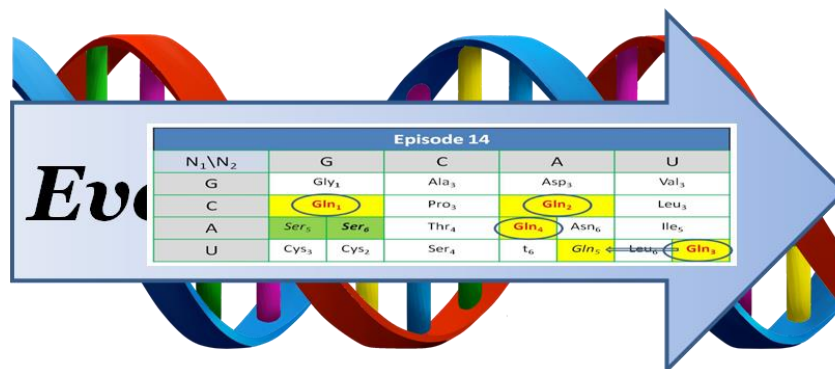# The Smooth Evolution of the Universal Genetic Code. Main Episodes

Piotr H. Pawłowski[1] ✎

[1]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland
Piotr H. Pawłowski

Abstract: The possible scenario of the origin and evolution of genetic code is proposed, being primarily implicated by the working hypothesis which states that the chronological order of amino acids evolutionary implementation monotonically correlates with their increasing mass. It fulfills the minimalistic claim of the smallest changes of the evolving system at increasing complexity, hereinafter called "the smooth evolution". The working hypothesis was postulated concerning the results of statistical analysis indicating a strong correlation between amino acid mass and the chosen parameters of contemporary genetic code, which are expected to change in a certain individual direction during the evolution of the initial genetic system. It was additionally supplemented by the most common hypotheses adopted from the literature, as stereochemical, 'frozen accident' and coevolutional. The developed scenario allows a detailed description of the twenty-two consecutive episodes of the history of code definition and the estimation of its dynamics. It reveals the main eras of evolution conditioned by the environmental and structural constraints. It also lets the estimation of the evolutionary frequencies of codon sense expansion, and redefinition. Dominating trends and amino acids were indicated. The underlying assumptions, limits, exceptions, and the future of the code evolution have been discussed.

Graphical Abstract

Keywords: Universal Genetic Code, Evolution, Theory

Introduction

The *genetic code* discovered in the mid-sixties of the last century (Nirenberg & Matthaei, 1961; Söll et al., 1965) are nearly universal relations between the sequence of bases in nucleic acid and the sequence of amino acids in a protein (Berg et al., 2002). It is responsible for the accurate transfer of information from the nucleotide triplets in DNA, through RNA, to synthesized protein, during the gene expression process, what guarantees the preservation of primary, and higher level, protein structure. The genetic code differing in the semantics from the contemporary version was probably originated in the initial phase of biological evolution, during the transition between RNA to DNA world. Its maturation could take place on the evolutionary pathway linking hypothetical minimal organism, MO, and the last common ancestor, LUCA, i.e. a minimal molecular assembly compatible with life (Forster & Church, 2006) and an entire minimum genome (Mat et. al., 2008).

Up to now, there is no one widely accepted theory of the origin and early evolution of genetic code. Since the geometric concept of triplet diamonds (Gamow, 1954), neglecting biochemical context, there have been many other hypotheses proposed to explain nature, and history, of genetic code origination. They usually focus on a single selected aspect, as: optimization improving conservation of kinetic

✎ **Piotr H. Pawłowski (Correspondence)**
☎ +

energy, transfer accuracy, and minimization of the harmful effect of genetic errors (Sonneborn, 1965; Woese, 1965; Itzkovitz and Alon, 2007; Guilloux and Jestin, 2007; Seligmann, 2011; Kumar and Saini, 2016), stereochemical affinity of codons and amino acids (Pelc & Welton, 1966), the randomness of abiotic allocation (Crick, 1968), coevolution *of the code and metabolism* (Wong, 1975), and the ancient proto-code (Koonin & Novozhilov, 2017). The hybrid combination of the above (Freeland, 2000; Chechetkin, 2006) was also proposed. In all cases, some evidence in support was seeking with the use of experimental tests, analysis of characteristic correlations or Monte Carlo simulations.

Many different single-factor criteria and multi-factor hypotheses about the chronological order of appearance of amino acids in the early evolution were reviewed and summarized in consensus ranking (Trifonov, 2000). Nine amino acids (Gly, Ala, Val, Asp, Glu, Pro, Ser, Leu, and Thr) of Miller's primordial environment are all ranked as topmost. Six of them are the lightest amino acids in the proteinogenic pool,

In the following paper, we propose to apply to some extent the former ideas together with the postulated work hypothesis of 'smooth evolution', declaring possible smallest, simplest and most effective changes in gradually erasing genetic code and its evolving semantics. The analysis of the over one thousand of evolutional scenarios allowed for the formulation of the set of the main essential rules of evolutionary priority and that way the anticipation of the scenario of the precedence and temporal evolution of discussed processes, divided into twenty-two chronological episodes. Each episode covers the

interval of time, including the existence of certain, parallel or competing, evolutionary processes, dependent on the previous history and determining the future of the system. This scenario, to the best of the author's knowledge, is currently the most elaborate concept. Despite this, it cannot be treated as the only one, but rather as the most possible attractor to which there were tending all the evolving early RNA - amino acid systems.

The obtained course of the evolution shows the main evolutionary phases, i.e., pre-code, proto-code pre-ribosomal (warm), proto-code pre-ribosomal (moderating), proto-code proto-ribosomal, and proto-code ribosomal eras, conditioned by the thermal and structural constraints. The estimation of the evolutionary frequencies of codon sense expansion, and amino acid redefinition allowed for the indication of dominating trends and the most important amino acids. The assumptions of the model, essential limits and possible exceptions from standard, as well as the future of the code evolution, have been finally discussed.

Methods
Empirical premises
If to analyze the present UGC regarding its possible initial evolution some general observations can be articulated. They are presented below.
Obs.1. The molecular weight of amino acid correlates with the strength of corresponding codon-anticodon hydrogen bonding.
Let's consider the eight RNA triplets containing only bases G and C, i.e., GGG, CGG, GCG... CCC, which are the most bounded triplets in mRNA - tRNA interactions (up to 9 H-bonds), coding four amino acids (Tab. 1).

Table 1. The most bonded and the weakest triplets of genetic code.

Codon	Amino acid	Mw [Da]	Mut. codon	Amino acid	Mw [DA]
GGG	Gly	75,1	AAA	Lys	146,2
CGG	Arg	174,2	UAA	Och	35911
GCG	Ala	89,1	AUA	Ile	131,2
GGC	Gly	75,1	AAU	Asn	132,1
CCG	Pro	115,1	UUA	Leu	131,2
CGC	Arg	174,2	UAU	Tyr	181,2
GCC	Ala	89,1	AUU	Ile	131,2
CCC	Pro	115,1	UUU	Phe	165,2

Let's also consider the corresponding triplets AAA, UAA, AUA... UUU as a result of the full set of possible point mutations, conserving the overall nucleotide structure (i.e., The number of carbon-nitrogen rings) but decreasing its bounding, namely, transitions G-A and C-U. They are the least bounded triplets (up to 6 H-bonds) coding six amino acids and one release factor. In each case the amino acids coded with the most strongly bounding triplets, i.e., Gly, Arg, Ala, Gly, Pro, Arg, Ala and Pro, have a smaller molecular mass than amino acids (or release

factor) coded by the corresponding weak triplets, i.e., Lys, Och, Ile, Asn, Leu, Tyr, Ile and Phe. The mentioned trend (c.c.= -0.996 for averages), weaker bounded codon - heavier object coded, is also clearly visible on the summary plot for all possible amino acid codons (Fig. 1). Shown linear regression may be a trace of hypothetical evolutionary course, leading from the strongly bounding codons and small amino acids coded towards the weak codons and heavy amino acids, at the increasing number of coded entities. The reverse course at decreasing the number

of coded amino acids was not included for further considerations. The association between codon-anticodon duplex stability and amino acid weight was earlier discussed by Seligmann (2010).

