

TOPOLOGY IN GENETIC CODE ALGEBRA

TAZID. ALI, CHANDRA KANTA PHUKAN

Abstract: Based on physico-chemical properties the four RNA bases can be obtained in two different orders. An addition operation can be defined on these sets of bases such that it is isomorphic to Z_4 , i.e., the group of integers (modulo 4). Depending on this addition operation, an operation is established on the set of codons, such that the set of codons is a group isomorphic to Z_{64} . In this paper we have equipped one of the groups (called primal group) with a topology based on the chemical similarity of the nucleotides such that it is a topological ring. This topological ring structure imposes a topological ring structure in the set of codons. The structure in the set of codons is in fact evolution algebra. We have also observed some interesting relation between the algebraic structure and biological properties of the codons.

Keywords: Algebra, Genetic Code, Topology

Introduction: Proteins are the basic constructional blocks and functional elements of living organisms. They are involved in all processes occurring in cells and tissues. Proteins are assembled in cell organelles called ribosomes that are located outside the cell nucleus; the process of their production is called translation. Proteins are chain of amino acids. Amino acids are synthesized by RNAs and RNAs are obtained from DNAs. A sequence of three nucleotides forms a unit called codon. A codon encodes an amino acid. Ribosomes are complexes of proteins and RNA. Two types of RNA molecules are involved in this process, mRNA (messenger RNA), which consists of long strands of RNA that carries a plan for the construction of protein. This construction is written in their codons. Other type of RNA is tRNA (transfer RNA), which consists of short RNA molecules, is dedicated to transport of a single unit of protein, amino acid. Proteins are molecules with a complicated three-dimensional structure, but they always have an underlying linear chain of amino acids as their primary structure. The structure of an amino acid contains a C_α carbon atom, C' carbon atom, an amine group NH_2 , a carboxyl group $COOH$ and a side chain R.

There are 20 different amino acids that occur in proteins, which differ from each other by the chemical structures of their R-groups. The properties of the amino acids are dictated by the chemical characteristics of their R- groups. Depending upon the size, structure, electrical properties and chemical properties and many other characteristics by which one can describe properties of amino acids in proteins

The origin of the genetic code is a notoriously difficult problem of biology, which has haunted biologists and mathematicians alike. Physical systems are governed by a number of conservations laws. Initially there are 64 codons as there are four bases, but only 20 amino acids and a set of stop codons. It suggests a symmetry breaking which requires the

existence of a potential to reduce the degeneracy. A group theoretical approach could be used to a great advantage to understand this degeneracy of the universal genetic code (Al-Zaharani et al., 2007). The group theoretical technique is indeed able to give a succinct account of many of the currently understood aspects of the evolution of the genetic code and the observed degeneracy structure of the codon-amino acid correspondences (Bashford et al., 1998).

Primal Genetic Code Algebraic Structure:

The order of the genetic code is strongly connected to the physico-chemical properties of the four bases, in particular to the base positions in the codon and to the codon-anticodon interactions, where chemical types (purine and pyrimidine) and number of hydrogen bonds have played an important role. The importance of the base position is suggested by the error frequency found in the codons. Errors on the third base are more frequent than on the first base, and, in turn, these are more frequent than errors on the second base. These positions, however, are too conservative with respect to changes in polarity of the coded amino acids (Sánchez et al., 2005).

Two orders of DNA bases can be obtained based on their physico-chemical properties. The chemical types (purine and pyrimidine) and the number of hydrogen bonds are the basic facts, which determine the two orders in the base set. Following are the important criteria that are applied to determine the order (Sánchez et al., 2005).

Chemical types cause the main difference between bases. The greatest difference between one element and the next serve as criterion to select arrangements. The starting base needs a minimum hydrogen bond number.

As a result, the two orders $\{A, C, G, U\}$ and $\{U, G, C, A\}$ in these base set are specified. It is possible to define a sum operation in several ways on the two sets. In Tables 2.1 A and B shown below are the sum tables of bases obtained from the two possible orders. These two groups are eventually isomorphic to Z_4

(group Z_4 of integers modulo 4).Therefore, there are two cyclic groups: the primal and the dual group, corresponding to the ordered sets {A, C, G, U} and {U, G, C, A}.

