

Combinatorics, Homology and Symmetry of codons

Vladimir Komarov

Department of Chemistry, St Petersburg State University, St Petersburg, Russia

*Corresponding author: kvsergeich@gmail.com

Abstract

In nuclear physics and elementary particle theory concept of unitary symmetry and the related idea of the hierarchy of interactions play a fundamental role [1, 2]. So the relative smallness of the electromagnetic and weak interactions as compared to the strong interaction of the nucleons in the nucleus can be considered a model of the nucleus in the limit of exact symmetry of the strong interactions. In this model, protons and neutrons are physically indistinguishable states of the nucleon, and the properties of the nucleus are invariant under isotopic transformations.

In the case of molecules, we can also talk about the hierarchy of interactions involved in their formation.

As an example of a "strong" interaction here we can point to energy of chemical bonds, which is 1-2 orders of magnitude more energy non-bonded interactions. Another example - when the energy of valence interactions is much greater than the energy of intermolecular bonds in the condensed medium.

Usually accounting of weak interactions in 'chemistry is performed by introducing a physical model of various perturbations. These perturbations typically are unmeasured parameters that; are essentially the fitting values.

However, in the preferred class of molecules can try to find such values of the parameter in the ratio in which the contributions of the "weak" interactions are compensated or negligible.

Symmetry approach is important in estimation of reliability of experimental data and to predict new values of a parameter.

The same, from the standpoint of finding a unitary symmetry, the approach would be interesting to extend to more complex molecules and molecular systems. Up until genetic.

The application of the previously developed concepts of symmetry to the codon is the purpose of this work.

Keywords: Codons, Combinatory, Homology, Homologous series, Unitary Symmetry.

1. Introduction

It was shown earlier [1,2,3,4,5,6,7], that for objects, starting with elementary particles and to complex molecules and their mixtures, combinatorial operations lead to the formation of a homologous series (series of combinatorial objects). These homologous series have a universal property - in the space of physical and chemical parameters is saved some of the parameter in relation to certain pairs of combinatorial objects. This versatile property has been called unitary symmetry $SU(n)$.

Mathematical interpretation of the $SU(5)$ as an example of the halogenated methane was given in [2]. In [7] were able to show the physical meaning of unitary symmetry for many physical and chemical properties that have a combinatorial nature. In contrast to the geometric symmetry which preserves the shape of the object, in the case of unitary symmetry is saved ratio parameters (features, functions) for homologous pairs of combinatorial objects.

Recall that for baryons [4] can be written, for example, equation (1):

Do Not Include Page Numbers

No Not Include a Header

$$RM(dss) - RM(dus) = RM(uss) - RM(uus)$$

where RM is a Rest mass of baryons (MeV/c^2), u,d,s - кварки.

Similar laws are known, for example, for the mirror nuclei (Equation 2):

$$E_b [(K)P + (K-1) 9N] - E_b [(K-1)P + (K)N] = E_b [(K-1)P + (K-2)N] - E_b [(K-2)P + (K-1)N]$$

where E_b is binding energy of protons (P) and neutrons (N) in mirror nuclei, K is a number of nucleons.

So much for the ionization energies of atoms was obtained [8], equation (3):

$$[I_i (A_Z) - I_i(A_{Z+1})] - [I_i (A_{Z+1}) - I_i (A_{Z+2})] = const$$

where I_i - ionization energy of (i) electron of the atom A_Z with atomic number Z.

For homologous series of halogenated methanes were obtained [8] 90 equations corresponding to the replacement of one atom in one molecule to another atom to form another molecule. An example of such equations is shown Table 1:

Table 1. The system of equations for the replacement F-H. Before each chemical compound in order to save space omitted designations of some physical or chemical parameter of the molecule (A). For some parameters, for which the geometric symmetry does not play a big role, the equation with (*) and without () can be combined.

Replacement: F \longleftrightarrow H	
1	$CF_3Cl - CF_3Br = CH_3Cl - CH_3Br$
1*	$CHF_2Cl - CHF_2Br = CH_2FC1 - CH_2FBr$
2	$CF_3Br - CF_3I = CH_3Br - CH_3I$
2*	$CHF_2Br - CHF_2I = CH_2FBr - CH_2FI$
3	$CF_3Cl - CF_3I = CH_3Cl - CH_3I$
3*	$CHF_2Cl - CHF_2I = CH_2FC1 - CH_2FI$
4	$CF_2Cl_2 - CF_2Br_2 = CHFC1_2 - CHFBr_2 = CH_2Cl_2 - CH_2Br_2$
5	$CF_2Cl_2 - CF_2I_2 = CHFC1_2 - CHFI_2 = CH_2Cl_2 - CH_2I_2$
6	$CF_2Br_2 - CF_2I_2 = CHFBr_2 - CHFI_2 = CH_2Br_2 - CH_2I_2$
7	$CFC1_3 - CFBr_3 = CHCl_3 - CHBr_3$
8	$CFBr_3 - CFI_3 = CHBr_3 - CHI_3$
9	$CFI_3 - CFC1_3 = CHI_3 - CHCl_3$

Not all of these and similar equations are completely accurate, but the reasons for the low invariance are the subject of additional studies of violations of a particular symmetry.

