Investigation Of The Relationship Between The Hydrophobicity Of An Amino Acid And Codon, Which Shall Encodes

IVAN TRENCHEV
Department of Electrical Engineering, Electronics and Automatics
University Center for Advanced Bioinformatics Research
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA

METODI TRAYKOV
Department of Informatics
University Center for Advanced Bioinformatics Research
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA

RADOSLAV MAVREVSKI
Department of Informatics
University Center for Advanced Bioinformatics Research
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA
radoslav_sm@abv.bg https://ais.swu.bg/profile/mavrevski

VIKTOR POPCHEV
South-West University "Neofit Rilski"
66 Ivan Mihaylov Str., 2700 Blagoevgrad
BULGARIA

Abstract: - In this article we did a brief review of mathematical models for studying the genetic code. We investigate how the hydrophilicity of amino acids effect of the structure. Presented a model, in this article, describe nucleotide sequences, about different levels of evolution of the proteins. Mathematical analysis of the construction and possible evolutionary scenario is presented in discussion. The code evolution is based on formal schemes whose relevance simple mutation and cardinality of synonymous set which codes one amino acid.

Key-Words: - Contemporary genetic code, optimality, simple mutation.

1 Introduction

Modelling of The Contemporary Genetic Code (CGC) is a subject of different methodical approaches in different areas of knowledge. The results, however, are related with the evolution of organisms in the nature. CGC is universal for all organisms [44], although some differences are discovered about some bacteria and mitochondria structures. These differences are in some codons, which encode another amino acid, but in general, it can be said, that CGC is universal [11, 12, 19, 21].

In the papers authors try to explain the relation between codon and amino acid (AA), as based on evolution or on the method of survival of the fittest [17, 18, 23-26]. Using computer simulations about the number of codons in one synonymous set and scoring of the number of possible mutations of all codons of the set, they prove that the number of codons changes in direct relation with the type of the amino acids (AA). The model, presented in this article, describe nucleotide sequences, about different levels of evolution of the proteins. The

following patterns of the model are present: fixing the possibilities about appearing of new protein in every generation about the whole target population; it is assumed that every position is independent by the other populations; the negative selection is only an object of modelling. This allows the authors to establish direct relation between codons and the type of the AA, and exactly the number of codons in the synonymous set is in direct relation with chemical properties of AA. In this research, done by Chechetkin and Lobzin [8, 9], an attempt is made, a relationship between the AA and number of codons, encoding one AA, to be interpreted by the sceptre of evolution, although they do not succeed to explain, by details, why, by example, a synonymous set with five codons does not exist, or why the number of codons in the synonymous sets in CGC is not more than 6 [7, 8]. Their model is very well structured and describes the changes of CGC by the sceptre of changes from lower to higher organisms.

Another interesting aspect of the study of properties of CGC is the relationship between the number of AA and the evolution of CGC. Ronneberg et al., [25] considered an interesting model of the evolution of CGC in which the code is complicated by the increasing number of AA i.e. appear a new AA and Nature creates a mechanism for it to be encoded [25].

They present interesting hypotheses about the change at codon. For example, the possible evolution of the codons GCA, this encodes AA - Alanine (Ala), in codons CAG, encoding Glycine (Gln) and others. The implemented model in the study did not consider some physic-chemical properties of AA, as they assumed that AA may be aromatic "replaced" with simply build AA and do not address the possible point mutations in different positions of codons. The computer simulations warrant the authors to express the opinion that one of the trends in the evolution of the code is that such nucleotides in the first two positions of codons encode the same AA. They find that the results they have made are theoretical and difficult to be tested in practice.

The idea, that CGC has been a subject of evolution, has been attractively presented by Vogel [40]. He believes that the early codes covered only a few AA and the number of codons has grown up with the evolution of the code. Many different hypotheses have been put forward to explain the possible evolution of the code to CGC and some of the current properties [10, 43-46]. They raise questions about the "optimality" of CGC, how the code is optimal in terms of minimization of errors in decoding or the number of synonymous codons in sets. How CGC was imposed because of its unique properties in respect of these criteria or the ability of evolution left in it. Other authors interpret the emergence of CGC with a scenario related to reducing the number of errors in simple mutations (TP, TPC) [35]. To verify this hypothesis, many CGC researchers calculate the number of possible point mutations for CGC and theoretical genetic code (TGK) and make comparisons between them [37-39] (. In most studies it is found that CGC is optimal and finding TGK with better properties than it's less possible.