Obs.2. The molecular weight of amino acid correlates with its distance from the center of the metabolic system.

Let's consider amino acid metabolism pathways (Fig.2) and the distance of a given amino acid from the arbitrary chosen central point of the glycolysis route (Glycerate -3-Phosphate), probably one of the earliest metabolic pathways. Discussed distance, describing amino acid peripherality, was defined by a minimal number of catalytic processes leading from the assumed center to the final product.

Summary analysis of mass vs. distance (Fig.3)

shows an overall positive trend (c.c.= 0.936), especially seen when amino acids are grouped in main similarity families.

Despite wide dispersion, the visible regularity may be the relict of the evolution of amino acid metabolism and its incorporation into genetic machinery. Coevolution of genetic code and metabolic pathways was postulated by Wong (1975, 2005).

Obs.3. The molecular weight of amino acid correlates with the packing of genetic code.
Let's consider the packing of genetic code, which means how many different entities (amino acids or release factors) can be coded due to the substitution of the nucleotide at the third position of the coding triplet (Tab. 2.).

Table 2. The packing of genetic code.

N1\N2	G	C	A	U
G	Gly	Ala	Glu||Asp	Val
C	Arg	Pro	Gln||His	Leu
A	Arg||Ser	Thr	Lys||Asn	Met||Ile||Ile
U	Trp||Opa||Cys	Ser	Amb||Och||Tyr	Leu||Phe

The columns and lines are assigned to the first and second position of the codon. The color of the field means the number of possible different entities coded due to the substitution of the nucleotide at third position: blue- 1, green- 2, red- 3.

The average mass of packed amino acids was analyzed (Fig. 4). Slightly manifested trend (c.c.= 0.996), the tighter packing - the heaviest amino acids, may be observed, probably as a result of combining wobble and crowding effects during the initial spread of the evolution.

Obs.4. The molecular weight of amino acid correlates with the accessibility of the codons.

Let's consider genetic codon during mRNA-tRNA triplet stacking in the translation process. The order and size of bases in the codon may facilitate or reduce the accessibility of the codon (Fig.5). Concerning the wobble hypothesis, the first and the second position of the codon play a more important role than the third position. The full approach always requires base pairing at first and second position, with the direction of the anticodon approach from 3' mRNA end. Then codon base at first position can be sometimes shaded by the base at the second position. This effect was assumed to be proportional to the difference in the molecular mass of code nucleotides 2th and 1nd (Tab. 3). Thus the accessibility measure was defined oppositely as the difference m.w.1 - m.w.2.

Table 3. The measure of codon accessibility. The difference in molecular mass of nucleotides at 1th and 2nd position, m.w.1 - m.w.2.

N1/N2	G	C	A	U
G	0	40,03	16	39,04
C	-40,03	0	-24,03	-0,99
A	-16	24,03	0	23,04
U	-39,04	0,99	-23,04	0

Field colors indicate the level of codon accessibility during the translation process: white – neutral, gold – facilitated, violet – reduced.

The average molecular mass of coded amino acids vs. the measure of the accessibility of the codon was analyzed. (Fig. 6), showing the trend (c.c.= -0.857), more accessible codons - lighter amino acids, which may be the result of the hypothetical evolutionary

strategy, first attributing easily accessible codons to the light amino acids. Properties contrasting 1st vs 2nd codon positions were mentioned correlated with genetic code history by Demongeot and Seligmann (2019).

General remarks
During the early evolution of genetic code a decrease in hydrogen bonding of codons, an increase in

metabolic peripherallity of amino acids, an increase in packing of code, and a decrease in accessibility of codons may be expected. These varying properties, as it was shown above, in overall, move in tandem with an increased molecular mass of amino acids. Such observation suggests the possibility of a sequential increase in the mass of new coded amino acids during the evolution of universal code. Additionally, the rational claim of minimal structural changes allows us to specify this suggestion and precise postulation of sequential increase in the mass of new amino acids, starting from the lightest one. This assumption was adopted in further considerations as a working hypothesis.

The minimalistic model of the emergence and development of a UGC is proposed. It presents the shortest route obeying the simplest possible modifications of the evolving coding system, gradually increasing its complexity, i.e., from the first codon, and first coded amino acid, towards the current universal code with 64-codons, coding 20 proteinogenic bricks. It postulates that this route approximates a certain universal bio-physico-chemical attractor, common for the all proto-organisms in the initial phase of the evolution of life, the history of which, in general, can be divided into the 22 chronologically consecutive episodes, essential for the development of a code in the form we know. Logically arranged episodes obey simultaneous processes, starting approximately at the same time, and running in the same period at a comparable rate, interrelated or not, sometimes competing. They mainly include the first definition and further redefinition of amino acid for a given codon, or the code expansion for a given amino acid, which led the emergence both of new coded amino acids, and used codons, and specified the dynamic constraints for the evolutionary pathways in the following episodes. Generally, one type of amino acid is introduced in a given episode, or amino acids of the same mass.

The precedence rules

According to the model, the order of the emergence of new amino acids, and newly defined codons, in the subsequent episodes, as well as the type of processes taking place, were regulated by the certain evolutionary dominating precedence rules (Tab. 4).

Table 4. The main rules of the evolutionary dominating precedence.

Main Rules	
1. First assignation of a free codon meaning (if possible)	
1.1	-defining the least mass unassigned amino acid
2. Codon redefinition (if possible)	
2.1	-of the amino acid from the same metabolic family
2.2	-possessing the closest mass
2.3	-assigned to the less accessible codon (exception: the first codon is conserved)
3. Unloaded tRNA or code expansion	
3.1	-via transition
3.2	-via transversion
3.3	-to the most accessible codon
Supplementary Rules	
1.	At least one codon representing a given amino acid should stay not redefined
2.	No more than four redefinition incidents per episode
3.	An amino acid with the five-membered nitrogen ring (Pro, His) can redefine only one codon
4.	Amino acid-containing sulfur (Cys, Met) can redefine only one codon
5	The same mass of redefining amino acids in a given episode

Results

The evolution of universal genetic code

A brief screenplay of the evolution of the UGC according to the proposed model was presented below (for detailed description, see supplementary material in Appendix 1 and Appendix 2). The main episodes were visualized as the movements and conversions on the board N1\N2 of nucleotide combinations (Fig. 8). Episodes 1 and 2 (E.1, E2.) introduce hypothetical pre-code starting conditions. Following episodes (E.3 - E.22) show evolutional progress due to the simplest possible implementation of the precedence rules (Tab. 4).

The presented model places the initial phase of the evolution of the genetic code in the hot era of the RNA world, dominated by self-replicating molecules of RNA, possessing some catalytic properties. They were prevailingly built of weak bonding light nucleotides A, and U, inducing the collective ephemeral interactions between the approaching RNA strands, or within a single RNA molecule. It could lead to the formation of the dynamic RNA sol-gel inter-phase areas, being natural reactors where the first biochemical reactions could be catalyzed. The dominant primary structure

of the ancient RNA is rich in poly-A and poly-U regions but also contains some amount of bases G, and C. They can form the first self-replicating proto-tRNA (E.1), and aminoacyl-tRNA synthetase proto-ribozyme (E.2).

Aminoacyl load of the proto-tRNA required strong and well-defined bonding. Due to some cooling, at the end of the hot pre-code era, the doublets of the strong bonding nucleotides in sequence (up to 6 H-bonds), became capable to preserve the thermal tearing, and provide the local, stable, complementary interaction between RNA strands. It enabled the more precise organization, function, and proper replication of some early RNA structures, which could drive complex physicochemical pre-life processes. The first ribozyme based aminoacyl-tRNA synthetase (Nick et al., 2000), utilizing the most frequent strongly bonding doublet joints, GG-CC, recognizes CC targeted proto-tRNA and transfers the simplest and stereochemically affined amino acid, glycine, to the 3' end of the future transporter (Fig. 7). The possible opposite configuration (e.g., CC- codon, GG anti-codon) implies the less curvature of the pro-tRNA anticodon loop, which hinders the exact approximation, recognition, and pairing (Duan, 2018). Likely in this way, with the synthesis of the first gly-proto-tRNA, the right evolution of the genetic code began (E.3). Final assignation to the weaker bonding codon doublets GA and AG ends an initial hot era.

