 5 : A set *V* is said to be a vector spaces over a field *F* if *V* is an abelian group under addition (denoted by +) and, if for each $\alpha \ \in F$ and $v \ \in V$, there is an element αv in *V* such that the following conditions hold for all α, β in *F* and all u, v in *V*

1. $\alpha(u + v) = \alpha u + \alpha v$
2. $(\alpha + \beta)\, u = \alpha u + \beta$
3. $\alpha(\beta u) = (\alpha\beta)u$
4. 1 *u* = *u*

3. Result

3.1. Representation of Genetic Code in a Three- Dimensional Space

To see the representation of the standard Genetic code, we investigate the structure of Algebra from *N*. Let *N* = *{C, U, A, G}* can be matched with $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(\bar{\mathbf{0}}, \bar{\mathbf{0}}), (\bar{\mathbf{0}}, \bar{\mathbf{1}}), (\bar{\mathbf{1}}, \bar{\mathbf{0}}), (\bar{\mathbf{1}}, \bar{\mathbf{1}})\}$ such that $\boldsymbol{C} = (\bar{\mathbf{0}}, \bar{\mathbf{0}})$, $\boldsymbol{U} = (\bar{\mathbf{0}}, \bar{\mathbf{1}})$, $\boldsymbol{A} = (\bar{\mathbf{1}}, \bar{\mathbf{0}})$, and$\boldsymbol{G} = (\bar{\mathbf{1}}, \bar{\mathbf{1}})$.

Based on Table 1 it appears that (*N, +)* is closed and associative, there is an identity element at *N*ie*C*, each element in *N* has an inverse ie$(C)^{-1} = C$, $(A)^{-1} = A$, $(U)^{-1} = U$, $(G)^{-1} = G$, so (*N*, +) abelian groups. Furthermore that the set of all codons or *NNN* forms a vector space over the field $\mathbb{Z}_2$.If the matching between *N* = *{C, U, A, G}* with the polynomial ring over $\mathbb{Z}_2$ i.e $\mathbb{Z}_2[x]/\langle x^2 + x + 1\rangle = \{0 + \langle x^2 + x + 1\rangle,\ 1 + \langle x^2 + x + 1\rangle,\ x + \langle x^2 + x + 1\ \rangle,\ x + 1 + \ \langle x^2 + x + 1\rangle\}$,this matching resulted in *N* having a four-element Galois field structure (GF(2^2)).

Table 1. Cayley table of (N ,+)

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

To facilitate the presentation of *NNN* in a three-dimensional space we use matching between $N = \{C, U, A, G\}$ with $\{0,1,2,3\}$ and with polynomial ring over $\mathbb{Z}_2$ so $C = 0 = 0$, $U = 1 = 1$, $A = x = 2$, dan $G = x + 1 = 3$. Thus it is concluded that DNA forms a vector space over a four-element Galois field $(\mathrm{GF}(2^2))$. [6].

Thus, a representation of *NNN* can be formed in a three-dimensional space called multicube*NNN*. The representation of *NNN* in a three-dimensional space which can be formed as many as 24 *NNN* multicubes relies on matching between *N* with $\{0,1,2,3\}$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$ (Joséet al, 2012: 125 and 141). For example, if $C = 0$, $U = 1$, $A = 2$, and $G = 3$ or if $C = (\bar{0}, \bar{0})$, $U = (\bar{0}, \bar{1})$, $A = (\bar{1}, \bar{0})$, dan $G = (\bar{1}, \bar{1})$ then the multycube*NNN* formed is called multicube *NNN* with the sequencing (*C, U, A, G*) . Figure 2 shows the multicube *NNN* by sequencing (*C, U, A, G*).