It is experimentally demonstrated that point mutations occur - and often at first and third position of codons than at codons second position [15, 16, 40]. This is an interesting experimental fact which is used in the study of "optimality" of CGC. Many authors describe the properties of CGC and make interesting hypotheses. For example, in the overview article [19]b, describes the properties of CGC in the following manner: (i) such codons, which have two identical nucleotides encode the same AA, and (ii) codons, which have nucleotide U in the second position code for hydrophobic AA and those with A in this position hydrophilic. The authors do not prove these assumptions, but simply find them and comment, it is necessary a thorough examination to be made.

In some studies, the authors calculated the cost function (criterion) and then compare the values of the function of TGK and CGC. For example, in articles [4, 13, 14] authors maintained the thesis that GC has evolved to minimize errors in transmitting the message (translation) and errors of simple mutations transversion and transition (TP, TPC). Results achieved in these studies do not support the original hypotheses about the influence of the chemical properties of AA on the structure of CGC. Other authors also argue that the number of codons in the synonym sets in CGC is not accidental [3. 41]. They examined the relationship between AA and the type of codons and biosynthetic connection, i.e. establish a link between the chemical structure of GC and codons, who puts it encodes. The authors of the chemical models explain in detail the effects of certain enzymes in the synthesis of proteins. The results obtained in these two papers confirm the relationship between AA and the type of codons, but they do not comment on the reason for this connection, what has been generated.

Other authors working on similar topics try to find evolutionary aspects of the relationship between codons and endogenous structures similar to an AA [39]. Many authors try to analyse these structures in terms of optimality CGC [2, 22], but in

most cases the results are unsatisfactory. That AA with similar structures and biochemical properties are coded by similar codons, explains the need for a depth study [9]. Other assumptions, made in some developments, are that likely minimize the error in simple mutations or errors in reading the code is a major trend in the development of the code [1, 10].

Evolutionary aspects in the sustainability of CGC are examined through analysis on the degeneracy of the code, for example, in research [34] it is presented a model for the structure of CGC as a link between two sets (one of 64 elements and one of 20 elements). They explore the degeneracy of the code, and the possible mutations in the third position. Thus, they concluded that a third item is particularly important for the code, while the first and second position does not differ significantly in importance. An interesting result is obtained from Crick [6], which examined the number and properties of codons, in terms of the question, why these codons encode the same AA and the importance of the positions in codons. While the relationship between AA and the codons is not fully explained, the study gives a fairly comprehensive assessment of possible causes of this link.

Another method for testing the synonym sets in CGC which is applied in biochemistry and molecular biology is to construct an acceptable algorithm for the origin of CGC and explain the relationship of AA and the synonym set of codons, making it to encode [37] i.e. to examine the mechanism of protein synthesis in different organisms and based on the peculiarities of these mechanisms draw conclusions about the origin of CGC. An interesting approach used by some authors is to try to find the reason for the similarity between codons in a synonym set. In this respect they offer a scenario for early coevolution in ribonucleotide world i.e. the evolution of AA and nucleotides is integral [9]. These developments, however, neglected the fact that the number of tRNK is greater than the number of AA.

In literature, some classic models that attempt to explain the "philosophy" of the evolutionary process are known. This example is a lethal model [33], on which he worked and Crick (1968) [6]. Under this model CGC have evolved through natural selection idea i.e. natural evolution - mutation appears in the code and this automatically leads to a change of the protein, but the body does not always accept the new protein and CGC changes to reduce errors in decoding hereditary information. The thesis of the deadly model: organisms with mutated nucleotides are in most cases non-viable. Woese (1967) [44] constructed a similar model called "mobile" because the translation said, explaining that a major factor in the development of CGC is the reduction of the transport error [43].