The further cooling initiates the proto-code pre-ribosomal warm era, when, besides the strongly bonding nucleotides, they are freely incorporating the weaker ones (E.4 - E.9). The following evolution of the first tRNA loading system with the two-positional code, obeying primarily mutation of codons and amino acid redefinition due to the close substitution, leads to the raising of functional precursors of the modern tRNA. In the beginning, they play structural, and catalytic, non-proteinobiotic role.

In the following proto-code pre-ribosomal moderating era (E.10 - E.11), the weakest doublets of A, and U, can be successfully incorporated into the code, too. It enables the further assignation of the aa-proto-tRNA complexes and the introduction of the blind termination codon UA in the synthesis of the small peptides due to the catalytic activity of the supramolecular ribozyme structures. The proteinobiotic role of the proto-tRNA begins to be manifested and the first translational system initiates.

Later, in the proto-code proto-ribosomal era (E.12 - E.16), due to the activity of the first proto-ribosomes the third position in the codons becomes significant, too. This allows the addition of the heavier amino acids to the existing coded collection and the

synthesis of the short polypeptides, some of them related to the maturating ribosomes. Initially, the role of the stop codon is naturally played by the sequences complementary to the anticodons belonging to the tRNAs not-charged with amino acids, which were the precursors of the protein release factors. The two protein-based release factors are activated in the following proto-code ribosomal maturation era (E.17- E.22) taking the three unused triplets. This closes the period of the rise and formation of the UGC and opens the phase of its global implementation in the synthesis of the first complex proteins.

Discussion
The action of life takes place in a narrow range of physicochemical parameters (Henderson, 1913). To meet them biological evolution at the beginning could not have an abrupt and discontinuous character. It seems, that the minimal change strategy proved to be very effective.

Present UGC is the set of rules determining the most persistent organization of protein structure - the order of amino acids. Certainly, it was not the case at the beginning of its formation when aa-proto-tRNA complexes do not play the transfer role, but mainly structural or catalytic one and the self -reference may be an important factor (Guimarães, 2013). The presented model shows a smooth passage from the first single coded amino acid, glycine, at that time were not building any protein, to today's entire system of codons specifying the basic components of life during the translation process. According to the proposed model, the proteinobiotic role of the codons raised in the middle of the evolutionary course. It was preceded by the creation of the certain pro-catalytic RNA structures, i.e. proto-ribozymes and proto-tRNA, synthesizing first aa-proto-tRNA complexes in earlier eras. Revealed pseudo-dynamics of the UGC formation is presented in Fig. 9.

It reflects the true dynamics only on the assumption that all episodes took the same period. At this constraint, high dynamics can be expected at the beginning of the code evolution (E.3 - E.6), and after initiation of the full ribosomal activity (E.17 - E.22). Redefinition of serine by proline (E. 6) initiates the phase of relatively low dynamics (E.7 - E.16), devoted to the creation of aa-proto-tRNA, small peptides, ribosomes, and short polypeptides.

Summarizing, the 2 specific definitions and 54 redefinitions of already assigned codon to a new amino acid, supplemented by 10 transitions, and 5 transversions, during the expansions of an assigned codon to a free one, were considered to create UGC. As a result, the most expansive amino acid turned out to be alanine, performing 4 expansions (Eps. 4

and 5). The isoleucine, was the most active in the codon sense redefinition, conquering 5 codons (Eps. 10 and 11).

The most important limitations of the model are related to the environmental, strictly thermal, conditions influencing thermostability (Gotoh, 1983; Xia et al., 1998). It describes hyperthermophilic genetic code origins based on the strong bonding duplexes G, and C, but it may not fully describe the evolution of the code of some exotic systems which were earlier adopted to the extreme thermal conditions or enter locally cold era, what allowed them to use the weaker bonding duplexes containing A, and U, and to follow the nonstandard evolutionary pathway. Although, in general, the proposed model assumes the cooling trend during the evolution of code meanings, it allows the hyperthermophilic maturation of genome compositions (Di Giulio, 2003). Also, environmentally driven, some specific differences in the maturation of ribosomes may be the source of the exceptions from universal standard observed today in some coding triplets, in the third position containing purines.

Some questions may arise regarding assumed precedence rules (Tab. 4). The priority of codon sense redefinition before its expansion, underline greater importance of the quality than the quantity at the beginning of the evolution of life. The rules 1.1 and 2.1-3 meet the empirical premises discussed in the earlier chapter. It is necessary to stress, that less accessible codons are favored in the redefinition as the target for new amino acids because its small use minimizes the possible changes in the evolving system. The exception obeys the first gly codons (GGN), conserved for the stability of the newly emerged system.

The accessibility of the bi-codons was estimated according to Table 3. In the case of comparison of codons with different bi-codons and partially distinguishable third position (i.e. Purines||Pyrimidines), the negative accessibility of a bi-codon was multiplicated by a factor of 2, and positive accessibility divided by a factor of two. In the analogous case, but for the codons with the same bi-codon it was assumed that the third position of a codon in proto-ribosomal eras favor triplet codons with a total number of three carbon rings at the first and the third position. This means that the mean crowding and rigidity in the vicinity area of the central nucleotide is preferred. This advantage could also be important in the case of codons UGA and UAA, but the accessibility of these codons was inhibited by the proto- ribosome activity. The assumption of domination of transition over transversion in code expansion and the first supplementary rule of the important rest (Tab. 4),

evokes fewer doubts. The limit of four redefinitions per episode was assumed to approximate the logistic and energetic capabilities of evolving live systems. This assumption makes each of the four nucleotides in the first and second position of the code capable to participate in the redefinition of the amino acid. Also, the limitation of the evolutional activity of the "exotic" amino acids containing sulfur, or 5-carbon ring, (Cys, Met, His, Trp) have practical reasons, related to the expected structural properties of the synthesized polypeptides, i.e. high flexibility and small rigidity. Finally, the constant mass of the redefining amino acids in a given episode expresses the working assumption that the mass of amino acid is the main factor determining the chronology of the evolution of code. With this assumption, a kind of the pseudo-rhythm of the evolutional episodes can be captured. Indicated association between amino acid weight and codon-anticodon duplex stability (Seligmann, 2010), and also a negative correlation between ''size/complexity'' rating and percentage occurrence in proteins (Dufton, 1997), further strengthens the main idea.

Based on the discussed model one may consequently predict that in the future the UGC should gain heavier than tryptophan non-standard amino acid, e.g. pyrrolysine (Pyl). Lighter selenocysteine is incorporated into some proteins by co-translational modification (Gonzalez-Flores at al., 2013) in the place assigned, until proto-code proto-ribosomal era (E.15), to cysteine. Fundamental change including pyrrolysine should be possible due to the redefinition of the meaning of the codons starting from CG doublet (Fig.10), or due to modification of the recognition in the area of the stop codon triplets. The second resolution is already tested in some bacteria and methanogenic archaea where pyrrolysine is defined by UAG codon, universally stop codon assigned to the release factors. Decreasing the number of stop codons seems not to be a generally good resolution for the organisms with big proteomes (Seligmann, 2019).

Referring the created model to the existing concepts, we may place hemoton like cyclic metabolic systems (Ganti 2003) in the proto-code pre-ribosomal warm era, MO like systems in the proto-code pre-ribosomal moderating era, and LUCA like systems in the proto-code proto-ribosomal era.

It was recently shown that the standard genetic code can evolve from a two-letter GC code without information loss or costly reassignments (Frank and Froese, 2018). An expanded (e. g. quadruplet) codons can be thought of in the context of expanding the existing genetic code (Chen & Schindlinger, 2010).

