

Free Triplet Conjecture and Equivalence Classes Derived Using Group Theory

Zhang Dakun^{*}, Song Guozhi and Huang Cui

School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin300387, China

Abstract: All proteins are made up of 20 different amino acids which contain 4 kinds of nucleotides. Three consecutive nucleotides on the gene, called triplet codons, are used to code an amino acid, and 64 triplet codons comprise the genetic code table. Central dogma (DNA-RNA-protein) has been acknowledged, but the process and mechanism of mRNA passing through the nuclear membrane still require further investigation. For these two problems mentioned above, this paper proposed a conjecture of nucleotide free triplet and obtained 20 equivalence classes of mapping from free triplet vertex set to nucleotide set using group theory. Whether the four numbers 3, 4, 20 and 64 have relevance are taken into consideration here. Subsequently, the numbers 3, 4, 20 and 64 were connected together which was important for the analysis of triplet code and protein composition.

Keywords: DNA/mRNA/protein, equivalence class, free triplet, group theory, nucleotide.

1. INTRODUCTION

The deciphering of the genetic code is one of the most outstanding achievements in the history of science. It not only provides a theoretical basis for the study of protein synthesis, but also confirms the precision of the central dogma. The presence of the genetic code has been further validated with the help of rapid development in molecular biology techniques such as DNA, RNA sequence determination and amino acid sequencing technology advances since the 1970s [1].

Though research on the decoding of genetic code has already had a history of more than forty years, interpretation of codon distribution and analysis of the nature and amount of amino acid code are still a challenge [2]. So the study of the genetic code continues to attract many scientists [3-7] and so many interdisciplinary research topics like life sciences, and molecular biology are to be further explored. Newton once said: "bold conjecture can be a great discovery." The emergence of any scientific theory in the history has bold conjecture and assumptions involved in the gestation process. For example, Einstein's general relativity theory, DNA double helix structure, triplet genetic code, *etc*, all have their origin aided by conjecture. The study on challenging interdisciplinary subjects like; life science and molecular biology, still needs bold conjecture and hypothesis. Galileo said. "The nature is a book written in Mathematical language", also it is necessary to use more mathematical methods in the study of life sciences and molecular biology [8, 9].

2. THE DERIVATION OF CONJECTURE OF NUCLEOTIDE FREE TRIPLET

2.1. Genetic Code Composition

The genetic code is a set of rules by which information encoded within the genetic material (DNA or mRNA) is translated into proteins. Each amino acid in the protein is expressed by codon composed of three consecutive nucleotides in the gene [10]. DNA has four different nucleotides, each containing four different bases: adenine (A), guanine(G), thymine(T) and cytosine(C). The order of these bases in the gene determines the sequence of amino acids in the protein which forms the primary structure of protein. When the nucleotide sequence of DNA transcribes to mRNA, uracil(U) replaces thymine(T). The 64 possible triplet codes are mentioned in the 9th reference.

The composition of triplet code was first proposed by astrophysicist Gamow in 1954 [11]. Though it was questioned by many biologists and geneticists, it was confirmed by the Institute of Science in 1961 (Crick et al first confirmed the correctness of the composition of triplet code from the perspective of genetics) [12].

2.2. Central Dogma and mRNA

1. Central Dogma of molecular biology

In 1958, Crick proposed to name the genetic pathway as central dogma, suggesting that chromosomal and DNA are the templates for the RNA molecule. Subsequently, It was transferred to the cytoplasm after synthesis to determine the amino acid sequence of the protein [12]. The DNA is the template used for self-replication and also for the synthesis of RNA (transcription), and RNA is the template for the synthesis of protein (translation). The latter two processes can only be executed in one direction.

