The Genetic Code: A Physicist's View Point

A.O. Al-Zaharani, M.S.G. Al-Ahmadi and K.A. Siddiqui*

Department of Physics, King Abdulaziz University, P.O. Box 80034, Jeddah 21589, Kingdom of Saudi Arabia

Abstract: The genetic code is degenerate. Degeneracy suggests an underlying symmetry of a system. The symmetry breaking process which generates the existing degeneracy of the code is important from a perspective of a physicist. The genetic code and its associated symmetries are described under a lie algebraic approach. A spin space equivalent to a physical Hilbert space of the system is developed for the discussion of the degeneracy of the genetic code.

Key Words: Genetic Code, Lie Algebra, Equivalent Spaces.

INTRODUCTION

Proteins are the structural elements and the machines of life; they form all the elements and perform the important tasks that a living system needs. Proteins are built from twenty different building blocks, amino acids and each amino acid is a triplet code of four possible bases. In constructing a protein, nature covalently links of the order of 100 to several hundred amino acids into a linear polypeptide chain. Even the most diverged organisms use the same set of 20 amino acids for de novo synthesis of virtually all proteins. The universal genetic code includes 20 common amino acids and a triplet of codons known as stop codons [1,2]. The origin of the genetic code is a notoriously difficult problem of biology, which has haunted biologists and mathematicians alike [3-5].

Physical systems are governed by a number of conservations laws. These laws are often constraints placed upon the behaviour of a system by an underlying symmetry. The definition of symmetry as invariance under a specified group of transformations allowed the concept to be applied much more widely not only to spatial figures but also to abstract or biological objects such as genetic code. 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 great advantage to understand this degeneracy of the universal genetic code.

It is known that a messenger ribonucleic acid (mRNA) is composed of a sugar spine along which are attached 20 different amino acids. Each of these is constructed utilizing four bases (nucleotides) taken three at a time in all possible permutations, thereby forming codon sequences. Each nucleotide is paired in deoxyribonucleic acid (DNA) with another. A known property of DNA and RNA is that the triplet sequences are precise and the codon assignments to amino acids are known.

*Address correspondence to this author at the Department of Physics, King Abdulaziz University, P.O. Box 80034, Jeddah 21589, Kingdom of Saudi Arabia

The DNA molecule is made of two linear chains of nucleotides wrapped in a double helix structure [6].

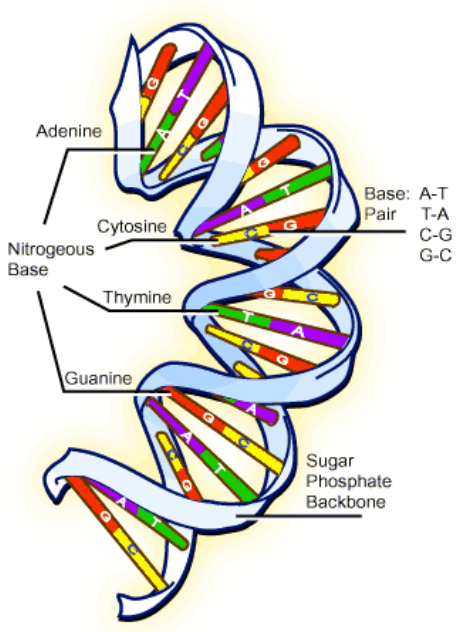


Fig. (1). DNA Double Helix.

Each nucleotide is characterized by one of the four elementary bases: Adenine (A) and Guanine (G) are purine, and Cytosine (C) and Thymine (T) are pyrimidine. DNA is localized in the nucleus of the cell and the transmission of the genetic information in the cytoplasm is achieved, by the ribonucleic acid, RNA. The operation is called the transcription. The assembly of proteins, which are built up from chains of amino acids is sometimes assisted by enzymes. The correspondence law between triplets of nucleotides, called codon, in the DNA sequence and the amino acids is called the genetic code. (A – T) and (C – G) are complementary bases, they always join each other along the double helix. As a codon is an ordered sequence of three bases, obviously there are 64 different codons. In RNA, Uracil (U) replaces Thymine (T).