Another popular model in literature is called "hypothesis Wobble "(Wobble hypothesis) (Crick, 1966) [5] under which, degeneracy of the code is the result of simplifying the relationship codon / anti-codon i.e. the number of tRNA to be minimized in the course of evolution which recognize an AA. According to this model type so tRNA depends on the first two nucleotides in codons. All of these models in different countries explain the evolution of the code, but the truth is probably somewhere in between [5].

Interesting study is made about microorganisms. Short life allows to study mutations in various research aspects. There is a plenty of experimental data which shows that mutations in bacteria are "useful" from a biological standpoint, because they allow genes to be expressed that trigger the synthesis of proteins which are useful for the conditions in which they are now. Thus, mutations are targeted to those genes which lead to changes in genotype and are current at the time of the study. These experiments demonstrate that simple organisms can design mutations depending on the pressure of the environment in which they are located, and it helps their survival. An interesting result, according to which, organisms adapt, depending on the environment. (This claim is contested in some scientific circles.) [40].

It should also be noted that in Economics more datasets now contain measures of genetic variation, therefore incorporating genomic data in economic analyses will become more common [31, 32, 41].

2 Problem Formulation

Let A, C, G, U and L are points to be the set of all three-letter words (codons) that encode twenty sites in the multiple united a_i and $a = \{a_1, a_2. ..., a_{20}\}$. Let $X = (X_1, X_2, ..., X_{20})$ is breaking many L \ M, where M is an arbitrary number of three codons, such that:

(1) $X_i \cap X_j = \varnothing$ за $i \neq j$;

(2) $\bigcup_{i=1}^{20} X_i = L \setminus M$;

(3) $| X_i | \neq \varnothing$.

Anyone correspondence C(X): X→a, between X and define a code K (X) (called theoretical genetic codel TGK) satisfying properties of CGC. This means that each a_i, i = 1, .., 20 is encoded by exactly one set of codons X_j, j = 1, .., 20, and these sets do not intersect (condition (1)). Moreover, these sets will call them synonymous, i.e. each of these codons encodes the same a_i. With K(L) denotes the set of all TGK K (X).

Let F(K(X), a_1, a_2, a_3, p_1, p_2, ..., p_{20}, b_1, b_2,) is a function of K (X) which accept the criteria for "optimality" of the code K (X). Depending on the parameters of the function, we will get different optimization problems as we examine the sustainability of TGK to mutations or simple relationship between the number of codons in the synonymous set and the probability of occurrence of AA in the average protein.

Denotes the probability of occurrence of a simple mutation in the 1st, 2nd and 3rd position of codons is a_1, a_2, a_3. Ivanov et all., (1989) [15] gave an assessment of them: 0.27, 0.19 and 0.54 respectively. Since AA constituting natural proteins are 20 in number, the probability of their occurrence denotes the average with p_1, p_2, ..., p_{20}. Since there are two types of simple point mutations - TP, TPC, with β_1 and β_2 to denote the probabilities of their occurrence, no matter what position. The ratio between β_1 and β_2 is: $\beta_1 = 2.\beta_2$

Our main goal is to solve and analyze the decisions of the following general discrete optimization problem:

F (K (X), a_1, a_2, a_3, p_1, p_2, … p_{20}, β_1,β_2) → extremum.

if K (X) ∈ K (L)

i.e. seek such TGK which is a minimum or maximum of the criterion of optimality F. Since CGC(L) obtained by comparing its TGK will give a clearer idea of its properties on its resistance to simple mutations.

To investigate the relationship between the type of codons and the affinity of CA to water molecules, i.e. properties of hydrophilic / hydrophobicity of CA and how much influence TGK, we consider the function:

$$F(x_{ij}) = \sum_{i=1}^{20}\left(\sum_{m=1}^{64} x_{im}\right)^{p_i}\left(\sum_{j=1}^{64}\sum_{k=1}^{64} P\left(x_{ij}, x'_{ik}\right)\left(h(a(x_{ij})) - h(a'(x'_{ik}))\right)^2\right) \to \max$$

as:

$$x = \left\{ x_{ij}, i = 1,..23; j = 1,..,64 \,|\, \sum_{i=1}^{23}\sum_{j=1}^{64} x_{ij} = 64,\ \sum_{j=1}^{64} x_{ij} \geq 1\ i = 1,..,20, \sum_{j=1}^{64} x_{ij} = 1,\quad i = 21,..,23, \right\}$$

In its use as the weight difference between the *h* hydrophobicity of AA and encoded by codons of TGK and hydrophobicity of AA and encoded by codons of CGC. Scale hydrophilic / hydrophobic of AA is from Epstein, (1966) [10]; Kyte et all. (1982) [20], $P(x_{ij}, x'_{ij})$ - probability of occurrence of simple mutations in codons by comparing codons that differ in one position, p_i - probabilities for the occurrence of AA in the average white: for bacteria and eukaryotes.