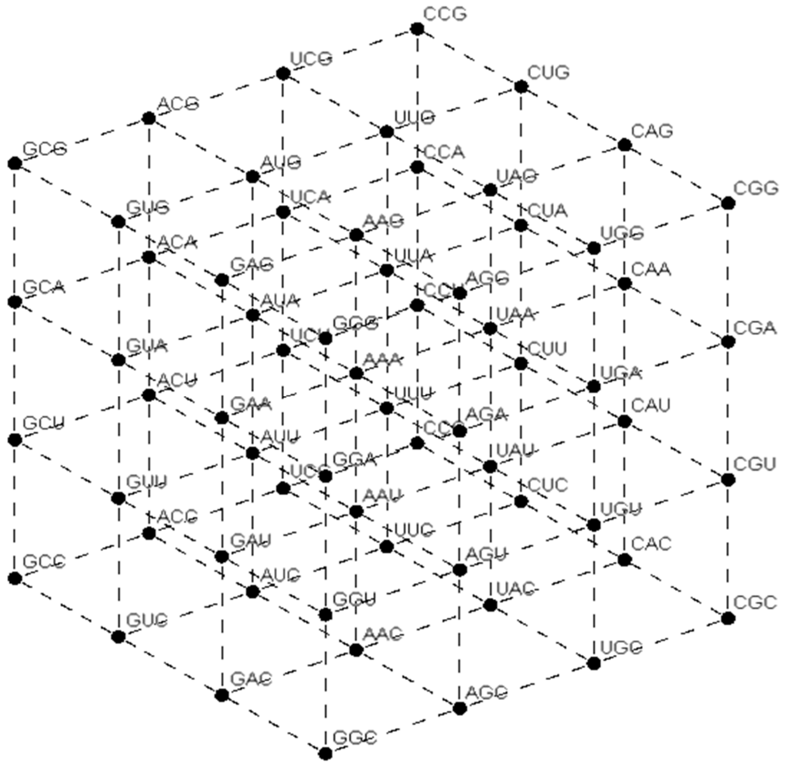


Figure 2. Multicube *NNN* by sequencing (*C, U, A, G*).

To see the mutations occurring in *NNN* will be used translations using vectors $(\bar{1},\bar{0}) \in \mathbb{Z}_2 \times \mathbb{Z}_2$. Translation of $(X_1, X_2, X_3) \in NNN$ with a vector of elements $\mathbb{Z}_2 \times \mathbb{Z}_2$ is defined as the translation of each nitrogen base (X_1, X_2, X_3) [6].

Suppose selected (*C, C, U*) ∈*NNN*, then the translations (*C, C, U*) are obtained as follows:
For *C* on *(C, C, U)*

$$\mathrm{T}_{\binom{\bar{1}}{\bar{0}}}(C) = \mathrm{T}_{\binom{\bar{1}}{\bar{0}}}\binom{\bar{0}}{\bar{0}} = \binom{\bar{0}}{\bar{0}} + \binom{\bar{1}}{\bar{0}} = \binom{\bar{1}}{\bar{0}} = A$$

For *A* on (*C, C, U*)

$$\mathrm{T}_{\binom{\bar{1}}{\bar{0}}}(U) = \mathrm{T}_{\binom{\bar{1}}{\bar{0}}}\binom{\bar{0}}{\bar{1}} = \binom{\bar{0}}{\bar{1}} + \binom{\bar{1}}{\bar{0}} = \binom{\bar{1}}{\bar{1}} = G$$

Thus, the translational result of (*C, C, U*) with vectors $(\bar{1},\bar{0}) \in \mathbb{Z}_2 \times \mathbb{Z}_2$ is (*A, A, G*).

3.2. Graph of The Set Of Amino Acids And Stopsignal

Representations of *NNN* can be formed as many as 24 multicube *NNN* with different ordering [6]. Based on Manhattan distance, two codons connected in multicube *NNN* have distance one, while biologically, two codons connected in multicube *NNN* indicate the occurrence of a single point mutation from one codon to another codon as shown in Figure 2 above.

Examples of two codons connected in the *NNN* multicube are codons (*C, U, G*) and (*C, A, G*) where the distance between them is :

$$\begin{aligned} \mathrm{d}\big((C,U,G),(C,A,G)\big) &= \mathrm{d}\big((0,1,3),(0,2,3)\big) \\ &= |0-0| + |1-2| + |3-3| \\ &= 0+1+0 = 1 \end{aligned}$$

While the distance between the codons (*A, U, G*) and (*C, A, G*) is

$$\begin{aligned} \mathrm{d}\big((A,U,G),(C,A,G)\big) &= \mathrm{d}\big((2,1,3),(0,2,3)\big) \\ &= |2-0| + |1-2| + |3-3| \\ &= 2+1+0 = 3 \end{aligned}$$

Thus it can be concluded that codons (*A, U, G*) and (*C, A, G*) are not connected.

It will be constructed graph $G_i$which is related to amino acids and stop signals where the set of all vertices of G_i is $A \cup \{S\}$and two different vertices are connected by an element side $E(G_i)$if it has a Manhattan distance of one. To make it easier to write the vertices, each amino acid element $A \cup \{S\}$is written in a three-letter symbol and the stopsignal is denoted by a *stop* so that the set of all vertices of $G_i = A \cup \{S\} =$ {Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Ile, Leu, Lis, Met, Phe, Pro, Ser, Thr, Try, Tyr, Val, *Stop*}.

