

Full Length Research Paper

An automorphism in the LK_ITB5a H110F gene mutation

Qonita Umami Safitri^{1*}, Elvira Kusniyanti¹, Irawati¹, Ilma Fauziah Ma'ruf² and Akhmaloka^{2,3}

¹Mathematics Department, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia.

²Biochemistry Research Group, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Indonesia.

³Department of Chemistry, Faculty of Science and Computer, Universitas Pertamina, Jakarta, Indonesia.

Received 10 January, 2020; Accepted 20 May, 2020

Sanchez et al. (2005) have shown the structure of vector space genetic code over the Galois field $GF(4)$ which relate to the physicochemical properties of the genetic code on proteins. From this vector space, an automorphism can be constructed to reflect the mutation process in the genetic code. This study investigates a type of transformation in the LK_ITB5a lipase gene mutation. The result of the study shows that there is a transformation matrix from the wild type gene lipase LK_ITB5a to mutant gene lipase LK_ITB5a H110F which is a diagonal matrix with non-zero determinant. This means that the transformation is an automorphism.

Key words: Automorphism, vector space, mutational pathways

INTRODUCTION

The genetic code is a set of instructions in genes that express amino acids from a row of 3 bases on the RNA (ribonucleic acid) strand. Each of these 3 bases is called a codon. The genetic code is involved in the process of protein synthesis to translate strands of RNA codons into amino acids that form a protein molecule (Crick, 1968). This protein plays a role in the process of living and the growth of organisms. The genetic code is composed of a combination of 4 RNA bases, namely Adenine (A), Guanine (G), Cytosine (C) and Uracil (U). The genetic code has been presented in tables adjusted to the order of the second bases. One of the orders standard genetic code is presented in Table 1.

Genetic code bases can be presented in sets $B = \{G, U, A, C\}$. Sanchez et al. (2005, 2006) proposed a vector space structure on the set B after being inspired by binary representations made by previous researchers,

such as Jimenez-Montano et al. (1996), Stambuk (2000) and Karasev and Stevanov (2001). The construction of the vector space begins with matching the set B with $\mathbb{Z}_2 \times \mathbb{Z}_2$. The matching is done based on the number of hydrogen bonds and the physicochemical properties of the bases. As a result, Sanchez et al. (2005) chose the representation $G = 00, U = 01, A = 10$ and $C = 11$.

The match induces a group isomorphism between $B, \mathbb{Z}_2 \times \mathbb{Z}_2$ and $\frac{\mathbb{Z}_2[x]}{\langle x^2+x+1 \rangle}$. Furthermore, group $\frac{\mathbb{Z}_2[x]}{\langle x^2+x+1 \rangle}$ is a ring and Galois field $GF(2^2)$. Hence, we can define an

isomorphism $\psi: B \rightarrow \frac{\mathbb{Z}_2[x]}{\langle x^2+x+1 \rangle}$, such that B also has a structure of Galois field $GF(2^2)$ (Sanchez et al., 2005; Sanchez and Grau, 2006).

Thereafter, the group of genetic code with 3 bases can

*Corresponding author. E-mail: qonita.usafitri@gmail.com.

Table 1. The standard genetic code.

First base position	Second base position								Third base position
	G		U		A		C		
¹G	GGG		GUG		GAG	E	GCG		G
	GGU	G	GUU	V	GAU	D	GCU	A	U
	GGA		GUA		GAA	E	GCA		A
	GGC		GUC		GAC	D	GCC		C
U	UGG	W	UUG	L	UAG	Stop	UCG		G
	UGU	C	UUU	F	UAU	Y	UCU	S	U
	UGA	Stop	UUA	L	UAA	Stop	UCA		A
	UGC	C	UUC	F	UAC	Y	UCC		C
A	AGG	R	AUG	M	AAG	K	ACG		G
	AGU	S	AUU		AAU	N	ACU	T	U
	AGA	R	AUA	I	AAA	K	ACA		A
	AGC	S	AUC		AAC	N	ACC		C
C	CGG		CUG		CAG	Q	CCG		G
	CGU	R	CUU	L	CAU	H	CCU	P	U
	CGA		CUA		CAA	Q	CCA		A
	CGC		CUC		CAC	H	CCC		C

¹ The one letter symbol of amino acids.

be constructed as a product of group B , namely $C_g = B \times B \times B$. It can be checked that $(C_g, +)$ is also a commutative group. So, we can define the vector space genetic code C_g over the Galois field $GF(2^2)$.

In addition, linear transformation can also be defined as a reflection of the genetic mutation process (Sanchez et al., 2005).