^{*}Address correspondence to this author at School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin 300387, China; Tel: +86 13752392145; E-mail: zhangdakun2013@163.com

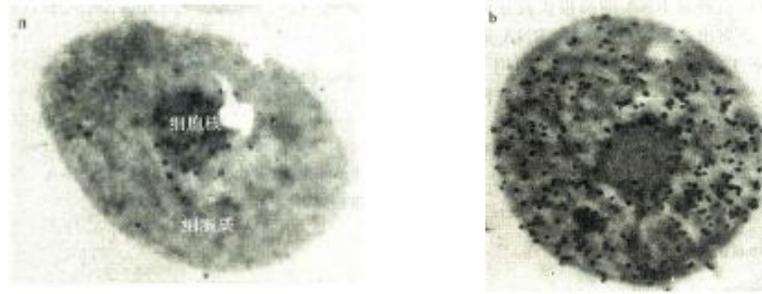


Fig. (1). Process of mRNA through nuclear membrane.

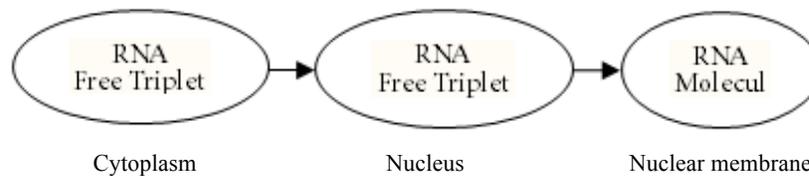


Fig. (2). Schematic diagram of the process of triplet through nuclear membrane.

2. The discovery of mRNA

Through indepth study it was discovered that RNA can be divided into three kinds, i.e. mRNA(messenger RNA), tRNA(transfer RNA) and rRNA(ribosomal RNA), with mRNA only accounting for a few percent in the cells. To code different proteins in the cells, the length of mRNA and composition of nucleotide vary widely. Within a given time, only a small section of mRNA binds to the ribosome, therefore, an mRNA can be examined simultaneously by multiple ribosomes.

3. mRNA moved to cytoplasm after synthesis in the nucleus

Until now, there is sufficient evidence to prove that RNA indeed moves from the DNA-containing nucleus to the cytoplasm containing ribosome [13]. Following this, the cells are briefly exposed to the radioactive labeling precursor, and an excess of unlabeled amino acids (pulse-labeled experimental) are added. mRNA synthesis manifests in a short period of time. The study shows that the mRNA synthesized in the nucleus. After about one hour, it was observed that most of the mRNAs left the nucleus and moved into the cytoplasm (Fig. 1) [12]. mRNA moves through the nuclear membrane in the protein complex called nuclear pore after being modified, which allows it to be expressed [13]. It is needed to be further explored in detail that how mRNA travels from the nucleus through the nuclear membrane to the cytoplasm.

2.3. The Derivation of Conjecture of Nucleotide Free Triplet

For the derivation of conjecture of nucleotide free triplet three consecutive nucleotides in the gene must have a triplet code and mRNA must move from the nucleus into cytoplasm through nuclear membrane. In this paper, the derivation

of conjecture of nucleotide free triplet was obtained at some points, and three consecutive nucleotides were bound as separate clusters. These clusters were made up of tiny triangles, called nucleotide free triplet composed of three nucleotides. For example, when mRNA moved through the nuclear membrane, three nucleotides formed a group and mRNA free triplet, and after free triplet moved through the nuclear membrane into cytoplasm, it restored the linear state and again synthesized mRNA molecules. The process of free triplet through nuclear membrane is shown in Fig. (2).

Cytoplasm Nucleus Nuclear membrane

3. TRIPLET CODE 20 EQUIVALENCE CLASSES DERIVED USING GROUP THEORY

3.1. Group Theory and Permutation Groups

1. Brief Introduction of Group Theory

Group theory is an important part of modern algebra. The concept of “group” was first put forward by the young French mathematician Galois (E. Galas, 1811-1832) and only defined the permutation group. Abstract group can be perceived as a collection of a class of objects, which have binary operator relations similar to multiplication; such operations satisfy the following properties:

⊖ Closure. If a and b are in the group, then $a \cdot b$ is also in the group;

⊖ Associativity. If a, b and c are in the group, then $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

⊖ There is an element I of the group such that for any element a of the group $a \cdot I = I \cdot a = a$;

⊖ For any element a of the group there is an element a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = I$.