It is important that these equations not only have great predictive power, but also the identification of false values previously measured parameters

It would be interesting to find similar pattern with more complex objects.

Do Not Include Page Numbers

No Not Include a Header

If we consider the triple nitrogenous bases in RNA - adenine (A), guanine (G), cytosine (C), uracil (U), then the combination of 3 with repetitions allowed, as is well known set of 64 codons that are completely "cover" all the vital for the synthesis of protein of 20 amino acids. These codons can be represented as a system of homologous series - the weight grating diagram corresponding to irreducible representations of the unitary group. (see Fig.1).

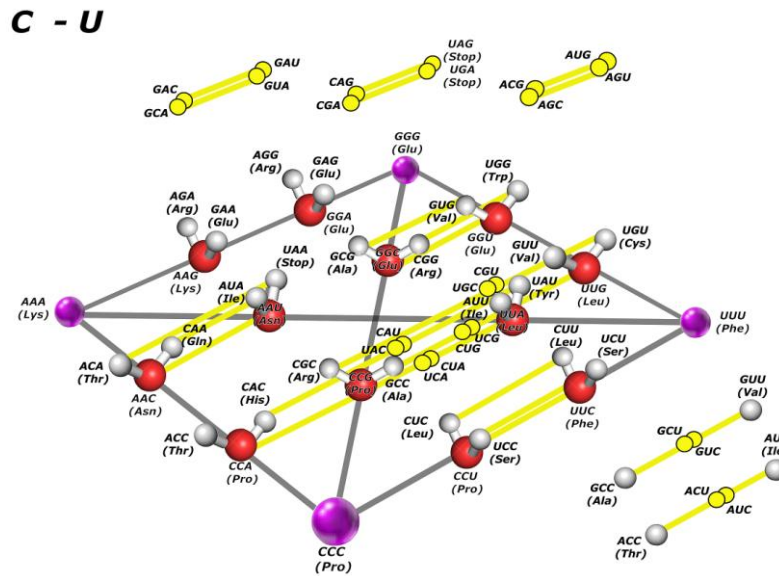


Fig.1. The formal structure of homologous series of codons. Codons, specify which amino acid (in brackets) will be added next during protein synthesis. Yellow lines show the "related" transitions from one codon to another codon while replacing (C - U).

It should be noted that this representation of the family of codons is more rational and promising scientific and educational sense than known tabular and diagramnye variants [10]. This view allows you to easily detect transitions from one codon to another, corresponding to the replacement of one nitrogenous base to another (see Fig 2).

**Homologous series of codons
when replacing cytosine (C) to uracil (U).**

C → U		A → G				
CCA	–	UCA	CCG	–	UCG	
		UCA			UCG – UUG	
		CUA			CUG – UUG	
ACC	–	ACU	GCC	–	GCU	
		ACU			GCU – GUU	
		AUC			GUC – GUU	
CAC	–	CAU	CGC	–	CGU	
		CAU			CGU – UGU	
		UAC			UGC – UGU	
ACA	–	AUA	GCG	–	GUG	
ACG	–	AUG	GCA	–	GUA	
AGC	–	AGU	GGC	–	GGU	
GAC	–	GAU				
AAC	–	AAU				
CAA	–	UAA				
CAG	–	UAG	}	CGG	–	UGG
CGA	–	UGA				
UCC	–	UCU				
CUC	–	CUU				
CCC	–	CCU				
UUC	–	UUU				

Fig.2. Homologous series of codons. Highlighted in red Stop codons.

By analogy with simpler molecules (combinatorial objects) for which the transition from one object to another, we observed invariance relations of a certain parameter (see. equation 1.2, 3 and Table 1), it would be logical to assume the existence of a certain invariance and in the case of biological combinatorial objects - codons. In this case the system of "equations", formally (from the standpoint of combinatorics) coincides with the system of real equations in Table.1 can be (in the case of replacement of C-U) is as follows (see. Table 2):

Table 2. All possible - equivalent and not equivalent - the replacement of one codon to another in a gene. In order to save space in front of each codon omitted a symbol of A, which characterizing the physical, chemical or biological parameter of both the codon and amino acid encoded by them. Sign (-), (=), (~) and the approximate equality are yet conditional and require strict biochemical and genetic interpretation.