The mutation is a change in the genetic material of organisms that are inherited to the next generation. The mutation process is called mutagenesis, the mutated organism is called the wild type, and the mutation results are called mutants (Marwadewi, 2017).

Mutations can occur naturally and artificially. Natural mutations occur due to natural factors or errors of replication in the process of meiosis, artificial mutations are mutations due to factors from outside the body of the organism, such as ionization, enzymes or viruses, and the process of using chemicals. One of the artificial mutations that can be used is the polymerase chain reaction (PCR).

PCR is an enzymatic DNA replication technique without using organisms (Hasibuan, 2015; Ma'ruf, 2017). PCR can multiply DNA in a short time, therefore, the use of PCR is still the choice for DNA analysis in various fields of biochemistry and molecular biology. Ma'ruf (2017) used PCR to mutate the LK_ITB5a lipase gene to get certain characteristics of mutant that will be needed in the industry. This study will investigate whether the claims made by Sanchez et al. (2005) can also apply to every type of mutation, including artificial mutations.

VECTOR SPACE AND LINEAR TRANSFORMATION ON GENETIC CODE

The mathematical framework of this paper will just be the vector space and linear transformation on genetic code. First, the vector space built from the set RNA bases B . Second, the vector space of genetic codes can be developed by the product of RNA bases vector space. Finally, we can construct the linear transformation to describe the mutational pathways on genetic code.

Vector space genetic code

Consider the RNA bases set B and $\mathbb{Z}_2 \times \mathbb{Z}_2$ with the mapping $f: B \rightarrow \mathbb{Z}_2 \times \mathbb{Z}_2$, where $f(G) = 00, f(U) = 01, f(A) = 10$ and $f(C) = 11$. It can be seen that f is a bijective function and a group isomorphism with the sum of the corresponding elements. Then, consider the ring $\mathbb{Z}_2[x]$ and ideal prime $\langle x^2 + x + 1 \rangle$ of $\mathbb{Z}_2[x]$. The ring factor $\frac{\mathbb{Z}_2[x]}{\langle x^2 + x + 1 \rangle} = \{\bar{0}, \bar{1}, \bar{x}, \overline{x+1}\}$ is a Galois field with 4 elements, denoted as $GF(2^2)$. Since $\frac{\mathbb{Z}_2[x]}{\langle x^2 + x + 1 \rangle}$ is a ring, then $\frac{\mathbb{Z}_2[x]}{\langle x^2 + x + 1 \rangle} +$, is a commutative group, so we can build a group isomorphism $:\mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow \frac{\mathbb{Z}_2[x]}{\langle x^2 + x + 1 \rangle}$, hence $\phi(00) = \bar{0}, \phi(01) = \bar{1}, \phi(10) = \bar{x}$ and $\phi(11) = \overline{x+1}$.

We define:

$$\psi: \frac{GF(2^2)}{a_0 + a_1 \bar{x}} \rightarrow \frac{B}{f^{-1}(\phi^{-1}(a_0 + a_1 \bar{x}))}, a_0, a_1 \in \mathbb{Z}_2 \quad (1)$$

Table 2. Cayley table for Galois Field of genetic bases.

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

with $f^{-1}: \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow B$ and $\phi^{-1}: \frac{\mathbb{Z}_2[x]}{(x^2+x+1)} \rightarrow \mathbb{Z}_2 \times \mathbb{Z}_2$ are bijective function.

Since ψ is a bijective function, there is $\psi^{-1}: B \rightarrow GF(2^2)$ such that for all $X, Y \in B$ we can define the sum “+” and product “.” as:

$$X + Y = \psi(\psi^{-1}(X) + \psi^{-1}(Y) \text{ mod } 2) \quad (2)$$

$$X \cdot Y = \psi(\psi^{-1}(X) \cdot \psi^{-1}(Y) \text{ mod } x^2 + x + 1) \quad (3)$$

From the Equations (2) and (3), we have a Galois Field Genetic Code Bases $GF(2^2)$. The operation result can be seen in Table 2. Based on the Galois Field of genetic code, we can define some vector spaces.

Definition 1 (Vector space of genetic code bases)

Commutative group $(B, +)$ is a vector space over Galois field $GF(2^2)$ with the sum and scalar product defined as follow:

$$X + Y = \psi(\psi^{-1}(X) + \psi^{-1}(Y) \text{ mod } 2) \quad (4)$$

$$\alpha \cdot X = \psi(\alpha \cdot \psi^{-1}(X) \text{ mod } 2) \quad (5)$$

for all $X, Y \in B, \alpha \in GF(2^2)$. Furthermore, the base of B is $\{U\}$, so $\dim(B) = 1$.