Received codes are presented in Table 1. and Table 2. TGK obtained decision function with parameter hydrophilic /hydrophobic of AA obtained after 10^7 Monte Carlo simulations.

Table 1. TGK obtained decision function with parameter hydrophilic / hydrophobic of AA obtained after 10^7 simulations in Monte Carlo for bacteria.

First position	Second position				Third position
	U	C	A	G	
U	Ser	Phe	Cys	Arg	U
	Ser	Phe	Cys	Tyr	C
	Ser	Leu	Stp	Stp	A
	Ser	Lys	Stp	Cys	G
C	Arg	His	Leu	Leu	U
	Arg	Leu	Leu	Leu	C
	Gln	Leu	Pro	Leu	A
	Arg	Leu	Leu	Leu	G
A	Thr	Ile	Ile	Arg	U
	Asn	Ile	Ile	Arg	C
	Arg	Trp	Ile	Ile	A
	Ile	Arg	Met	Ser	G
G	Gly	Asp	Val	Val	U
	Asp	Asp	Val	Val	C
	Glu	Glu	Ala	Arg	A
	Glu	Glu	Val	Arg	G

Table 2. TGK obtained decision function with parameter hydrophilic / hydrophobic of AA obtained after 10^7 simulations in Monte Carlo for eukaryotes.

First position	Second position				Third position
	U	C	A	G	
U	Trp	Asn	Phe	Thr	U
	Glu	Ile	His	Arg	C
	Ala	His	Stp	Stp	A
	Ser	Arg	Stp	Arg	G
C	Lys	Met	Asp	Cys	U
	Met	Ser	Tyr	Gln	C
	Cys	Leu	Gly	Leu	A
	Arg	Thr	Ala	Ser	G
A	Ala	Val	Arg	Leu	U
	Trp	Glu	His	Pro	C
	Tyr	Arg	Ile	Asn	A
	Arg	Arg	Arg	Ser	G
G	Ala	Trp	Pro	Met	U
	Asp	Ala	Tyr	Asn	C
	Glu	Ala	Arg	Glu	A
	Asp	Arg	Val	Ser	G

3 Problem Solution

In so far analyzed three cases, we tested CGC only in terms of simple mutations and probabilities of occurrence of AA in the average protein, without comparing with TGK and CGC. A similar approach was used in articles [14, 20]. This option has two advantages: simplicity of function and concentration on exploring the properties of CGC. A major shortcoming in this case is impossible to compare a TGK and CGC within one criterion. Therefore, we modify our criterion and that of the Gilis et all., (2001) [13], by adding thereto, and the probabilities of occurrence of AA in the average protein. Another modification is a change in the probability matrix - P (a, a'). In contrast to the Gilis (2001), however [13], we used the fixed probability of occurrence of simple mutations in codons by comparing codons that differ in one position, the TGK and one of CGC. When using the modified feature in the study of optimal properties of CGC against simple mutations and some properties of AA, we examined the properties of CGC with different probabilities for the occurrence of AA in the average protein of microorganisms and eukaryotes. We tried to find the optimal solution function of the whole set of TGK, i.e. polyhedron gene. Due to the size of the task - it has 1472 variables, we could not find the extremum of the function, which for now remains an open question. At the same time, however, we described a new criterion for testing the properties of CGC.

For a more detailed study of the properties of CGC, we have applied this criterion in two sub cases: a) with a weight matrix depending on the chemical properties of AA, and b) with a weight matrix, the elements of which are likely an AA to be replaced by another. From the results it can be assumed that CGC probably evolved to a point where decoding errors in simple mutations are minimal, but the code is embedded option to change. Our results confirm the hypothesis made by Vogel (1998) [44] that evolution in the organism world is the result of changes in the code, but the species have evolved in the direction of reducing errors in decoding the hereditary information [42].