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

A: Primal algebra B: Dual algebra

Table 2.1 Sum Operation of four bases of the DNA.

Some Mathematical Definitions:

Definition 3.1 Let (A, \bullet) be an algebra over ring R. If it admits a countable basis $x_1, x_2, \dots, x_n, \dots$, such that

$$x_i \bullet x_j = 0 \text{ if } i \neq j$$

$$x_i \bullet x_i = \sum_k a_{i,k} x_k \text{ for any } i$$

we then call this algebra an evolution algebra (Tian , 2007).

Definition 3.2 A topological group is a set G with two structures

- (i) G is a group with respect to \bullet
- (ii) G is a topological space,

such that the two structures are compatible i.e., the multiplication map $\mu : G \times G \rightarrow G$ and the inversion map $\nu : G \rightarrow G$ are both continuous. In this definition the set $G \times G$ carries the product topology.

Definition 3.3 A topological ring is a set R with two structures

- (i) R is a ring with respect to the operation “+” and “ \bullet ”
- (ii) R is a topological space,

such that the two structures are compatible i.e., the addition map $\eta : R \times R \rightarrow R$, the multiplication map $\mu : R \times R \rightarrow R$ and the inversion map $\nu : R \rightarrow R$ (additive inverse) are continuous. In this definition the set $R \times R$ carries the product topology

Definition 3.4 (Balachandran, 1999) A algebra A over K is called a topological algebra if it is equip with a topology such that

- (i) The map $(x, y) \rightarrow x + y$ of $A \times A \rightarrow A$ is continuous.
- (ii) The map $(\lambda, x) \rightarrow \lambda x$ of $K \times A \rightarrow A$ is continuous.
- (iii) The map $(x, y) \rightarrow x y$ of $A \times A \rightarrow A$ is continuous.

Algebraic Structures On Genetic Code: Many attempts have been made to introduce a formal characterization of the genetic code (Balakrishnan, 2002; Bashford et al., 1998; Bashford et al. 2000; Beland and Allen, 1994; Bertman and Jungck, 1979; Jiménez-Montaño, 1996, 1999; Karasev, and Stefanov, 2001; Siemion et al., 1995). Many of these models highlight a quantitative relationship between codons expressed through the hydrogen bonds and the chemical types of bases. Eventually, these formal

descriptions suggest that the hydrogen bond number and the chemical type should be enough to obtain a “natural order” in the codon set (Sánchez et al., 2005). Recently, using these elemental properties of bases Sánchez et al.(2004a, 2004b) proposed a Boolean structure of the genetic code in which the partial order of the codon set and the Boolean deductions between codons are connected to the physico-chemical properties of amino acids. Here, nevertheless, an effort has been made to show that by using the same base properties, it is possible to deduce a different codon order and a different algebraic and topological structure of the genetic code.

4.1 Ring Structure on Genetic Code: In this paper an attempt has been made to create an alternative algebraic structure of genetic code other than primal or dual, by upholding the same order of DNA bases as in the primal genetic code structure. Here, the addition operation of the group of primal order $P = \{A, C, G, U\}$ is well indistinguishable with the primal genetic code. However a newly defined product operation is considered in the set $\{A, C, G, U\}$ of primal order of genetic code structure as

	+	A	C	G	U		\bullet	A	C	G	U
A	A	C	G	U		A	A	A	A	A	A
C	C	G	U	A		C	A	C	G	U	U
G	G	U	A	C		G	A	G	A	G	G
U	U	A	C	G		U	A	U	G	C	C

Table 4.1 Sum and Product Operations on $P = \{A, C, G, U\}$

With these two binary operations the set $P = \{A, C, G, U\}$ fulfil the axioms of a ring structure. In the ring $(P, +, \bullet)$, A is the additive identity and C is the multiplicative identity. For all $x, y \in P$ the product binary operation satisfies the commutative law: $x \bullet y = y \bullet x$. So, P has commutative ring structure with identity element. Binary operations have been defined over here so as to set an isomorphism between $(P, +, \bullet)$ and ring of integer modulo 4 i.e., $(Z_4, +, \bullet)$ ring . The isomorphic image of A, C, G and U are 0, 1, 2 and 3 respectively.