In the contemporary to some other models, assuming

the initial setup of few essential codons and simple amino acids coding the first small proteins (Trifonov and Bettecken, 1997), the proposed "smooth" model considers the evolution starting from a single codon, despite its nonproteinobiotic role. From E.17 the difference in methodology of these attempts is not so important, because the initiation of the ribosome maturation, offers the function of the full three base-codon recognition, and thus the translation of simple proteins. Accepted in the model the division of the evolution into 22 episodes carries some dose of uncertainty regarding exact instantaneous chronology. The aim was to approximate, the evolving population of real individuals which chose evolutionary routes in each episode, in the special individual order. The proposed "smooth" order of integration of amino acids compared to those consistent with the consensus ranking and the rules of abiotic origin, thermostability, complementarity, and processivity ((Trifonov, 2000) were presented in Tab. 5.

Table 5. The order of integration of amino acids

	Model	
Ordinal number	"Smooth evol."	"Consensus 4 rules"
1	Gly	Gly
2	Ala	Ala
3	Ser	Val
4	Pro	Asp
5	Val	Pro
6	Thr	Ser
7	Cys	Glu
8	Ile	Leu
9	Leu	Thr
10	Asn	Arg
11	Asp	Asn
12	Gln	Lys
13	Lys	Gln
14	Glu	Ile
15	Me	Cys
16	His	His
17	Phe	Phe
18	Arg	Me
19	Tyr	Tyr
20	Trp	Trp

The thirty percent identity is observed, especially manifested at the beginning of the evolution.

Some indirect vote for the correctness of the presented concept comes from the known phenomenon of the decrease in G+C content with the decrease in the genome size. Also, the mitochondrial tRNA sequences have a relatively smaller G+C content that their cytoplasmic counterparts. This is in agreement with the taken assumption that early RNA was dominated by A and U basis. Thus the best candidate for the first perceptible codon was respectively rare and strongly bonding doublet GG. It prevents the informatics shock and thermal catastrophe at the beginning of code evolution.

Conclusions

The proposed model is the first such global approximation of the discussed phenomenon of genetic code creation, bringing out the essence of the biological evolution: continuity, smooth and order, and revealing its full chronology. The author allows himself to believe that this model supported by a strong criticism and the further investigated details will bring a better understanding of the most important eon in the evolution of life, the natural creation of the UGC.

Appendix 1
A symbolic description of the evolution of the standard genetic code

The main pathways and processes of evolution were considered (Tab. 6). Applied symbol meaning: => expansion of the code, =>> ∥ code divergence due to the ribosomal recognition ability of the third position of the codon, _ code definition, / code redefinition, the index indicates the version of the given coding system, proto-tRNA, or proto-ribozyme, bold font – preferred version.

Table 6. The main pathways and processes of UGC evolution

I. Pre-code (hot) Era		
Episode	Process	Codons
E.1	$=> t_1$	AAN=>GAN=>GGN
	$=> t_1$	AAN=>AGN=>GGN
E.2	Rz_1	GGN
E.3	Rz_1: Gly + t_1 = Gly$_1$	GGN
	Gly$_1$=>Gly_2	GGN=>GAN
	Gly$_1$=>Gly_3	GGN=>AGN
II. Proto-Code Pre-Ribosomal warm Era		
Episode	Process	Codons
E.4	Ala$_1$/Gly_3	AGN
	Ala$_2$/Gly_2	GAN
	Ala$_2$=>Ala_3	GAN=>GCN
	Ala$_1$=>Ala_4	AGN=>UGN
E.5	Ser$_1$/Ala$_4$	UGN
	Ser$_2$/Ala$_1$	AGN
	Ala$_3$=>Ala_5	GCN=>GUN
	Ala$_3$=>Ala_6	GCN=>ACN
	Ser$_1$=>Ser_3	UGN=>CGN
E.6	Pro$_1$/Ser$_3$	CGN
	Pro$_1$=>Pro_2	CGN=>CAN
	Ser$_1$=>Ser_4	UGN=>UCN
E.7	Val$_1$/Ala$_2$	GAN
	Val$_2$/Ala$_6$	ACN
	Val$_3$/Ala$_5$	GUN
	Val$_4$/Pro$_1$	CGN
	Pro$_2$=>Pro_3	CAN=>CCN
E.8	Thr$_1$/Val$_4$	CGN
	Thr$_2$/Val$_1$	GAN
	Thr$_3$/Val$_2$	ACN
	Thr$_4$/Pro$_2$	CAN
	Pro$_3$=>Pro_4	CCN=>CUN
E.9	Cys$_1$/Ser$_1$	UGN
III. Proto-Code Pre-Ribosomal moderating Era		
Episode	Process	Codons
E.10	Ile$_1$/Thr$_1$	CGN
	Ile$_2$/Thr$_4$	CAN
	Ile$_3$/Thr$_2$	GAN
	Ile$_4$/Pro$_4$	CUN
	Thr$_3$=>Thr_5	ACN=>AUN
E.11	Ile$_5$/Thr$_5$	AUN
	Leu$_1$/Ile$_1$	CGN
	Leu$_2$/Ile$_2$	CAN
	Leu$_3$/Ile$_4$	CUN
	Leu$_3$=>Leu_4	CUN=>UUN
IV. Proto-Code Proto-Ribosomal Era		
Episode	Process	Codons
E.12	Asn$_1$/Ile$_6$	AAN
	Asn$_2$/Ile$_3$	GAN
	Asn$_3$/Leu$_1$	CGN
	Asn$_4$/Leu$_2$	CAN
	t_2	UAN
	Leu$_4$=>>**Leu$_6$**‖Leu$_5$	UUN=>>UUR‖UUY

E.13	Asp_1/Asn_3	CGN
	Asp_2/Asn_4	CAN
	Asp_3/Asn_2	GAN
	Asp_4/Leu_5	UUY
	$t_2=>>t_4\|t_3$	UAN=>>UAR|UAY
	$Cys_1=>>Cys_3\|Cys_2$	UGN=>>UGR|UGY
	$Asn_1=>>Asn_5\|Asn_6$	AAN=>>AAR|AAY
E.14	Gln_1/Asp_1	CGN
	Gln_2/Asp_2	CAN
	Gln_3/Asp_4	UUY
	Gln_4/Asn_5	AAR
	$Gln_3=>>Gln_5$	UUY=>>UAY
	$Ser_2=>>Ser_5\|Ser_6$	AGN=>>AGR|AGY
E.15	Lys_1/Gln_5	UAY
	Lys_2/Gln_1	CGN
	Lys_3/Gln_3	UUY
	Lys_4/Gln_4	AAR
	$Ile_5=>>Ile_7\|Ile_8$	AUN=>>AUR|AUY
	$Gln_2=>>Gln_7\|Gln_6$	CAN=>>CAR|CAY
	$Asp_3=>>Asp_5\|Asp_6$	GAN=>>GAR|GAY
E.16	Glu_1/Gln_6	CAY
	Glu_2/Lys_2	CGN
	Glu_3/Lys_3	UUY
	Glu_4/Asp_5	GAR
	$t_4=>>t_5\|-$	UAR=>>UAG|UAY
	$Cys_3=>>-\|-$	UGR=>>UGG|UGA
	$Ile_7=>>Ile_9\|Ile_{10}$	AUR=>>AUG|AUA
V. Proto-Code Ribosomal Maturation Era		
Episode	Process	Codons
E.17	Met_1/Ile_9	AUG
	$t_5=> t_6$	UAG=>UGG
	r	UAA UGA
E.18	His_1/Glu_1	CAY
	Rz_2	UGG
E.19	Rz_2: <u>Phe</u> + t_6 = <u>Phe</u>$_1$	UGG
	Phe_2/Glu_2	CGN
	Phe_3/Glu_3	UUY
	Phe_4/Lys_1	UAY
E.20	Arg_1/Phe_1	UGG
	Arg_2/Phe_4	UAY
	Arg_3/Phe_2	CGN
	Arg_4/Ser_5	AGR
	$t_5=>-$	UAG
E.21	Tyr_1/Arg_1	UGG
	Tyr_2/Arg_2	UAY
	$r=>s$	UAR
E.22	Trp_1/Tyr_1	UGG

Appendix 2

A detailed story of the evolution of the standard genetic code

I. Pre-code Hot era.

E.1. Proto-tRNAs.