Next will be determined the number of $G_i$graphs that can be formed based on multycube *NNN*. Representation of *NNN* that can be formed as many as 24 graphs, with different sorting. Because for every two graphs each obtained from two multicube *NNN* with the same reverse sequencing it can be concluded that the $G_i$graphs that can be formed based on multycube*NNN* are as many as twelve graphs.

Figure 2 and Table 2 it is found that the Ala vertices coded by codons (G,C,C), (G,C,U), (G,C,A),and(G,C,G), respectively conected with the following vertices.

1. $(G,C,C)$is connected to(G,U,C), (A,C,C), and(G,C,U), and has Manhattan distance is one.
   ((G,U,C), encodes Val and d ((G,C,C), (G,U,C))= 1 then the vertex of Ala is connected with Val in graph G_1.
   (A,C,C) encodes Thr and d ((G,C,C), (A,C,C)= 1 then the Ala vertex is connected with Thr in the graph G_1.
   Since(G,C,U) and (G,C,C), have Ala code, it can not be deduced from this case.

Table 2. Element of NNN/R_1With Amino Acids and Stop signal

Class equivalence	**Elemen of Class equivalence**	**Elemen of$A \cup \{S\}$**
$[(G,C,C)]$	(G,C,C), (G,C,U), (G,C,A), and (G,C,G)	Alanin (Ala)
$[(C,G,C)]$	(C,G,C), (C,G,U), (C,G,A), $(C,G,$G), (A,G,A),and(A,G,G)	Arginin (Arg)
$[(A,A,C)]$	(A,A,C) and (A,A,U)	Asparagin (Asn)
$[(G,A,C)]$	(G,A,C) and (G,A,U)	Asam aspartat (Asp)
$[(U,G,C)]$	(U,G,C) and (U,G,U)	Sistein (Cys)
$[(G,A,A)]$	(G,A,A) and (G,A,G)	Asam glutamat (Glu)
$[(C,A,A)]$	(C,A,A) and (C,A,G)	Glutamin (Gln)
$[(G,G,C)]$	(G,G,C), (G,G,U), (G,G,A), and (G,G,G)	Glisin(Gly)
$[(C,A,C)]$	(C,A,C) and (C,A,U)	Histidin (His)
$[(A,U,C)]$	(A,U,C), (A,U,U), and (A,U,A)	Isoleusin (Ile)
$[(C,U,C)]$	(C,U,C),(C,U,U),(C,U,A), (C,U,G), (U,U,A), and (U,U,G)	Leusin (Leu)
$[(A,A,A)]$	(A,A,A) and (A,A,G)	Lisin (Lis)
$[(A,U,A)]$	(A,U,G)	Metionin (Met)
$[(U,U,C)]$	(U,U,C) and (U,U,U)	Phenylalanin (Phe)
$[(C,C,C)]$	(C,C,C), (C,C,U), (C,C,A), and (C,C,G)	Prolin (Pro)
$[(U,C,C)]$	(U,C,C),(U,C,U),(U,C,A), (U,C,G), (A,G,C), and (A,G,U)	Serin (Ser)
$[(A,C,C)]$	(A,C,C), (A,C,U), (A,C,A), and(A,C,G)	Threonin (Thr)
$[(U,G,G)]$	(U,G,G)	Triptofan (Try)
$[(U,A,C)]$	(U,A,C) and (U,A,U)	Tirosin (Tyr)
$[(G,U,C)]$	(G,U,C), (G,U,U), (G,U,A), and(G,U,G)	Valin (Val)
$[(U,A,A)]$	(U,A,A), (U,A,G), and (U,G,A)	*Stop signal* (*stop*)

2. $(G,C,U)$is connected to (G,U,U), (A,C,U), (G,C,C),and(G,C,A), and has Manhattan distance is one.
   (G,U,U), encodes Val and d ((G,U,U), (G,C,U))= 1 then the Ala vertex is connected with Val in graph G_1.
   (A,C,U), encodes Thr and d ((G,C,U), (A,C,U)), = 1 then the Ala vertex is connected to Thr in the graph G_1.
   Because (G,C,U), (G,C,A), and (G,C,C),both encode Ala, it cannot be deduced from this case.
3. (G,U,A), is connected to(G,C,A), (A,C,A), and(G,C,U), and has Manhattan distance is one.

(G, U, A), encodes Val and d $((G, C, A), (G, U, A))$= 1 then the Ala vertex is connected with Val in graph G_1..
(A, C, A), encodes Thr and d $((G, C, A), (A, C, A))$ = 1 then the Ala vertex is connected to Thr in the graph G_1.
Because (G, C, A), and (G, C, U)encode Ala then it can not be deduced from this case.