The Cartesian product of the ring P i.e., $P \times P \times P$ apparently provide genetic code (set of codons), denote it by C_G . where,

$$P \times P \times P = \{(X, Y, Z) : X, Y, Z \in \{A, C, G, U\}\} \text{ i.e., } C_G = \{XYZ : X, Y, Z \in \{A, C, G, U\}\},$$

Each codon of the form XYZ is associated with the element (X, Y, Z) of $P \times P \times P$. So, a one to one correspondence can be established between set $P \times P \times P$ and C_G . Note that, since P is compatible with ring structure thus $P \times P \times P$ will also possess ring structure with the binary operations defined by

$$(X, Y, Z) + (X', Y', Z') = (X+X', Y+Y', Z+Z')$$

$$(X, Y, Z) \bullet (X', Y', Z') = (X \bullet X', Y \bullet Y', Z \bullet Z').$$

Again, P is isomorphic to Z_4 so, $P \times P \times P$ is isomorphic to $Z_4 \times Z_4 \times Z_4$, consequently the elements of C_G and $Z_4 \times Z_4 \times Z_4$ acquire one to one correspondence. For example an element $GAC \in C_G$ has correspondence with the element $(2, 0, 1) \in Z_4 \times Z_4 \times Z_4$.

4.2 C_G as Module over P: Since, Z_4 is a ring with identity, so it can be treated as an additive abelian group and so, $Z_4 \times Z_4 \times Z_4$ can also be viewed as an additive abelian group with usual addition operation on $Z_4 \times Z_4 \times Z_4$. Now, considering Z_4 as a ring and $Z_4 \times Z_4 \times Z_4$ as an additive Abelian group the mapping

$$f : Z_4 \times (Z_4 \times Z_4 \times Z_4) \rightarrow Z_4 \times Z_4 \times Z_4$$

is such that $f(x, u) = x u = u x$, for all $x, y \in Z_4$ and for all $u, v \in Z_4 \times Z_4 \times Z_4$ gives a module structure on $Z_4 \times Z_4 \times Z_4$. Analogously, in the set of codons C_G a module structure over P can be achieved. It can be shown that $Z_4 \times Z_4 \times Z_4$ is a free module over Z_4 . In a similar way, a free module structure can be obtained in the set of codons C_G over P. Here, the basis of C_G can be represented by the elements CAA, ACA, AAC.

We have established the following

Result 4.1 For every pair (α, β) of elements of C_G , there is an endomorphism $f : C_G \rightarrow C_G$, such that $f(\alpha) = \beta$ if and only if, the order of β is a divisor of the order of α .

Proof: Since, $C_G \cong Z_4 \times Z_4 \times Z_4$ it is sufficient to prove the result for $Z_4 \times Z_4 \times Z_4$.

Order of any element $(a, b, c) \in Z_4 \times Z_4 \times Z_4$ is the highest order of the components. Since, elements of Z_4 have order 1, 2 or 4, any elements of $Z_4 \times Z_4 \times Z_4$ will be of order 1, 2 or 4. Again, any elements of $\alpha \in Z_4 \times Z_4 \times Z_4$ can be expressed as

$$\alpha = 2^{2-m} \alpha_i, \text{ where } \alpha \text{ is of order } 2^m \text{ (} m = 1, 2 \text{) and } \alpha_i \text{ is of order } 4.$$

Let $\alpha, \beta \in Z_4 \times Z_4 \times Z_4$ and let $o(\alpha) = 2^r, o(\beta) = 2^s$. suppose \exists an endomorphism of $f : Z_4 \times Z_4 \times Z_4 \rightarrow Z_4 \times Z_4 \times Z_4$ such that $f(\alpha) = \beta$.