The emergence of the short self-replicating RNAs, precursors of tRNA, with anticodon region capable of slight temporary binding to the complementary codons GAN and AGN of other RNA strands (initially, only the first two positions in a codon are essential for complementarity). They likely originated from the short poly-U strands, then partially mutated to the proto- tRNA strands (t_1), strongly bonding GG doublets in GGN triplets of catalytic structures. Other pathways supporting input to t_1 were also possible. Pathways for the emergence of proto- tRNAs specific to other codons were not thoroughly analyzed. They were likely evolved from t_1.

E.2. Aminoacyl-tRNA synthetase proto-ribozymes.
The emergence of self-replicating aminoacyl-tRNA

synthetase proto-ribozymes (Rz_1), strongly bonded to the codon GGN of proto-tRNA, in the sol-gel reactor.

E.3. Glycine.
The first specific recognition, of the proto- t-RNA (t_1) and glycine (Gly_1), by some proto-ribozyme (Rz_1) and following aminoacylation. Next, the self-replicating ribozyme and proto-tRNA system, using GGN codes for recognition, multiplies and partially evolves to the systems recognizing weaker binding triplets GAN and AGN (first codon expansion).

II. Proto-Code Pre-Ribosomal Warm Era.
E.4. Alanine.
The first redefinition of two codons (Ala_1/Gly_3, Ala_2/Gly_2). Next, the first expansion (transversion) of the coding systems to the codons GCN (Ala_3) and UGN (Ala_4) occurred.

E.5. Serine.
The first redefinition of two codons (Ser_1/Ala_4, Ser_2/Ala_1) and the expansion (transition) of three codes ($Ala_3 => Ala_5$, $Ala_3 => Ala_6$, $Ser_1 => Ser_3$) occurred.

E.6. Proline.
The redefinition of the code (Pro_1/Ala_2) and the expansion of two codes ($Pro_1 => Pro_2$, $Ser_1 => Ser_4$) occurred.

E.7. Valine.
The redefinition of four codons (Val_1/Ala_2, Val_2/Ala_6, Val_3/Ala_5, Val_4/Pro_1) and the expansion of one code ($Pro_2 => Pro_3$) occurred.

E.8. Threonine.
The redefinition of four codons (Thr_1/Val_4, Thr_2/Val_1, Thr_3/Val_2, Thr_4/Pro_2) and the expansion of one code ($Pro_3 => Pro_4$) occurred.

E.9. Cysteine.
The redefinition of one codon ($Cysr_1/Ser_1$) occurred.

III. Proto-Code Pre-Ribosomal Moderating Era.
E.10. Isoleucine.
Weak binding codons began to be useful. The redefinition of four codons (Ile_1/Thr_1, Ile_2/Thr_4, Ile_3/Thr_2, Ile_4/Pro_4,) and the expansion of one code ($Thr_3 => Thr_5$) occurred.

E.11. Leucine.
The redefinition of four codons (Ile_5/Thr_5, Leu_1/Ile_1, Leu_2/Ile_2, Leu_3/Ile_4) and the expansion of two codes ($Ile_3 => Ile_6$, $Leu_3 => Leu_4$) occurred. Some supramolecular molecules composed of the amino acid loaded proto-tRNA and the ribozymes get organized and begun to play some translational role in the simple processes of the synthesis of the first peptides along with the RNA matrix.

IV. Proto-Code Proto-Ribosomal Era.
E.12. Asparagine.
The redefinition of four codons (Asn_1/Ile_6, Asn_2/Ile_3, Asn_3/Leu_1, Asn_4/Leu_2) took place and the

initiation of the activity of the proto-ribosomes occurred. Proto-tRNAs (t_2), with anticodon region capable of strong binding to the complementary codons UAN marked the position of the first stop codons and started to play the role of proto–release factor. The third position of the codon begun to be essential for proto ribosomes enabling further differentiation of the aminoacyl-tRNA synthetase ribozymes, and the ribosomal distinguishing between different ribozyme coding systems, initially recognizing the last purines and the last pyrimidines (UUR||UUY). Systems with five carbon rings in the codon were preferred and overexpressed (**Leu_5**||Leu_5). The weak binding, next, the less accessible, codons were evolved first.

E.13. Aspartic acid.
The redefinition of four codons (Asp_1/Asn_3, Asp_2/Asn_4, Asp_3/Asn_2, Asp_4/Leu_5) took place. The last one utilizes the mining of the third position of the codon, Further three distinctions between third position purines and pyrimidines (t_4||t_3, **Cys_3**||Cys_2, Asn_5||**Asn_6**) occurred.

E.14. Glutamine.
The redefinition of four codons (Gln_1/Asp_1, Gln_2/Asp_2, Gln_3/Asp_4, Gln_4/Asn_5) took place. Then, one transitional expansion ($Gln_3 => Gln_5$) occurred. One distinction between third position purines and pyrimidines (Ser_5||**Ser_6**) occurred, too.

E.15. Lysine.
The redefinition of four codons (Lys_1/Gln_5, Lys_2/Gln_1, Lys_3/Gln_3, Lys_4/Gln_4) took place. Three distinctions between third position purines and pyrimidines (Ile_7||**Ile_8**, **Gln_7**||Gln_6, Asp_5||**Asp_6**,) occurred.

E.16. Glutamic acid.
The redefinition of four codons (Glu_1/Gln_6, Glu_2/Lys_2, Glu_3/Lys_3, Glu_4/Asp_5) took place. Further distinctions between third position guanines and adenines occurred (AUG||**AUA**, **UAG**||UAA, **UGG**||UGA) leading to the differentiation of a codons (Ile_9||**Ile_{10}**), partial reduction of proto-tRNA based stop codons (t_5), and the deactivation of one coding system (Cys_3).

V. Proto-Code Ribosomal Maturation Era.
E.17. Methionine.
The redefinition of one codon (M_1/Ile_9) and one transitional expansion of proto-tRNA ($t_5 => t_6$) occurred. The first protein-based release factor (r) took the two codons (UAA, UGA).

E.18. Histidine.
The redefinition of one codon (His_1/Glu_1) and the reactivation of post-citosinic ribozyme Rz_2 capable to recognize tRNAs (t_6) and slightly bent to the codon UGA, occurred.

E.19. Phenylalanine.
The specific recognition of the amino acid, phenylalanine (F_1), by some ribozymes (Rz_2) and then the aminoacylation of a targeted tRNAs (t_6)

took place. Next, the redefinition of three codons (Phe_2/Glu_2, Phe_3/Glu_3, Phe_4/Lys_1) occurred.

E.20. Arginine.

The redefinition of four codons (R_1/F_1, Arg_2/Phe_4, Arg_3/Phe_2, Arg_4/Ser_5) occurred. Ribosomal recognition of proto-tRNA release factor (t_5) was inhibited due to the optimization of the arg-tRNA reception.

E.21. Tyrosine.

The redefinition of two codons (Y_1/R_1, Tyr_2/Arg_2) occurred. The second protein-based release factor (s), the derivative of r, took the two codons (UAG, UAA).

E.22. Tryptophan.

The redefinition of one codon (W_1/Y_1) occurred.

Abbreviations

UGC: the universal genetic code.
Obs. - observation
E. - episode

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and material
Not applicable.

Competing interests
The author declares that he has no competing interests.