Other side, in our study is to analyze the power of synonym sets in the light of some chemical properties of the AA as: molecular weight, potential energy and affinity to water (hydrophilic and hydrophobic). For this purpose, we added to our criterion and the number of synonym sets and probability of occurrence of AA in the average protein. The research, that is done for two kinds of probabilities for the occurrence of AA in the average white: for bacteria and eukaryotes, allows for a fuller examination of certain characteristics of CGC as the number of codons in the synonym set type of codons for an AA, etc.

With a fixed probability for simple mutation matrix P (a, a '), we "eliminate" a degenerate TGK. TGK example with the following capacities synonym sets: one subset with 42 codons, while the other 19 synonym sets by 1 codon. These fixed probabilities simplify the study of the function and received her results are close to the CGC. In developing Gilis et all., (2001) probabilities of simple mutations in different positions of codons are not fixed and are calculated directly dependent on the number of possible mutations in simple synonym sets. This is an advantage, because the value of a simple mutation probability is calculated and depend on the power of synonym sets, however, these probabilities are fixed in nature, as are evidenced by other authors [15, 16].

The results obtained with the modified function are not sufficiently informative to analyze the problems associated with probabilities of occurrence of AA in the average protein and the mechanism of evolution of organisms. This shows that the question of the level of complexity for the Study of CGC, and the factors that will participate in modeling stays open.

Some biological studies consider a hypothetical possibility, the probability of occurrence of AA in the average protein by mutations is not only point to has changed in the course of evolution [17, 27-30, 41]. The results of our studies provide mathematical proofs for this kind of experiments. Based on a simple function similar to that in our study Gilis et all., (2001) [13] commented that the probabilities of occurrence of AA in the average protein acts as an "evolutionary pressure" as a factor which is a function for conversion of organic compounds and any code. In this regard, he formulated two contrasting versions of the evolution of the code:

CGC is a function of probability of occurrence of AA in the average protein and transformed into their "evolutionary pressure" and- probability of occurrence of AA in the average protein is a function of the transformation of genetic code. Therefore, we can assume that CGC and probabilities of occurrence of AA in the average protein are closely related, which enhances the relationship codon / AA. Confirmation of these findings are mathematically proven in the development of Perlwitz et all. [24]. Therefore, inclusion of probabilities of occurrence of an AA in the average protein, in simplified and complicated in function, through which examine the code is an essential step in this study. It allows for a more detailed study of the properties of the code and search for parallels with the purely biological experiments.

The literature also pointed to "other" parameters that play the role of evolutionary pressure for CGC [22]. Broadly covering all mechanisms that encode genetic information and support. For example, CGC is obviously related to translational apparatus consisting of ribosomes and tRNA, whose action we have described here schematically by the probabilities p (c `/ c) (Pelc, 1965; King, Jukes, 1969) [18, 23] and weight matrix dependent on some chemical properties of AA. All these mechanisms may have evolved in parallel with the evolution of CGC during the early stages of formation of life [23, 27]. Our results give reason to conclude that the relationship between AA and the codon who encodes it, is specific, i.e. that the first two letters of triplets determine the type of tRNK. In a similar theme work [1, 25], but they do not explain this connection in CGC. The relationship is not absolute, because the number of all the variations for the first two positions of codons is 24 i.e. 16 and the number of tRNK is greater than 20 in many organisms. These differences are related to the third position in codons, which is not clear if related to the type of tRNK.

Another important property that we stake in modeling is the affinity of the AA molecule to water. For this purpose, we used in the criterion values F of hydrophilic / hydrophobic of AA, coded by TGK and CGC. Our results show a possible link between the type of AA and the second position of codons, namely AA encoded by adenine nucleotides in the second position of codons are hydrophobic, hydrophilic by uracil. Other authors have worked with this property of AA [20, 39] show the same but with other methods. Hypothesis that AA with similar properties are coded by similar codons of first and second position is intriguing but needs to be investigated more thoroughly.