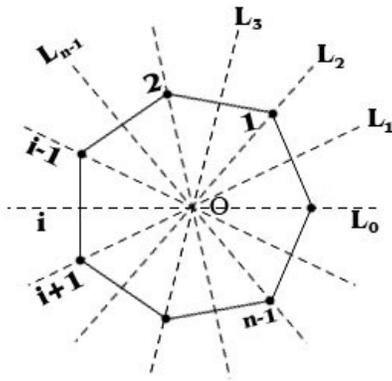


Fig. (3). Regular n quadrate.

In general, group theory is a universal tool to describe the symmetry ?? In the late 19th century, group theory was applied to study crystal structure, which had a wide range of applications in many areas. In modern physics, group theory has become a powerful tool for the study of elementary particles, and quantum mechanics. Algebra also experienced a renaissance due to the introduction and development of the concept of group theory [14].

2. Permutation Groups

In theoretical research and experimental application, permutation groups hold immense importance . On one hand, any finite group can be expressed by them; On the other hand, permutation groups are used to solve the problem of algebraic equations using square root. Permutation groups play an important role in the certification and application of Burnside lemma and Pólya theorem,. Permutation groups have a wealth of practical background, for example, these groups are used for the observation of proofs and application of polygon symmetry by observing all possible rotation and flipping that let polygon back to the original state, to obtain the permutation group called dihedral group [15].

3.2. Dihedral Group

Suppose $X=\{0,1,2,\dots,n-1\}$ (without loss of generality, $0\sim n-1$ was taken as sequence number) is the vertex set of the regular $n(n\geq 3)$ quadrate and arranged counter-clockwise, as shown in Fig. (3).

When the regular n quadrate was rotated according to $2\pi/n$ counter-clockwise, vertex i moved to the position originally occupied by vertex $i+1(\text{mod } n)$, so this rotation resulted in the conversion on X , marked as

R_1 :

$$R_1 = \begin{pmatrix} 0 & 1 & 2 & \dots & n-1 \\ 1 & 2 & 3 & \dots & 0 \end{pmatrix} \tag{1}$$

The conversion according to $2k\pi/n$ marked R_k :

$$R_k = \begin{pmatrix} 0 & \dots & n-1 \\ k & \dots & k+n-1 \end{pmatrix} \tag{2}$$

Where the addition and subtraction are the modulo n operation (as the same for the entire paper); R_1 is the identity; and R_k can be shown as:

$$R_k(i) = k + i, i = 0,1, \dots, n-1 \tag{3}$$

Another conversion is reflection in the symmetric axis according to π , named reflectivity conversion. Because of n symmetric axis, the axis was marked through vertex 0 as L_0 , and the axis through the vertex of the midpoint of edge marked $[0,1]$ as L_1, \dots , until L_{n-1} . The corresponding reflectivity conversion was marked as M_0, M_1, \dots, M_{n-1} . For instance,

$$M_n = \begin{pmatrix} 0 & 1 & \dots & n-1 \\ 0 & n-1 & \dots & 1 \end{pmatrix} \tag{4}$$

We can prove that

$$M_k = k + n - i, k = 0,1, \dots, n-1 \tag{5}$$

Let

$$D_n = \{R_k, M_k | k = 0,1, \dots, n-1\} \tag{6}$$

Following this, the D_n was closed under the composite operation of the conversion with the identity R_0 and each element was inverted, as a result, D_n formed the dihedral group [16].

3.3. Role of Group on Collection and Necklace Problem

The concept and theory of roles of groups in the collection provide theoretical basis for their application, and also provide powerful tools for the analysis of finite group. Let $X=\{1, 2, \dots, n\}$, G was a permutation group on X , for any $g \in G$ and $x \in X$, call $g(x)$ was the group element with g acting on x , and group G acting on collection X . X is called the target set. Here, $g(x)$ represents a reversible transformed g corresponding to X . The notion of permutation group on the role of the target set can be popularized to pop groups, and both Burnside lemma and Pólya theorem can solve the problem of track number (equivalence class) of collection acted on the group [15].

Necklace problem is described as making an n bead necklace using beads of m colors, so how many different types of necklace can be made?