Definition 2 (Vector space genetic code)

Commutative group $(C_g, +)$ is a vector space over Galois field $GF(2^2)$ with the sum and scalar product defined as follow:

$$X + Y = (X_1 + Y_1, X_2 + Y_2, X_3 + Y_3) \quad (6)$$

$$\alpha \cdot X = (\alpha \cdot X_1, \alpha \cdot X_2, \alpha \cdot X_3) \quad (7)$$

for all $X = X_1X_2X_3, Y = Y_1Y_2Y_3 \in C_g, \alpha \in GF(2^2), X_i, Y_i \in B, i = 1, 2, 3$. Furthermore, the standard base of C_g is $B = \{UGG, GUG, GGU\}$, so $\dim(C_g) = 3$.

The vector space genetic code C_g over Galois field $GF(2^2)$ is called VG . Finally, this structure can be extended to vector space S with N genetic codes, so we have the vector space genetic code $3N$ -dimension.

Definition 3 (Vector space of 3N-dimension genetic codes over Galois Field $GF(2^2)$)

Let S be a product of N groups of $(C_g, +)$, that is:

$$\begin{aligned} (S, +) &= (C_g, +) \times (C_g, +) \times \dots \times (C_g, +) \quad (N \text{ times}) \\ &= (B, +) \times (B, +) \times (B, +) \times \dots \times (B, +) \quad (3N \text{ times}) \end{aligned} \quad (8)$$

where for all $g \in S, g = (c_1, c_2, \dots, c_N), c_i \in C_g, i = 1, 2, \dots, N$ and for all $c_i = X_{i1}X_{i2}X_{i3}, X_{ij} \in B, j = 1, 2, 3$. Then, S is a $3N$ -dimension vector space over Galois field $GF(2^2)$, denoted by $(VG)^N$ or $(B^3)^N$. The standard base of S is $B = \{e_1, e_2, \dots, e_{3N}\}$, where $e_k = Y_1Y_2Y_3 \dots Y_{3N}$, with $Y_k = U, Y_l = G, \forall l \neq k, k, l = 1, 2, \dots, 3N$.

Vector space VG have a normal subspace $C_G = \{X_1GX_3: X_1, X_3 \in B\}$ and $C_{GG} = \{GGX_3: X_3 \in B\}$. Quotient space VG/C_G divide VG based on amino acids polarity, and VG/C_{GG} divide VG based on the same or similar amino acids (Aisah et al., 2015; Jose et al, 2012). Besides, as a group, C_g has normal subgroup $C_{GGA}, C_{GGU}, C_{GGC}$ which divide C_g as follow:

- 1) Quotient group C_g/C_{GGA} divide C_g based on base type purine and pyrimidine.
- 2) Quotient group C_g/C_{GGU} divide C_g based on nucleotides chemical classification, amino and keto.
- 3) Quotient group C_g/C_{GGC} divide C_g based on nucleotides biological classification strong and weak.

Quotient space preserves the structure and physicochemical property of genetic code. According to the result, we can describe that the genetic code table is not randomly arranged (Crick, 1968). These results also suggests the strong connection between codon algebraic structure with the physicochemical property.

Automorphism on vector space of genetic code

The simplest mutation process is point mutation. A point mutation can be seen as the local endomorphism on the vector space of genetic code as in Definition 4 and Theorem 5 (Sanchez et al., 2005).

Definition 4 (Local endomorphism)

Let S be a vector space over the Galois field $GF(2^2)$, with

$\dim(S) = N$. An endomorphism $f: S \rightarrow S$ will be called local endomorphism if there is $k \in \{1, 2, \dots, N\}$ and there are $\alpha_{ik} \in GF(p^n)$ (Sanchez et al., 2005), such that:

$$f(e_i) = \alpha_{ik}e_k + e_i, \quad i \neq k \quad (9)$$

and

$$f(e_k) = \alpha_{kk}e_k \quad (10)$$

If $\alpha_{kk} \neq 0$, then f is local automorphism, endomorphism f is diagonal local automorphism if $f(e_k) = \alpha_{kk}e_k$ and $f(e_i) = e_i$, for all $i \neq k$.

This means that f is local endomorphism if one of the bases has changed, namely:

$$\begin{aligned} f(x_1, x_2, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_N) \\ = (x_1, x_2, \dots, x_{k-1}, y_k, x_{k+1}, \dots, x_N) \end{aligned}$$

where $y_k = \sum_{i=1, i \neq k}^N \alpha_{ik}x_i$.