Then

$$o = f(o) = f(2^r \alpha) = 2^r f(\alpha) = 2^r \beta \Rightarrow o(\beta) / 2^r = o(\alpha).$$

Conversely, let $o(\alpha) / o(\beta)$.

$$\text{We have } \alpha = 2^{2-r} \alpha_1 \text{ and } \beta = 2^{2-s} \beta_1,$$

where, $\alpha_1, \beta_1 \in Z_4 \times Z_4 \times Z_4$ are of order 4 each.

Now, we can construct two basis $(\alpha_1, \alpha_2, \alpha_3)$ and $(\beta_1, \beta_2, \beta_3)$ of $Z_4 \times Z_4 \times Z_4$ by suitably choosing α_2, α_3 and β_2, β_3 in $Z_4 \times Z_4 \times Z_4$. Then there exists an automorphism say $g : Z_4 \times Z_4 \times Z_4 \rightarrow Z_4 \times Z_4 \times Z_4$ such that $g(\alpha_i) = \beta_i$.

Since, $o(\beta) = o(\alpha), s \leq r$. Consider the map $f = 2^{r-s} g$. Then clearly g is an endomorphism on $Z_4 \times Z_4 \times Z_4$ and

$$f(\alpha) = 2^{r-s} g(\alpha) = 2^{r-s} g(2^{2-r} \alpha_1)$$

$$= 2^{r-s} 2^{2-r} g(\alpha_1) = 2^{2-s} \beta_1 = \beta.$$

Hence the result

Corollary 4.1 For any $\alpha, \beta \in C_G$ there exists a diagonal endomorphism $f : C_G \rightarrow C_G$ if and only if for every pair of coordinates $\alpha_i, \beta_i \in P$ of the vectors α and β the inequalities $m_{\beta_i} \leq m_{\alpha_i}$ holds [Here, $2^{m_{\beta_i}}$ is the order of β_i etc].

4.3 Evolution Algebra on C_G : We have Z_4 is a commutative ring and $Z_4 \times Z_4 \times Z_4$ is a module over Z_4 , which satisfies the condition $\alpha(xy) = (\alpha x)y = x(\alpha y)$ for $\alpha \in Z_4$ and $x, y \in Z_4 \times Z_4 \times Z_4$.

So, $Z_4 \times Z_4 \times Z_4$ module can be extended to the structure of algebra. Similar situation will also happen in the genetic code structure. Ultimately genetic code C_G can be viewed as an algebra. Now, it is observed that the basis $B = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ of $Z_4 \times Z_4 \times Z_4$ shows some important properties, which eventually leads us to see $Z_4 \times Z_4 \times Z_4$ to be an evolution algebra (Definition 3.1). Because, $(Z_4 \times Z_4 \times Z_4, \bullet)$ is an algebra over ring Z_4 and it admits a countable basis $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ such that

$$x_i \bullet x_j = 0 \text{ if } i \neq j \text{ since,}$$

$$(1, 0, 0) \bullet (0, 1, 0) = (0, 0, 0)$$

$$(1, 0, 0) \bullet (0, 0, 1) = (0, 0, 0)$$

$$(0, 1, 0) \bullet (0, 0, 1) = (0, 0, 0) \quad \text{and}$$

$$x_i \bullet x_i = \sum_k a_{i,k} x_k \text{ for any } i \text{ since,}$$

$$(1, 0, 0) \bullet (1, 0, 0) = (1, 0, 0) = 1(1, 0, 0)$$

$$(0, 1, 0) \bullet (0, 1, 0) = (0, 1, 0) = 1(0, 1, 0)$$

$$(0, 0, 1) \bullet (0, 0, 1) = (0, 0, 1) = 1(0, 0, 1).$$

Consequently, one important algebraic structure, evolution algebra is found in genetic code.

Topology On Genetic Code: As we have mentioned that because of chemical similarity, the bases A, C, G, U can be divided into two categories $\{A, G\}$ as purines and $\{C, U\}$ as pyrimidines. Taking these two categories as open sets, a topology can be equipped on the base set $P = \{A, C, G, U\}$ defined as $\mathfrak{T} = \{\emptyset, P, \{A, G\}, \{C, U\}\}$. With respect to this topology P is a topological algebra.