0250-5363/07

It can be noticed that except the three triplets, TAA, TAG and TGA, each of the 61 others is related to an amino acid. Three codons mentioned above are called Nonsense or Stop codons, and their role is to stop the biosynthesis [7]. 20 different amino acids can be distinguished. It implies that the genetic code as given in Table **1**, is degenerate. There are 3 sextets (Arginine, Leucine, Serine), 5 quadruplets (Alanine, Glycine, Proline, Threonine, Valine), 2 triplets (Isoleicine, Stop codons), 9 doublets (Asparagine, Aspartic acid, Cysteine, Glutamine, Glutamic acid, Histidine, Lysine, Phenylalanine, Tyrosine), 2 singlets (Methionine, Tryptophane). This degeneracy of the genetic code suggests a primordial symmetry breaking and group theoretic approach can be used to understand the behaviour of the system. The possibility of underlying continuous symmetries being involved has been discussed by many authors [8,9].

A GROUP THEORETIC APPROACH

The key idea of using group theoretical symmetry breaking to model the dynamical process of code evolution is an extremely powerful one. The 64 codons can be considered as a 64 dimensional irreducible representation of some appropriately chosen Lie group.

It is assumed that, in the beginning of the evolution, they were all supposed to code for one amino acid. It implies that symmetry breaking has occurred in stages to reduce the degeneracy to achieve the present status. If four bases, A, C, T and G are represented by four vectors in a 4-dimensional space then triplets of codons will generate a 64 dimensional space.

If One Consider the Following Vectors

$$A=\begin{pmatrix}1\\0\\0\\0\end{pmatrix} \qquad C=\begin{pmatrix}0\\1\\0\\0\end{pmatrix} \qquad T=\begin{pmatrix}0\\0\\1\\0\end{pmatrix} \qquad G=\begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

And the raising and lowering operators, L_+ and L_-, operating in a 4 dimensional space, then these operators obey the commutation rules as follows:

$$[\, L_Z\, ,\, L_+\,] = L_+,\ [L_Z,\, L_-\,] = L_-\, ,\ [L_+\, ,\, L_-\,] = 2\, L_Z$$

A generalization of these relations defines a general Lie algebra as a set of large matrices [10]. Now we have to look for a group that has a 64 dimensional representation. The Cartan classification theorem [11] states that there are only nine different families of Lie algebras labeled as A_n, B_n, C_n, D_n, E_6, E_7, E_8, F_4 and G_2. Only 8 of these have a 64 dimensional representation. They are given as A_1, A_2, A_3, B_6, C_2, C_3, D_7 and G_2. These algebras are named as SU(2), SU(3), SU(4), SO(13), Sp(4), Sp(6), SO(14) and G2. It is known that algebra of rank r breaks in algebra of rank smaller or equal to r. Eventually the chains of all these algebras stop at SU(2) group, which is the group of the matrices corresponding to operators L_Z, L_+ and L_-. These operators relate the bases amongst themselves.

Hornos & Hornos (1993) [12] adopted the 64 dimensional irreducible representation of the Symplectic group,

Table 1. A Standard Genetic Code

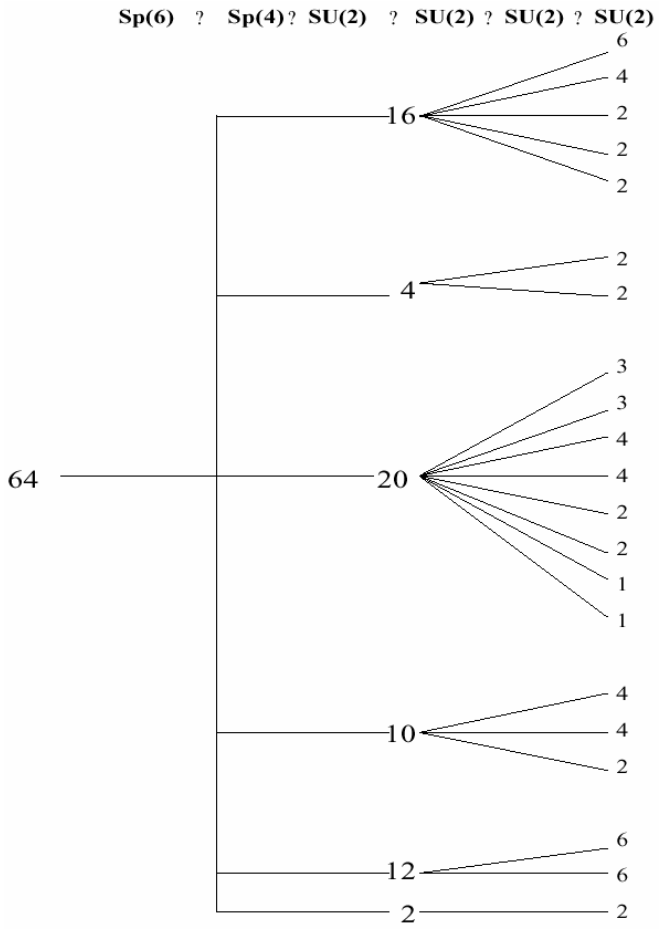
TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC		TCC		TAC		TGC	
TTA	Leu	TCA		TAA	stop	TGA	stop
TTG		TCG		TAG		TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC		CCC		CAC		CGC	
CTA		CCA		CAA	Gln	CGA	
CTG		CCG		CAG		CGG	
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC		ACC		AAC		AGC	
ATA		ACA		AAA	Lys	AGA	Arg
ATG	Met	ACG		AAG		AGG	
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC		GCC		GAC		GGC	
GTA		GCA		GAA	Glu	GGA	
GTG		GCG		GAG		GGG	