When solving the necklace problem using Pólya theorem, initially the n -bead necklace is abstracted towards regular polygon of n vertices, and regular polygon exports the dihedral group, the equivalence necklace program number can be derived using Pólya theorem.

3.4. Nucleotide Free Triplet Vertex Set and Free Triplet Coding Complete Works

Nucleotide free triplet was abstracted to equilateral triangle (Fig. 4), whose vertex set can be defined as $X=\{1,2,3\}$. The set composed of four nucleotides is defined as $B=\{U,C,A,G\}$ Each mapping $f: X \rightarrow B$ represents a free triplet coding (it represents a label necklace in necklace problem). Let $\Omega=\{f|f: X \rightarrow B\}=B^X$ i.e. the coding completes the function of all free triplets, apparently $|\Omega|=|B|^{|X|}=4^3=64$,

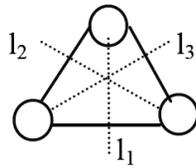


Fig. (4). Abstract model of free triplet.

Table 1. Elements of D3 group and the rotation numbers.

Elements of D ₃ Group	Group Elements Rotation Resolving Expression	Rotation Numbers in Group Elements $\lambda(g_k)$
g_0	(1) (2) (3)	3
g_1	(1 2 3)	1
g_2	(1 3 2)	1
g_3	(1)(2 3)	2
g_4	(2)(1 3)	2
g_5	(3)(1 2)	2

they are fit with the triplet genetic code set (see in the 9th reference) exactly when listed one by one.

3.5. 20 Equivalence Classes of Triplet Code Derived Using Group Theory

The dihedral group (D₃ group) was calculated corresponding to the triplet code, and the group elements R_k , $M_k(k=0,1,2)$ as $g_k(k=0,1,\dots,5)$ were represented as shown in Table 1 [17]. The equivalence class number of triplet code can be solved directly using Pólya theorem.

Pólya Theorem: Let G be a permutation group of the set of n objects, and n objects be dyed using m colors, then the number of nonequivalent colorings is given by

$$L_G = \frac{1}{|G|} \sum_{g_k \in G} m^{\lambda(g_k)} \tag{7}$$

Where $G = \{g_1, g_2, \dots, g_p\}$, $\lambda(g_k)$ is the number of rotations in the permutation g_k [15].

Solving the triplet code equivalence class is equivalent to the necklace problem of three beads dyed by four colors, where $m=4$, $|G|=6$, $\lambda(g_k)$ ($k=0,1, \dots, 5$) as shown in Table 2, these values are substituted in equation (7)

$$L_{D_3} = \frac{1}{6} (4^3 + 2 * 4^1 + 3 * 4^2) = 20 \tag{8}$$

Therefore, the triplet code set (64 code) can be divided into 20 equivalence classes using group theory.

Table 2. The triplet code table divided by 20 equivalence classes.

Equivalence Classes Number	Representative Elements	Triplet Code in Equivalence Classes	Code Number in Equivalence Classes	Subtotal
1	UUU	UUU	1	4
2	CCC	CCC	1	
3	AAA	AAA	1	
4	GGG	GGG	1	
5	UCC	UCC,CCU,CUC	3	9
6	UAA	UAA,AAU,AUA	3	
7	UGG	CGG,GGU,GUG	3	
8	CUU	CUU,UUC,UCU	3	9
9	CAA	CAA,AAC,ACA	3	
10	CGG	CGG,GGC,GCG	3	
11	AUU	AUU,UUA,UAU	3	9
12	ACC	ACC,CCA,CAC	3	
13	AGG	AGG,GGA,GAG	3	
14	GUU	GUU,UUG,UGU	3	9
15	GCC	GCC,CCG,CGC	3	
16	GAA	GAA,AAG,AGA	3	
17	UCA	UCA,CAU,AUC,ACU,UAC,CUA	6	24
18	UCG	UCG,CGU,GUC,GCU,UGC,CUG	6	
19	UAG	UAG,AGU,GUA,GAU,UGA,AUG	6	
20	CAG	CAG,AGC,GCA,GAC,CGA,ACG	6	

3.6. Representative Elements of 20 Equivalence Classes and Division of Equivalence Classes

64 triplet codes were divided into 20 equivalent classes and the results are shown in Table 2.