Theorem 1

For every single point mutation that change the codon α_i of the wild type gene $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_i, \dots, \alpha_N)$ (α not the zero vector in vector space), by the codon β_i of the mutant gene $\beta = (\alpha_1, \alpha_2, \dots, \beta_i, \dots, \alpha_N)$ (Sanchez et al., 2005), there is:

- 1) At least a local endomorphism f such that $f(\alpha) = \beta$.
- 2) At least a local automorphism f such that $f(\alpha) = \beta$.
- 3) A unique diagonal automorphism f such that $f(\alpha) = \beta$ if and only if the codon α_i and β_i of the wild type and mutant genes, respectively, are different of GGG .

Proof (Sanchez et al., 2005)

Note that the purpose of the genetic code endomorphism study is to find out the mutation process in the genetic code so that we can understand the conditions before and after the mutation. Therefore, the selection of endomorphism is not arbitrary. To fulfill this goal, the endomorphism of f chosen must have a reversal, that is, there is f^{-1} so that $f \circ f^{-1} = id$. Endomorphism which has this property is automorphism. Furthermore, the set of automorphisms is a group with composition operations.

Proposition 1. Let $(C_g, +)$ be a group of genetic code. Set $\mathbb{G} = \{f \in Aut(C_g)\}$ is a group with composition (" \circ ") .

Proposition 2. Let f be an endomorphism genetic code, and let P be the matrix representation of f . Then f is automorphism if and only if $\det(P) \neq 0$.

Furthermore, a set $\mathbb{G}' = \{F \in End((VG)^N)\}$ also applies to the same thing, which is, for each $F \in \mathbb{G}'$ there is $P \in M_N(C_g)$ so that $F(X) = XP$, for all $X \in (C_g)^N$. In other word, the representation matrix $P = (P_{ij})$, with $P_{ij} \in M_3(B)$ and $P_{ij} \in End(VG_i, VG_j)$. As a result, representation matrix P is automorphism, if $P_{ii} \in Aut(\mathbb{G}_i)$ (Sanchez et al., 2006).

Let $St(k) \subset \mathbb{G}$ be a stabilizer subgroup that fix base position k , that is,

$$St(k) = \{f \in \mathbb{G}: f(X_1X_2X_3) \in S_k\} \quad (11)$$

where S_k is the subset of codon conserving the same base position $k \in \{1, 2, 3\}$.

Note that $St(k)$ preserves one base, while local endomorphism changes one base. As a result, every $f \in St(k)$ can be seen as a composition of 2 local endomorphism $g, h \in End(C_g)$, so that $f = g \circ h$. Hence, if P, Q, R are the matrix representation for f, g and h , respectively, then $P = QR$. If g and h is local diagonal automorphism, then so is f .

AUTOMORPHISM ON LIPASE LK_ITB5a H110F GENE MUTATION

Lipase is one of the enzymes used in the industry. Generally, industries need lipases with certain characteristics. One example is lipases that are resistant to high pH needed for processing textile waste (Ma'ruf, 2017).

In this study, the lipase gene used was LK_ITB5a lipase which had a 99% identity with lipase *Pseudomonas Stutzeri* LipC (Ma'ruf, 2017). The LK_ITB5a lipase gene clone consists of 936 base pairs encoding 311 amino acids and 1 start codon. The LK_ITB5a lipase was mutated by the PCR approach through site-directed mutagenesis on the residual oxyanion hole H¹¹⁰ substituted to F¹¹⁰. This mutation is done to see the effect of these mutations on the character of lipase. The results of the chemical analysis showed that the mutation on LK_ITB5a wild type into the LK_ITB5a H110F mutant caused a change in the interaction between the enzyme and the substrate so as to increase the enzyme specifications.

Next, an algebraic approach will be observed, which is by seeing the mutation process as a linear transformation in the vector space of genetic code, namely:

1. The lipase gene LK_ITB5a has 936 bases that encode 1 start codon and 311 amino acids. Then, we have $N = 312$, so that $(VG)^{312}$ is a vector space $3N$ -dimension over the Galois field $GF(2^2)$.

2. Changes in the genetic code in the LK_ITB5a H110F mutant actually occur in codons 115 and 111 (because the first codon is the start codon, it does not encode

amino acids), that is, the missense mutation H110F on codon 111 and silent mutation T114T on codon 115, then based on Theorem 1 there is a local endomorphism in the genetic code vector space.

3. Consider the missense mutation on 111th codon, $CAC \rightarrow TTC$ (UUC).

Based on the earlier mentioned points in the study, there is $f \in St(3)$. This can also indicate the presence of local endomorphism $\tau, \sigma \in End(VG)$, where $\tau(X_1X_2X_3) = X_1Y_2X_3$ and $\sigma(X_1X_2X_3) = Y_1X_2X_3$, so that $f = \sigma \circ \tau$.