Sp(6), of quaternionic matrices. Sp(6) chain Sp(4) ⊗ SU(2) is the one that reproduces the genetic code in four stages. Sp(6) breaks down to the following chain

Sp(6) ⊃ Sp(4) ⊗ SU(2) ⊃ SU(2) ⊗ SU(2) ⊗ SU(2)

Finally, the second and the last SU(2) breaks down, SU(2) → O(2) and it decomposes the 64 fold degeneracy of the genetic code into the existing one as shown in Table **2**.

Table 2. Lie Algebraic Degeneracy of the Code



AN EQUIVALENT SPIN SPACE APPROACH

It is known that every elementary particle has a specific value of spin *s*. For example, pi mesons have spin 0; electrons have spin 1/2; photons have spin 1; *deltas have spin 3/2*; gravitons have spin 2; and so on. For a particle of spin *s*, there is a corresponding m_S which takes (-s, -s+1, -s+2, ….., +s) values.

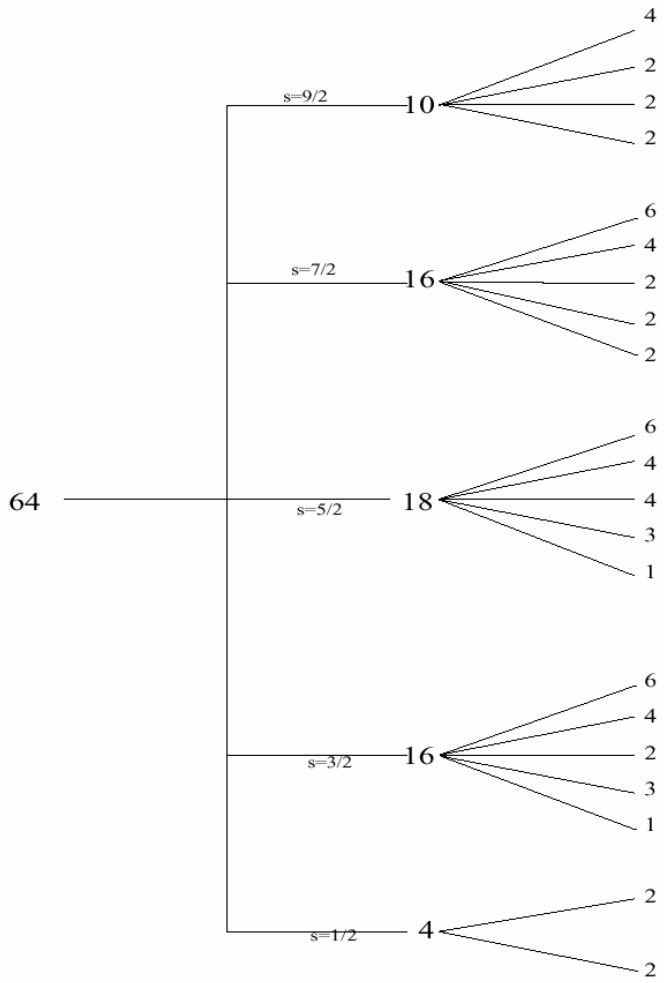
Quantum dynamics is inherently unseparable from quantum measurement, atleast for biological systems. A possible approach is to generate a 64 dimensional equivalent spin space which will represent the codon space. It can be generated as a direct product space of three particles of spin 3/2 interacting together. Hence, 3/2 ⊗ 3/2 ⊗ 3/2, will generate a 64 dimensional affine Hilbert space. It has already been said that this basic space is arising from the group SU(2). The spin-spin interaction among the particles reduces the 64 dimensional space into a set of one 10-dimensional space, two 8-dimensional spaces, three 6-dimensional spaces, four 4-dimensional spaces and two 2-dimensional spaces.