Initiating, equivalence, any triplet code in every equivalence class can be used as representative element, As shown in Table 2, triplet codes in every equivalence class had the same nucleotide composition. It is important for the analysis of triplet code and protein composition.

CONCLUSION AND FUTURE WORK

In this paper, the nucleotide free triplet conjecture of nucleotide is proposed and the research based on the abstract model of nucleotide free triplet using group theory method was conducted. It has been found that the total number of mapping from free triplet set of vertices to nucleotide collection was 64, which matches exactly with the element number of the triplet code set (64). 64 triplet codes were divided into 20 equivalence classes using group theory (Pólya Theorem), which were different from the triplet code set for 20 amino acids, but the inner link needs to be further studied in depth.

Studies have shown that there are 20 kinds of amino acids in nature and each amino acid can be represented by more than one triplet code. The triplet codes representing the same amino acid are generally different in the third base, which is called “silent codon”. Chinese writer Lu Xun said that “To explode in silence, or to die in it”. The secret role of “silent codon” may one day be found like recessive gene.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (NSFC) (61272006).

REFERENCES

- [1] Zhu Y, Li Y, Zheng Z. Modern Molecular Biology, 6th ed, Beijing, China: Higher Education Press 2011; pp.108-9.
- [2] Baranov PV. Codon size reduction as the origin of the triplet genetic code. *Plos One* 2009; 4: 1-9.
- [3] Petra S, Lucila OM. A robust feature selection method for novel pre-microRNA identification using a combination of nucleotide-structure triplets. *IEEE 2nd Conference on Healthcare Informatics, Imaging and Systems Biology: California, 2012*; p. 61.
- [4] Wang K, Schmiel WH, Chin JW. Reprogramming the Genetic Code: From Triplet to Quadruplet Codes. *Angew Chem Int Ed Engl* 2012; 51(10): 2288-97.
- [5] Fernando A, Michael F. Symmetry breaking in the genetic code: finite groups. *Math Comp Model* 2011; 53: 1469-88.
- [6] Tsvi T. A colorful origin for the genetic code: Information theory, statistical mechanics and the emergence of molecular codes. *Phys Life* 2010; 7: 362-76.
- [7] Wu B. Evolution of the genetic triplet code *via* two types of doublet codons. *J Mole Evol* 2006; 63: 54-64.
- [8] Sherwood GB. Functional dependence and equivalence class factors in combinatorial test designs. *IEEE 7th International Conference on Software Testing, Verification and Validation Workshops: United Kingdom 2014*; pp. 108-17.
- [9] Lu Y, Cha J. A fast algorithm for identifying minimum size instances of the equivalence classes of the pallet loading problem. *Eur J Oper Res* 2014; 237: 794-801.
- [10] Petsko GA, Ringe D, Xiaochun GT. Proteins structure and function. Beijing: Science Press 2009, pp. 6-7.
- [11] Zheng T, Yao F. Science and the mysteries of life. Knowledge Beijing, Press 2003, pp. 28-29.
- [12] Waston JD. Molecular biology of the gene. 6th ed. Beijing: Science Press 2010, pp. 40-60.
- [13] Sylvain WL, Wang Y. Introduction to molecular biology. Beijing: Chemical Industry Press 2008, pp.34-6.
- [14] Li W. Introduction to the history of mathematics. 3rd ed. Beijing: Higher Education Press 2005, pp. 211-2.
- [15] Vanlint JH. Combinatorial mathematics tutorial. Beijing: Mechanical Industry Press 1993, pp. 189-207.
- [16] Hu G. Application of modern algebra. 2nd ed. Beijing: Tsinghua University Press 1999; pp. 47-9, 103-8.
- [17] Zhang D. Study of key technology of three-dimensional for LEO/MEO satellite networking based on virtual reality. MS thesis, Shenyang: Northeastern University 2004, pp. 26-7.

Received: May 16, 2015

Revised: August 23, 2015

Accepted: September 31, 2015

© Dakun *et al.*; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.