Topology On Amino Acids: All 64 possible 3-letter combinations of the DNA coding units A, C, G and T are used either to encode one of 20 amino acids or as one of the three stop codons. Most amino acids have multiple codons. For that reason the amino acids divide the set of codons into disjoint equivalent classes.

Since multiple codons encode the same amino acids, we get a many-one mapping from the set of codons C_G i.e., the set of amino acids say A_M , which also contains stop codons. This mapping is called the canonical map. So it is natural to equip the set of amino acids by the quotient topology.

We have observed the following

i. All hydrophilic amino acids are zero divisors

ii. Inverse of hydrophobic amino acids are hydrophobic

/ stop codons.

iii. Inverse of hydrophilic amino acids are hydrophilic/ stop codons

iv. Stop codons form a basis of C_G .

v. Amino acids coded by only two codons are either both zero divisor or both non-zero divisors.

Conclusion: Some well known algebraic structures: groups, rings, modules and algebras that have been built from the four DNA bases and from which later developed the Genetic Code algebraic structures. We have introduced an algebraic structure on C_G induced from that of the primal group P . We have also introduced a topology on the primal group by considering the purines and pyrimidines as open sets.

References :

1. Al-Zaharani, A. O., Al-Ahmadi, M. S. G., and Siddiqui, K. A. (2007): The Genetic Code: A Physicist's View Point, Karachi University Journal of Science, 35,1-41.
2. Balachandran V.K. (1999): Topological Algebras, Narosa Publishing House.
3. Balakrishnan, J.(2002): Symmetry scheme for amino acid codons, Phys. Rev. E, 65, 021912-5
4. Bashford, J.D. and Jarvis P.D.(2000): The genetic code as a periodic table, Biosystems 57, 147-161.
5. Bashford J. D., Tsohantjis, I., and Jarvis, P. D. (1998): A supersymmetric model for the evolution of the genetic code, Proc. Natl. Acad. Sci. USA, 95, 987-992.
6. Beland, P. and Allen, T.F. (1994): The origin and evolution of the genetic code, J Theor Biol.170, 359-365.
7. Bertman, M.O. and Jungck, J.R. (1979): Group graph of the genetic code, J. Hered. 70, 379-384.
8. Jiménez-Montaña, M.A. (1999): Protein Evolution Drives the Evolution of the Genetic Code and Vice Versa. Biosystems 54, 47-64.
9. Jiménez-Montaña, M.A., (1996): The hypercube structure of the genetic code explains

This induces topology on C_G such that C_G is a topological algebra. There is another natural way of introducing a topological structure on the set of codons. That is by considering a pretopology where vicinity of each codon is the set of all its one-error mutants. It would be interesting to study relation between this pretopology and the algebraic structure on C_G .

Acknowledgement :

The first author is thankful to the Department of Science and Technology, Govt. of India for providing financial support under a research project.

- conservative and non-conservative amino acid substitutions in vivo and in vitro, Biosystems 39, 117-125.
10. Karasev, V.A. and Stefanov, V.E. (2001): Topological Nature of the Genetic Code, J. Theor. Biol. 209, 303-317.
11. Sánchez, R., Morgado, E., and Grau, R. (2004b): A Genetic Code Boolean Structure. I. The Meaning of Boolean Deductions. Bull. Math. Biol. (article in press).
12. Sánchez, R., Morgado, E., and Grau, R. (2004a): The Genetic Code Boolean Lattice. MATCH Commun. Math. Comput. Chem 52, 29-46.
13. Sanchez, R., Morgado, E. and Grau, R., (2005): Gene Algebra from a genetic Code algebra structure, J. Math. Biol, 51, 431 - 457.
14. Siemion, I.Z., Siemion, P.J., Krajewski, K., (1995): Chou-Fasman conformational amino acid parameters and the genetic code, Biosystems 36, 231-238.
15. Tian, J.P., (2007): Evolution Algebras and their Applications Springer Science + Business Media.

¹Associate Professor, Department of Mathematics, Dibrugarh University. Dibrugarh, Assam, India, email: tazidali@yahoo.com

² Research Scholar, Department of Mathematics, Dibrugarh University. Dibrugarh, Assam, India, email: chandra_kanta25@yahoo.com