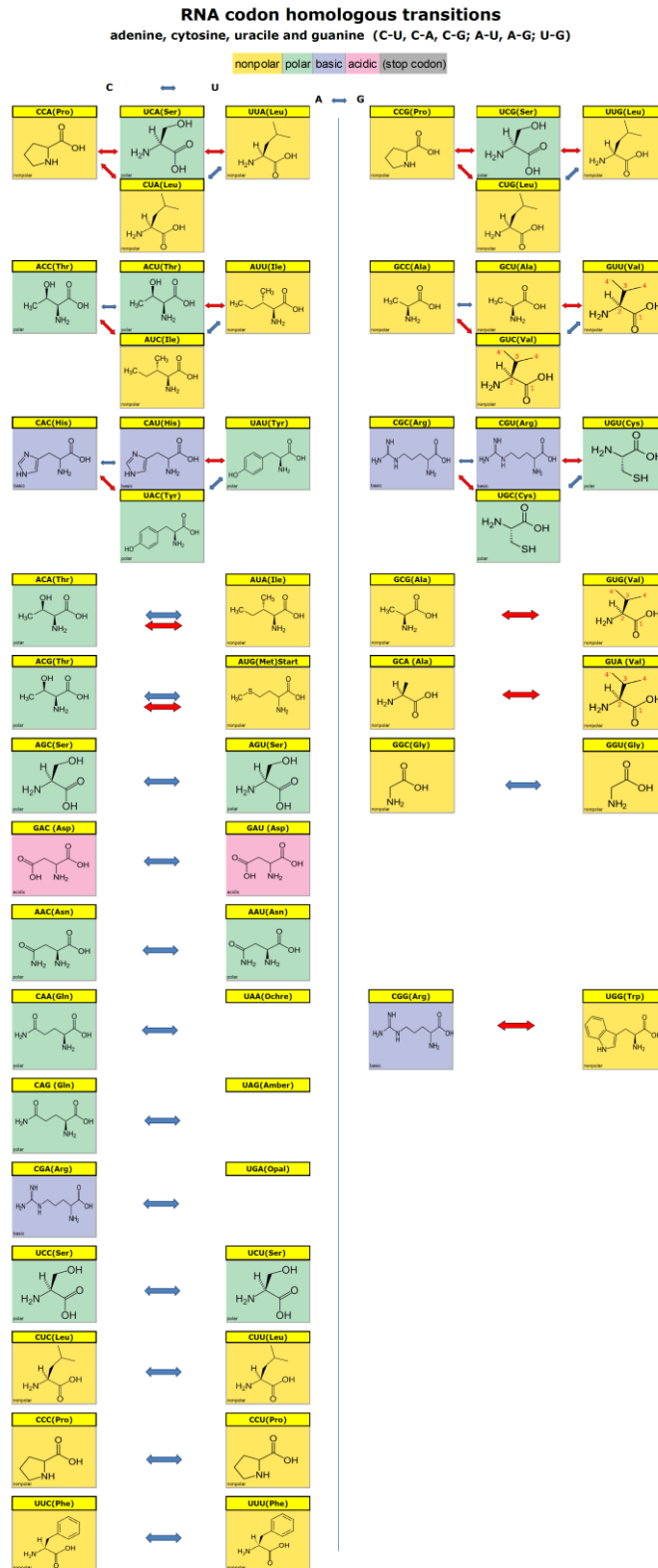
No Not Include a Header

	C		U		encoded amino acids						
1	CCA	-	UCA	=	CCG	-	UCG	Pro	Ser		
2			UCA	-	UUA	=	UCG	-	UUG	Ser	Leu
3			CCA	-	CUA	=	CCG	-	CUG	Pro	Leu
4			CUA	-	UUA	=	CUG	-	UUG	Leu	Leu
5	ACC	-	ACU	≈	GCC	-	GCU			Thr/Ala	Thr/Ala
6			ACU	-	AUU	≈	GCU	-	GUU	Thr/Ala	Ile/Val
7			ACC	-	AUC	≈	GCC	-	GUC	Thr/Ile	Ala/Val
8			AUC	-	AUU	≈	GUC	-	GUU	Ile/Val	Ile/Val
9	CAC	-	CAU	≈	CGC	-	CGU			His/Arg	His/Arg
10			CAU	-	UAU	≈	CGU	-	UGU	His/Arg	Tyr/Cys
11			CAC	-	UAC	≈	CGC	-	UGC	His/Tyr	Arg/Cys
12			UAC	-	UAU	≈	UGC	-	UGU	Tyr/Cys	Tyr/Cys
13	ACA	-	AUA	≈	GCG	-	GUG			Tyr/Ala	Ile/Val
14	ACG	-	AUG	≈	GCA	-	GUA			Thr/Ala	Met/Val
15	AGC	-	AGU	≈	GGC	-	GGU			Ser/Glu	Ser/Glu
16	GAC	-	GAU	=	const					Asp	Asp
17	AAC	-	AAU	=	const					Asn	Asn
18	CAA	-	UAA	}							
19	CAG	-	UAG				CGG	-	UGG		
20	CGA	-	UGA								
21	UCC	-	UCU	=	const					Ser	Ser
22	CUC	-	CUU	=	const					Leu	Leu
23	CCC	-	CCU	=	const					Pro	Pro
24	UUC	-	UUU	=	const					Phe	Phe