Authors' contributions
PHP -100%

References

1. Berg J.M., Tymoczko J.L., Stryer L. (2002). Biochemistry. 5th edition. New York: W. H. Freeman; 2002. Section 5.5. Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point. Available from: http://www.ncbi.nlm.nih.gov/books/NBK22342/
2. Chechetkin V.R. (2006). Genetic code from tRNA point of view. J Theor Biol. 2006;242(4):922–934.
3. Chen I.A., Schindlinger M. (2010). Quadruplet codons: One small step for a ribosome, one giant leap for proteins. An expanded genetic code could address fundamental questions about algorithmic information, biological function, and the origins of life. Bioessays. 2010 Aug; 32(8): 650–654.
doi: 10.1002/bies.201000051
4. Crick F.H. (1968). The origin of the genetic code. J Mol Biol. 1968;38(3):367–79.
5. Duan C., Huan Q., Chen X., Wu S., Carey L.B., Xionglei, H., Wenfeng, Q. (2018). Reduced intrinsic DNA curvature leads to increased mutation rate. Genome Biology 19, 132. Available from: https://doi.org/10.1186/s13059-018-1525-y
6. Dufton M.J. (1997). Genetic code synonym quotas and amino acid complexity: cutting the cost of proteins? J Theor Biol. 1997;187:165–173.
7. Demongeot, J., & Seligmann, H. (2019b). Bias for 3'-dominant codon directional asymmetry in theoretical minimal RNA rings. J Comput Biol, in press.
8. Freeland S.J., Knight R.D., Landweber L.F., Hurst L.D. (2000). "Early fixation of an optimal genetic code". Molecular Biology and Evolution. 17 (4): 511–18.
9. Forster A.C., Church, G.M. (2006). Towards synthesis of a minimal cell. Molecular Systems Biology 2:45.
10. Frank A., Froese T. (2018). The Standard Genetic Code can Evolve from a Two-Letter GC Code Without Information Loss or Costly Reassignments. Orig Life Evol Biosph. 48(2):259-272. doi: 10.1007/s11084-018-9559-4
11. Guilloux A., Jestin J.L. (2012). The genetic code and its optimization for kinetic energy conservation in polypeptide chains. Biosystems. 2012;109(2):141–144.
12. Gamow, G. (1954). Possible relation between deoxyribonucleic acid and protein structures. Nature, 1954. 173:318.
13. Gánti Tibor (2003): Chemoton Theory. Vol. I.Theory of Fluid Machineries. Kluwer Academic/Plenum Publishers, New York
14. Gánti Tibor (2003): Chemoton Theory. Vol. II. Theory of Living Systems. Kluwer Academic/Plenum Publishers, New York
15. Di Giulio M. (2003). The late stage of genetic code structuring took place at high temperature. Gene. 2003;261:189–195.
16. Gonzalez-Flores JN, Shetty SP, Dubey A, Copeland PR. (2013). The molecular biology of selenocysteine. Biomol Concepts. 4(4):349–365. doi:10.1515/bmc-2013-0007
17. Gotoh, O., (1983). Prediction of melting pro®les and local helix stability for sequenced DNA. Adv. Biophys. 16, 1±52.
18. Guimarães RC (2013) Formation of the genetic code – a review and update as of November 2012. http://www.icb.ufmg.br/labs/lbem/pdf/GMRTgeneticodeNov12.pdf Original publications available from: https://www.researchgate.net/profile/Romeu_Guimaraes
19. Henderson L.J., The fitness of the environment: an inquiry into the biological significance of the properties of matter The Macmillan Company, 1913.
20. Itzkovitz S., Alon U. (2007). The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Research, 17(4), 405–412.
21. Koonin E.V. Novozhilov A.S. (2017). Origin and Evolution of the Universal Genetic Code; Annual Review of Genetics. 2017. Vol. 51:45-62
22. Kumar B.,Saini S., (2016) Analysis of the optimality of the standard genetic code. Mol. BioSyst., 2016,12, 2642-2651. doi:10.1039/C6MB00262E
23. Mat W.K., Xue H., Wong J.T. (2008). The genomics of LUCA. Front Biosci. 1;13:5605-13.
a. Nick L., B. Yoshitaka, Wei. Kenneth J.W. Szostak, S. Hiroaki. (2000). Ribozyme-catalyzed tRNA aminoacylation. Nature Structural Biology. 7, 28. Nature America, https://doi.org/10.1038/71225
24. Nirenberg, M. W. and Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *PNAS*, *47*(10), 1588-1602. Available from:
a. http://dx.doi.org/10.1073/pnas.47.10.1588
25. Pelc S.R., Welton M.G.E. (1966). Stereochemical relationship between coding triplets and amino acids. Nature. 1966;209:868–872.

26. Seligmann H. (2010). Do anticodons of misacylated tRNAs preferentially mismatch codons coding for the misloaded amino acid? BMC Mol Biol. 2010;11:41.

a. Seligmann H. (2011). Error compensation of tRNA misacylation by codon-anticodon mismatch prevents translational amino acid misinsertion. Comput Biol Chem. 2011; 35:81–95.

27. Seligmann H. (2019). Localized Context-Dependent Effects of the "Ambush" Hypothesis: More Off-Frame Stop Codons Downstream of Shifty Codons. DNA Cell Biol. 38(8):786-795. doi: 10.1089/dna.2019.4725.

28. Sonneborn T.M. (1965). Degeneracy of the genetic code: extent, nature, and genetic implications. Evolving genes and proteins. 1965:377–397.

29. Söll, D., Ohtsuka, E., Jones, D. S., Lohrmann, R., Hayatsu, H., Nishimura, S., and Khorana, H. G. (1965). Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *PNAS*, *54* (5), 1378-1385. Available from:

a. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC219908/.

30. Trifonov E.N., (1997). T. Bettecken. Sequence fossils, triplet expansion, and reconstruction of earliest codons. Gene. 1997: 205, Issue: 1-2, Page: 1-6.

31. Trifonov E.N. (2000). Consensus temporal order of amino acids and evolution of the triplet code. Gene. 2000;261:139–151. doi: 10.1016/S0378-1119(00)00476-5.

32. Woese C.R. (1965). On the evolution of the genetic code. Proc Natl Acad Sci USA. 1965;54(6):1546–1552. Available from:

a. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC300511/

33. Wong J.T.F. (1975). A Co-Evolution Theory of the Genetic Code. Proceedings of the National Academy of Sciences. 1975;72(5):1909–1912. Available from:

34. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC432657/

35. Wong, J.T.F. (2005). The coevolution hypothesis at age thirty. Bioessays 27, 416-426.

36. Xia, T., SantaLucia, J., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., Turner, D.H. (1998). Thermodinamical parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 37, 14719±14735.

Figure legends

Fig.1 The average mass of coded amino acids for the groups of codons bounding by a given number of hydrogen bonds (from 6 to 9 HB). The hypothetical course of the evolution was indicated by the arrow.

Fig.2. Metabolism of amino acids. An empty arrow indicates the center of the metabolic pathways system, arbitrarily chosen within the glycolysis route.

Fig.3. The average mass and metabolic distance of amino acids from the assumed center of metabolic system for the distinguished similarity families: Ser –serine, Pyr – Pyruvate, Glu- glutamate, Asp-aspartate, Aro – aromatic and His – histidine family. The hypothetical course of the evolution was indicated by the arrow. The colors are as in Fig.2.

Fig.4. The average mass of coded amino acids for the groups of codons of different density of packing.

The hypothetical course of the evolution was indicated by the arrow. The colors are as in Tab. 2.

Fig.5. Accessibility of the codon by the anticodon during mRNA-tRNA triplet stacking in the translation process. Broken lines mark the available space for triplets fully approach (a). The order and size of the bases may influence codon accessibility positively (b), or negatively (c).

Fig.6. The average molecular mass of coded amino acids vs. the measure of the accessibility of the codon. Colors meaning as in Tab. 3. The hypothetical course of the evolution was indicated by the arrow.

Fig.7. Hypothetical picture of the synthesis of the first tRNA precursor. In the gel-like reactor RNA strands and amino acid molecules could meet closely together (a). This could result, both in a complementary, and stereochemical, fitting (b) initiating proto-tRNA loading (c, d). Fig.8. The evolution of UGC. Main episodes (E.1 - E.22). Rows\columns define nucleotides in the first\second position of a codon. The third position of a codon, if essential, is defined in the divided field, respectively in the order: GA for purines, and CU for pyrimidines. Color font, meaning: red- the code definition (underlined) or redefinition (encircled). Field color meaning: gray- the codon hardly or unstable binding in the hot, warm or moderating era, green- new ribosomal property related to the recognition of the third position by the ribosome. The arrows indicate the main routes of codon expansion. Font meaning: bold- preferred systems when the third position becomes essential, italic-new expanded code for the amino acid, tRNA anticodon or the recognition codon of the ribozyme. The empty field manifests as a nonsense codon or a stop codon.