Let:

$$P = \begin{bmatrix} U & G & G \\ G & C & G \\ G & G & U \end{bmatrix} \text{ and } Q = \begin{bmatrix} A & G & G \\ G & U & G \\ G & G & U \end{bmatrix}$$

be the matrix representation of τ and σ , respectively. Then, matrix representation of f is

$$R = PQ = \begin{bmatrix} A & G & G \\ G & C & G \\ G & G & U \end{bmatrix}.$$

It can be calculated that $det(R) = U \neq G$, so the endomorphism of f is automorphism. More specifically, the endomorphism f is a diagonal automorphism.

4. Consider the silent mutation on 115th codon, $ACC \rightarrow ACA$.

Changes occur in only the 3rd base, and also preserve the 1st and 2nd base, meaning there is local endomorphism $\tau \in End(VG)$, so that $\tau(X_1X_2X_3) = X_1X_2Y_3$, with the matrix representation

$$W = \begin{bmatrix} U & G & G \\ G & U & G \\ G & G & C \end{bmatrix}.$$

It can be calculated that $det(W) = C \neq G$, then endomorphism τ is a local diagonal automorphism. It also appears that $\tau \in St(1) \cap St(2)$. This mutation is most common (Sanchez et al., 2005; Sanchez and Grau, 2006).

Now, consider the mutation process from wild type gene LK_ITB5a to mutant gene LK_ITB5a H110F. Then there is $F: (VG)^{312} \rightarrow (VG)^{312}$ with matrix representation $P = (P_{ij})$, where $P_{111,111} = R, P_{115,115} = W, P_{ii} = I_3$, and $P_{ij} = [G]_{ij}, \forall i \neq 115, i \neq 111, j \neq i$, with

$$I_3 = \begin{bmatrix} U & G & G \\ G & U & G \\ G & G & U \end{bmatrix}$$

Since $det(P) = C \neq G$, so that F is an automorphism.

Conclusion

From the observations, it was found that the artificial mutation in the lipase LK_ITB5a H110F can also be seen as a linear transformation in the vector space of the genetic code. Particularly, the linear transformation in LK_ITB5a H110F is an automorphism.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENT

The authors appreciate the Genetic and Molecular Biotechnology Research Group, School of Life Science and Technology, Institut Teknologi Bandung, for lipase LK_ITB5a data support and all information about DNA.

REFERENCES

- Aisah I, Kurniadi E, Carnia E, Ula N (2015). Representasi mutasi kode genetik standar berdasarkan basa nukleotida. *Jurnal Matematika Integratif* 11(1):25-34.
- Crick F (1968). The origin of the genetic code. *Journal of Molecular Biology* 38(3):367-379. [https://doi.org/10.1016/0022-2836\(68\)90392-6](https://doi.org/10.1016/0022-2836(68)90392-6)
- Hasibuan E (2015). Peranan teknik polimerase chain reaction (PCR) terhadap perkembangan ilmu pengetahuan, Universitas Sumatera Utara. <http://repository.usu.ac.id/handle/123456789/61186>
- Jose MV, Morgando ER, Sanchez R, Govezensky T (2012). The 24 possible algebraic representation of the standard genetic code in six or in three dimensions. *Advanced Studies in Biology* 4(3):119-152.
- Jimenez-Montano MA (1996). The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions *in vivo* and *in vitro*. *Biosystems* 39(2):117-125.
- Karasev VA, Stevanov VE (2001). Topological Nature of the Genetic Code. *Journal of Theoretical Biology* 209(3):303-317.
- Marwadewi DA (2017). Mutasi genetik. Faculty of Animal Husbandary, Udayana University, Bali, Indonesia.
- Ma'ruf IF (2017). Site directed mutagenesis (H110F) and its role toward activity of LK_ITB5a lipase. Master's Program Thesis, Institut Teknologi Bandung.
- Sanchez R, Perfetti LA, Grau RH, dan Morales ERM (2005). A new DNA sequences vector space on genetic code galois field, *MATCH Communication in Mathematics and in Computer Chemistry* 54:3-28.
- Sanchez R, dan Grau R (2006). A novel algebraic structure of genetic code over the galois field of four DNA bases, *Acta Biotheoretica* 54(1):27-42.
- Sanchez R, Grau R, dan Morgando E (2006). A novel Lie algebra of the genetic code over the galois field of four DNA bases. *Mathematical Biosciences* 202(1):156-174. <https://doi.org/10.1016/j.mbs.2006.03.017>
- Stambuk N (2000). Universal metric properties of the genetic code. *Croatica Chemica Acta* 73:1123-1139.