Indeed, for a number of transitions observed preservation of the acid-alkaline nature (see. Table 3).

Table 3. All possible codons homology when replacing the C - U. The red arrows indicate the substitution of one codon to another, in which the synthesized amino acid changes. The blue arrows indicate the substitution of one codon to another, in which the synthesized acid remains unchanged.

No Not Include a Header



In some cases, replacement of a single base in the codon to another, leading to the formation of the homologous codon does not change (is invariant with respect to such replacement) the main function of codon - amino acid coding.

For instance, the codons AGC and AGU, encode serine, replacing one to another will not result in a change of primary structure of the protein. However, may affect (for all whether genomes and on all living beings?) On the amount of protein in the cell, as the amount of transfer RNA that

Do Not Include Page Numbers

No Not Include a Header

serve these codons may differ very greatly in the cell, which will slow down or accelerate the fusion protein and, consequently, a change amount thereof .

If on-site codon CCA, including in protein encoding amino acids Proline, appears codon UCA, then the protein will not Proline and Serine. And in the case of on-site codon CCA another codon - CUA, then the protein will be an amino acid Leucine.

How will this affect the structure of the synthesized protein and its subsequent functions, development, growth and functioning of the organism as a whole?

It is known that in this case, the answers can be varied from a complete loss of function of the protein to a full save with all intermediate states (for all whether genomes and on all living creatures?). Most often a lack of effect or a slight change in the function, the so-called gene polymorphism.

Similar phenomenon was also observed in the case of halogenated methanes. With respect to the properties of one (thermodynamic parameters) of 1, 1 *, 2 * and 2, 3 and 3 in Table. 1 can be combined in pairs. This means that the replacement of one element to the other changes the spatial symmetry of the molecules, but does not affect the ratio of, for example, enthalpies.

Anyway, combinatorial representation of codons confirms the already known regularities in the changes of both the genes and their function.

But, more importantly, point to the discovery of new laws in the functioning of genomes and mutational processes.

Acknowledgment

The author expresses his gratitude to the head of the department of Molecular and Radiation Biophysics of The Petersburg Nuclear Physics Institute Dr. V.G. Kopolev

References

- [1] M. Gell-Mann, A Schematic of Baryons and Mesons. *Physics Letters* **8** (3): 214–215. 1964
- [2] Komarov V.S, Lyakhovskii V.D. "Unitary symmetry of molecules". *Journal of Chemical Physics (USSR)*, v. 5 & 7, pp. 914-924. 1986.
- [3] Komarov V.S., Evseyeva I.U. Proceedings of the Academy of Sciences, (USSR), v. 279, № 4, (1984), p. 928
- [4] Komarov V.S., Evseyeva I.U. Journal of Physical Chemistry, (USSR), v. 59, № 5, (1985), p.1246
- [5] Komarov V.S., Evseyeva I.U., Stoyanovskii M.O. Journal of Physical Chemistry (USSR), v. 59, № 7, (1985), p. 1829
- [6] Komarov V.S. , Stoyanovskii M.O. Proceedings of the Academy of Sciences.(USSR), v. 285, № 4, (1985), p. 933.
- [7] Komarov V.S. Higher Standards for Chemistry Education in XXI century. [viXra:1409.0085](https://arxiv.org/abs/1409.0085) submitted on 2014-09-12 02:42:51,
- [8] Komarov V.S. Proceedings of European conference on research in chemistry education 2014, "New trends in research-based chemistry education", Jyväskylä
- [9] Komarov V.S. Higher Standards for Science Education in XXI century. Proceedings of International Conference on 21st Century Education 2014, Dubai, U.A.E
- [10] https://en.wikipedia.org/wiki/Genetic_code.