4. (G, C, G) connected to(G, U, G), (A, C, G), and(G, C, A), and has Manhattan distance is one.
   (G, U, G), encodes Val and d $((G, C, G)(G, U, G))$ = 1 then the vertex Ala is connected with Val in the graph G_1.
   (A, C, G), encodes Thr and d $((G, C, G), (A, C, G))$ = 1 then the Ala vertex is connected to Thr in the graph G_1.
   Because (G, C, G) and (G, C, A), encode Ala, it cannot be deduced from this case.

Based on 1 until 4 it can be concluded that the Ala vertex is only connected with the Val and Thr vertices in the G_1 graph. The same is done for the other twenty vertices to obtain the graph G_1 shown in Figure 3.

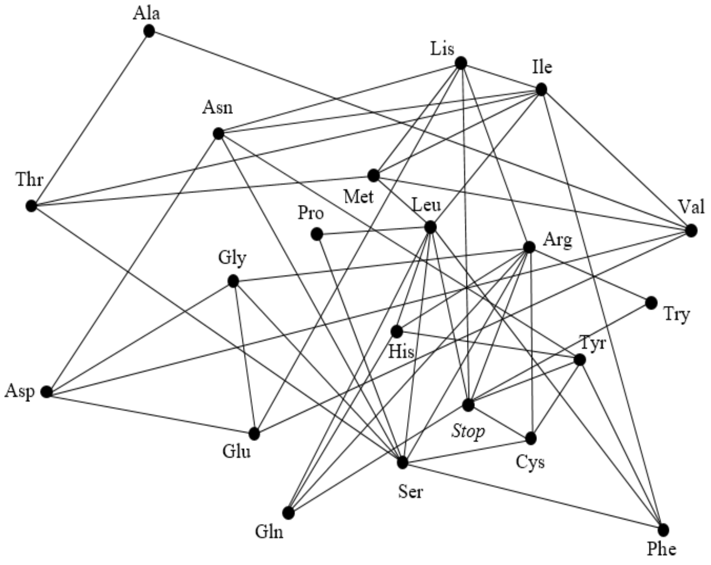


Figure3. Graph G_1 from multicube *NNN*

4. Conclusion

1. The set of DNA has the structure of Algebra as Group, Galois Field and Vector space
2. By using linear transformation and with the help of Geogebra software can be seen representation of mutations geometry in the form of multicube
3. Based on the multicube formed, the chemical compounds of the codon can be represented in graph form so that it can easily see the relationship between the encoded compound

Acknowledgments
The authors thank to the University of Padjadjaran, due to the funding of this work through the scheme of Academic Leadership Grant (ALG) with contract number 855/UN6.3.1/PL/2017.

References

[1] Anton H 2010 *Elementary Linear Algebra Applications Version* (Canada: John Wiley & Sons, inc)

[2] Akhtar A dan Ali T 2014 Anlysis of Unweigthed Amino Acids Network*International Scholarly Research Notices*

[3] Elrod S L, Stansfield WD 2010*Schaum's Outlines: Genetics*. (New York: Mc Graw Hill)

[4] GalianJ *Contemporary Abstract Algebra*.Sixt Edition 2006 (New York: Houghton)

[5] Jimēnez-Montaño M A, de la Mora-Basañez C R., and Pöschel T 1996 *The hypercube Structure of The Genetic Code Explains Conservative and Non-conservative Amino Acid Substitutions in vivo and in vitro*. *Biosystems***39** pp 117-125

[6] José M V, Morgado E R, dan Govezensky T 2011 Genetic Hotels for The Standard Genetic Code: Evolutionary Analysis Based Upon Novel Three-dimensional Algebraic Models*Bull. Math. Bio*, **73** pp 1443-1476

[7] José M V, Morgado E R, Sánchez R, and Govezensky T 2012 The 24 Possible Algebraic Representations of The Standard Genetic Code in Six and Three Dimensions*Advanced Studies in Biology***4** pp 119-152

[8] José M V, Morgado E R, Guimarães R C, Zamudio G S, Tores S, Bobadilla J R, Sosa D 2014Three-Dimensional Algebraic Models of the tRNA Code and 12 Graph for Representing the Amino Acids*Life*, **4** pp 341-373

[9] Madhulatha T S 2012An Overview on Clustering Methods*IOSR J of Engineering* 2(4)pp 719-725

[10] Marks A D, Smith C, dan Liebermann M 2005*Basic Medical Biochemistry: A Clinical Approach.*Lippinoott Williams and Wilkins

[11] Available from: https://cnx.org/contents/GFy_h8cu@9.87:QEibhJMi@8/The-Genetic-Code[cited 15 June 2017]