4 Conclusion

The notion of optimality CGC may have different contexts in biology, biochemistry and molecular mechanics, mathematics and bioinformatics. As it is seen from the perspective of evolutionary development of species optimum is the code that is not absolutely resistant mutations. From a mathematical point of view, the minimum number of mutations, i.e. the error means minimize the effects of errors, which is an optimization problem. Optimality criteria for further complicate the problem because they are different, such as: length of the codons, the number of letters (nucleotides), number of codons in the synonym set, the position of mutation in codons and others and because their combination is a part of modeling.

Acknowledgements

This work is partially supported by the project of the Bulgarian National Science Fund, entitled: Bioinformatics research: protein folding, docking and prediction of biological activity, NSF I02 /16 12.12.14.

References:

[1] Alff-Steinberger, C. The genetic code and error transmission. Proc. Natl. Acad. Sci.U. S. A. 64, 1969, 584–591.

[2] Ahmad, M. Jung, T. and Bhuiyan A. From DNA to protein: Why genetic code context of nucleotides for DNA signal processing?

Biomedical Signal Processing and Control, 34, 2017, 44-63.
[3] Amirnovin, R. An analisys of the metabolic theory of the origin of the genetic code. J Mol Evol 44, 1997, 473-476.
[4] Ardell, D. H. On error minimization in a sequential origin of the genetic code. J. Mol. Evol. 47, 1998, 1–13.
[5] Crick, F. H. Codon--anticodon pairing: the wobble hypothesis. J. Mol. Biol. 19, 1966, 548–555.
[6] Crick, F. H. The origin of the genetic code. J. Mol. Biol. 38, 1968, 367–379.
[7] Chechetkin, R. and Lobzin V. Stability of the genetic code and optimal parameters of amino acids. Journal of Theoretical Biology, 269, 1, 2011, 57-63.
[8] Chechetkin, R. and Lobzin, V. Local stability and evolution of the genetic code. Journal of Theoretical Biology, 261, 4, 21, 2009, 643-653.
[9] Di Giulio, M. On the RNA world: Evidence in favor of an early ribonucleopeptide world. J. Mol. Evol. 45, 1997, 571–578
[10] Epstein Role of the amino-acid'code and selection for conformation in the evolution of proteins., Nature 210, 1966, 25-28
[11] Freeland, S. J. Knight, R.D. and Lanweber, F. Measuring adaptation within the genetic code. Tibs 25, 2000, 44-45.
[12] Freeland, S. J. and Hurst, L. D. The genetic code is one in a million. J. Mol. Evol. 47, 1998, 238–248.
[13] Gilis, D. Massar, S. Cerf, N. and Romman, M., Optimality of the genetic code with respect to protein stability and amino-acid frequencies. Genome Biology, 2(11), 2001, 1-12.
[14] Haig, D. and Hurst L. D. A quantitative measure of error minimization in the genetic code. J. Mol. Evol. 33, 1991, 412–417
[15] Ivanov, O. Ch. Genetic code and point mutations. Studia Biophysica, 129, 1989, 63-65.
[16] Ivanov, O. Ch., Milanov, P. and Kenderov, P. Genetic code optimality from mathematical and evolutionary point of view. Compt. Rend. Acad. Sci. Bulg., 40, 1987, 25-32.
[17] Keeling, P. J. Genomics: Evolution of the Genetic Code. Current Biology 26, 18, 2016, 851-853.
[18] King, J. L. and Jukes, T. H: 1969. Non-Darwinian evolution. Science, 164, 2016, 788-798.
[19] Knight, R. D. Freeland, S. and Landweber F. 1999. Selection, history and chemistry: the three faces of the genetic code. Trends Biochem. Sci. 2, 2016, 241–247.
[20] Kyte, J. and Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 12, 1982, 159-158.
[21] Kuruoglu, E. E. and Arndt P. F. The Information Capacity of the Genetic Code: is the Natural Code Optimal? Journal of Theoretical Biology 419, 2017, 227-237.
[22] Nitta, I. Kamada, Y. Noda, H. Ueda, T. and Watanabe, K. Reconstitution of peptide bond formation with Escherichia coli 23 Sribosomal RNA domains. Science 281, 1998, 666–669.
[23] Pelc, S. R. Correlation between coding-triplets and amino acids. Nature 207, 1965, 597-599.
[24] Perlwitz, M. Burks, and Waterman, S. Pattern analysis of the genetic code. Appl. Math. 9. 1988, 7–21.
[25] Ronneberg, T.A. Landweber, L.F. and Freeland S. J. Testing a biosynthetic theory of the genetic code: Fact or artifact? PNAS, 97, 25, 2000, 13690 -13695.
[26] Rodin, A. S. and Rodin, S. N. The Universal Genetic Code and Non-Canonical Variants, Brenner's Encyclopedia of Genetics (Second Edition), 2013, 263-264.
[27] Santos, J. and Monteagudo A. Study of the genetic code adaptability by means of a genetic algorithm. Journal of Theoretical Biology, 264, 2010, 854-865.
[28] Sawyerr, B. A. Adewumi, A. O. and Ali, M. M. Real-coded genetic algorithm with uniform random local search. Applied Mathematics and Computation, 228, 2014, 589-597.
[29] Sciarrino, A. and Sorba, P. Codon–anticodon interaction and the genetic code evolution. Biosystems, 111, 3, 2013, 175-180.
[30] Salemne, F. R. Miller, M. D. and Jordan, J. R. Structural convergence during protein evolution. Proc Natl Acad Sci USA, 74, 1977, 2820-2824.
[31] Stavrova, E. „Conventional and shadow banking sector – comparative aspects of the post-crisis period in time of the currency board Bulgaria 'case", International Scientific conference "Innovation in science and education" March 22-24, 2017 - Prague, Czech Republic, http://ojs.journals.cz/index.php/CBUIC/article/view/1114/pdf, 2017, 453-457
[32] Stolonal, D. Tax structure and economic growth: Evidence from the European Union, Journal Contadina y Administration, vol. 62(3), 2017, 1041–1057, https://doi.org/10.1016/j.cya.2017.04.006