3/2 ⊗ 3/2 ⊗ 3/2 = 3/2 ⊗ [3 ⊕ 2 ⊕ 1 ⊕ 0] = [9/2 ⊕ 7/2 ⊕ 5/2 ⊕ 3/2]

⊕ [7/2 ⊕ 5/2 ⊕ 3/2 ⊕ 1/2] ⊕ [5/2 ⊕ 3/2 ⊕ 1/2] ⊕ 3/2

Finally an exchange type Hamiltonian, $H_{ex} = \Sigma_{i,j} S_i \bullet S_j$, will break the symmetry further and generate the required degeneracy of the genetic code as given in Table **3**.

Table 3. Degeneracy as Generated by Equivalent Spin Space



CONCLUSIONS

Physicists pursuing the fundamental reasons why the genetic code is just the way it is, they found an outstanding

feature of the genetic code - Degeneracy. The degeneracy is associated with a consequence of symmetry. A lie algebraic approach to genetic code provides a set of symmetry breaking as satisfying the possible degeneracy of the code. This approach does not generate a detailed microscopic biological, physical and chemical analysis of the genetic code. It is established that symmetry considerations alone can not replace a microscopic model but just establish a general background. It is emphasized that group theoretic analysis provides only tentative codon assignments.

A generic Hamiltonian operator has been invoked to provide a semi quantitative measure. As Hamiltonians and irreducible representations of a group, both require a space, therefore, an equivalent space approach has been adopted. It is expected that a spin Hamiltonian operating in the equivalent space will generate the known degeneracy of the genetic code. The detailed formulation of the spin Hamiltonian is not the target of the present study but it can be pursued further.

Developing a microscopic model remains the major task for further work. A more ambitious project is to correlate the numbers that define each amino acid with important protein properties such as folding. The field being so controversial, it is only natural that we do not want to claim to have solved the problem in the present study. Solving this problem in more specific terms, however, is only a challenge at this time, requiring further investigations.

REFERENCES

[1] Crick, F.H.C., Barnett, L., Brenner, S. and Watts-Tobin, R.J. (1961) Nature, 192, 1227.

[2] Crick, F.H.C. (1966) Sci. Am., 215, 55.

[3] Shimizu, M. (1982) J. Mol. Evol., 18, 297.

[4] Shimizu, M. (1983) Origin and Evolution of Hereditary-Metabolic Systems, ISAS Research Note 220.

[5] Jukes, T.H. (1983) Evolution of Genes and proteins. In: Evolution of the Amino Acid Code, M. Neir and P.K. Kohn (Eds.), Sinauer Assoc. Inc., Sunderland, Massachusetts.

[6] Woese. C.R. (1967) The Genetic Code, Harper and Row, New York.

[7] Jestin, J.L. (2006) C.R. Biologies, 329, 168.

[8] Bashford, J.D. and Jarvis, P.D. (2000) Biosystems, 57, 147.

[9] Forger, M., Hornos, Y.M.M., and Hornos, J.E.M. (1997) Phys. Rev. E, 56, 7078.

[10] Bashford, J.D., Tsohantjis, I. and Jarvis, P.D. (1997) Phys. Lett. A233, 481.

[11] Mckay, W. and Patera, J. (1981) Tables of Dimensions, Indices, and Branching Rules for Representations of Simple Lie Algebras, Marcell Dekker, New York.

[12] Hornos, J.E.M. and Hornos, Y.M.M. (1993) Phys. Rev. Lett., 71, 4401.