[33] Sonneborn, T. M. Degeneracy of the genetic code: extent, nature and the genetic implications. In Evolving Genes and Proteins. Edited by Bryson, V, Vogel HJ. New York: Academic Press, 1965, 377-397.

[34] Suzuki, Y. and Gojobori, T. A method for detcting Positive Selection at Single Amino acid Sites. Mol. Biol. Evol. 16(10), 1999, 1315-1328.

[35] Szathmáry, E. The emergence, maintenance, and transitions of the earliest evolutionary units. Oxf. Surv. Evol. Biol. 6, 1989, 169–205.

[36] Szathmáry, E. and Smith, M. J. The major evolutionary transitions. Nature, 374, 1995, 227–232

[37] Szathmáry, E. and Maynard, S. J. From replicators to reproducers: the first major transitions leading to life. J. Theor. Biol. 187, 1997, 555–571.

[38] Tlusty, T. A colorful origin for the genetic code: Information theory, statistical mechanics and the emergence of molecular codes. Physics of Life Reviews, 7, 3, 2010, 362-376.

[39] Taylor, F., and Coates, D. The code within codons. Bio Systemsp 22, 1989, 177-187.

[40] Trencheva T., Denchev S. Intellectual Property Knowledge at the University's Information Environment: A Comparative Study Open Journal of Applied Sciences, 4, 2014, 130-136, http://file.scirp.org/Html/4-2310217_43621.htm

[41] Trendcheva, M. Innovative moments in credit risk analysis and management - Eleventh International Scientific Conference of Young Scientists "The Economy of Bulgaria and the European Union: Competitiveness and Innovation" at UNWE, Sofia, 15 December 2015, 521-528, ISBN: 978-954-8590-35-8.

[42] Vogel, G. Tracking the history of the genetic code. Science, 281, 1998, 329-331.

[43] Woese, C. The Genetic Code: The Molecular Basis for Genetic Expression. New York: Harper and Row, 1967, 57-63.

[44] Woese C. R., On the evolution of the genetic code. Proc Natl Acad Sci. USA, 54, 1965, 1546-1552.

[45] Wong, J. T. A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. U. S. A. 72, 1975, 1909–1912.

[46] Zhang, Z. and Yu J. Does the Genetic Code Have A Eukaryotic Origin? Genomics, Proteomics & Bioinformatics, 11, 1, 2013., 41-55.