Fig.9. Pseudo-dynamics of UGC formation. The molecular weight of evolutionary new amino acid coded versus the episode number. At the bottom main eras are indicated: pre-code hot era (red), proto-code pre-ribosomal warm era (orange), proto-code pre-ribosomal moderating era (gray), proto-code proto-ribosomal era (blue), proto-code ribosomal maturation era (green). Symbols represent the main events: the first coded amino acid (gold circle), the beginning, and clarification of the ribosomal codon third position recognition function (small and big diamond), the incorporation of the release factor (square).

Fig.10. The future of genetic code evolution. Nonstandard amino acid pyrrolysine (Pyl) assigned due to the redefinition of the bi-codon CG.
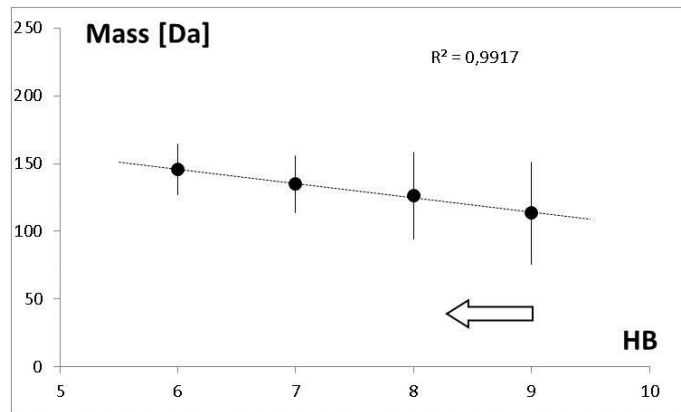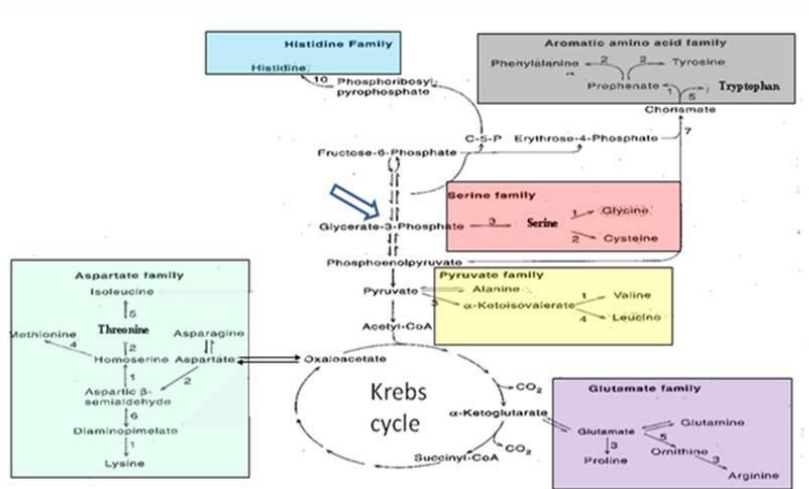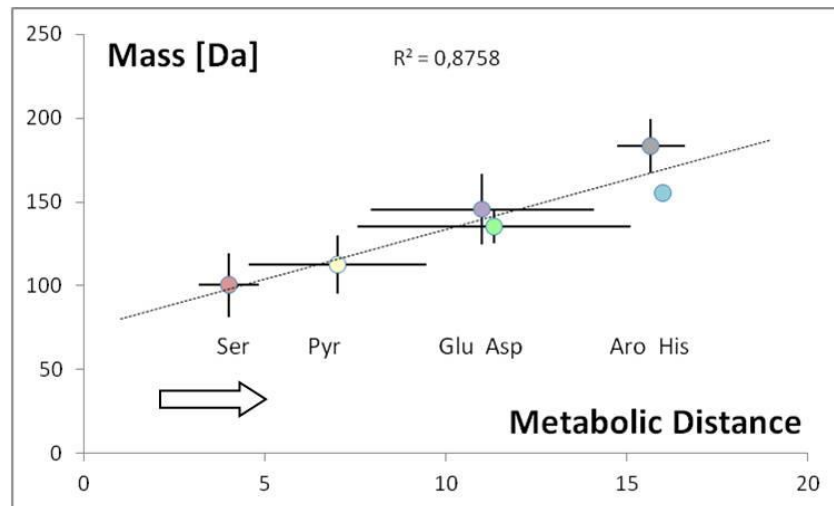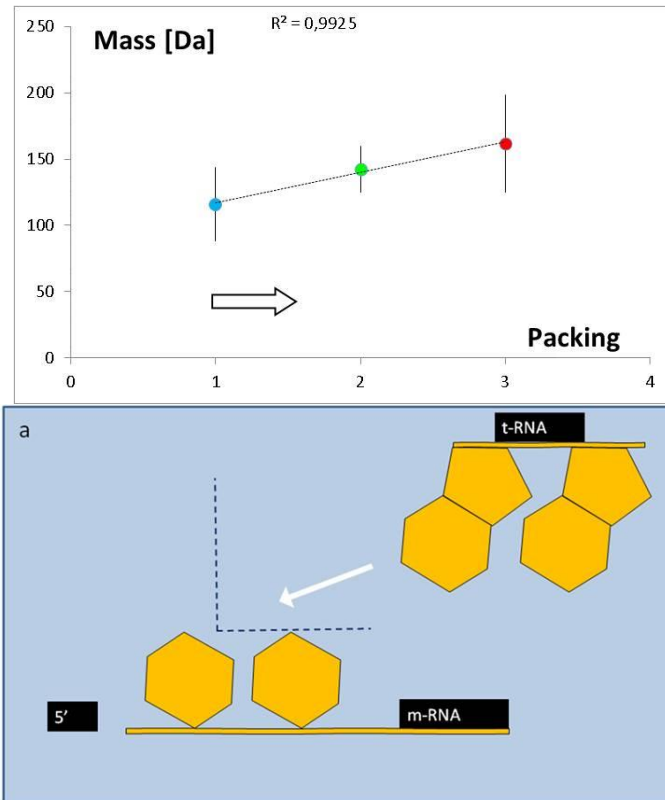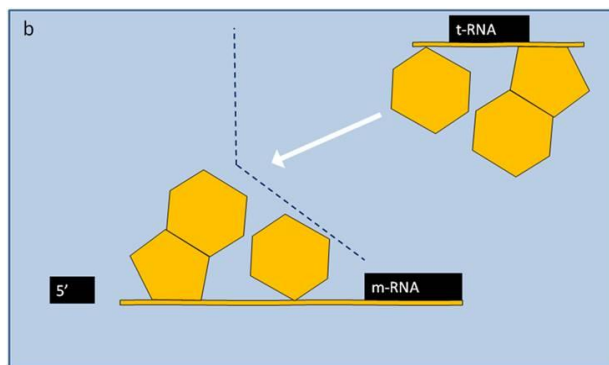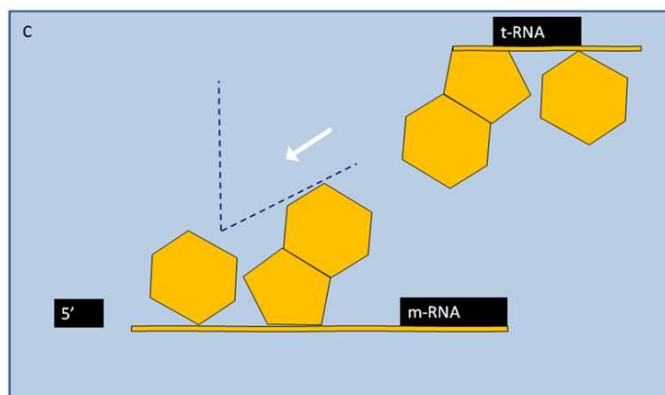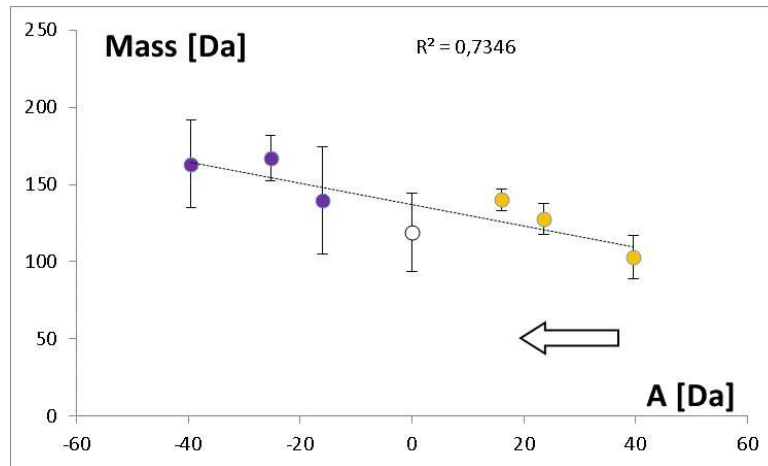
Figures

Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

Fig. 5a.

Fig. 5b.

Fig. 5c.

Fig. 6.
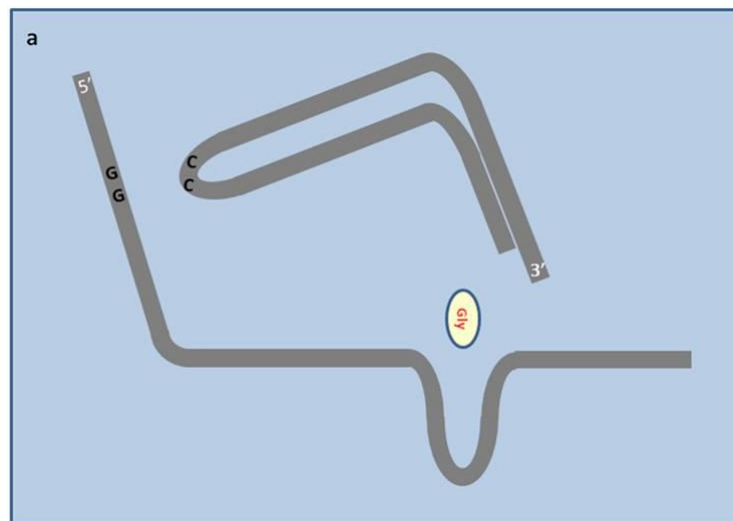
Fig. 7a.
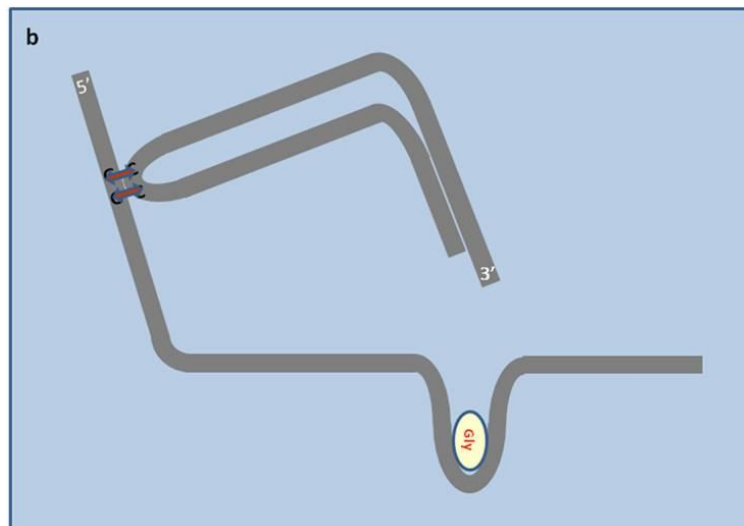
Fig. 7b.
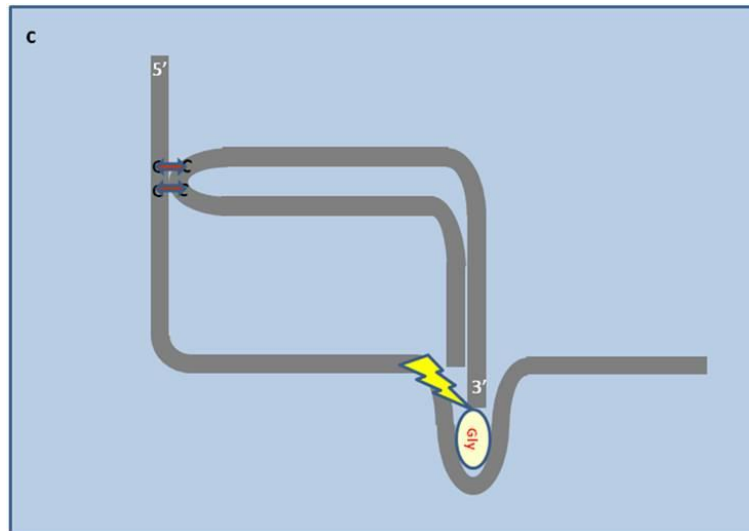
Fig. 7c.
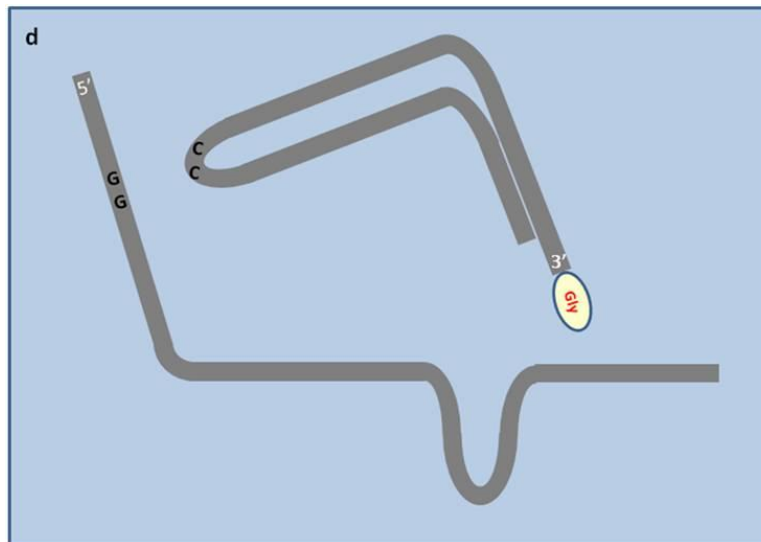
Fig. 7d.

Episode 1				
$N_1 \backslash N_2$	G	C	A	U
G	t_1			
C				
A				
U				

Fig. 8. E. 1.

Episode 2				
$N_1 \backslash N_2$	G	C	A	U
G	$t_1\ Rz_1$			
C				
A				
U				

Fig. 8. E. 2.

Episode 3				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1		Gly_2	
C				
A	Gly_3			
U				

Fig. 8. E.3.

Episode 4				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Ala_2	
C				
A	Ala_1			
U	Ala_4			

Fig. 8. E. 4.

Episode 5				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Ala_2	Ala_5
C	Ser_3			
A	Ser_2	Ala_6		
U	Ser_1			

Fig. 8. E5.

Episode 6				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Ala_2	Ala_5
C	Pro_1		Pro_2	
A	Ser_2	Ala_6		
U	Ser_1	Ser_4		

Fig. 8. E. 6.

Episode 7				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Val_1	Val_3
C	Val_4	Pro_3	Pro_2	
A	Ser_2	Val_2		
U	Ser_1	Ser_4		

Fig. 8. E. 7.

Episode 8				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Thr_2	Val_3
C	Thr_1	Pro_3	Thr_4	Pro_4
A	Ser_2	Thr_3		
U	Ser_1	Ser_4		

Fig. 8. E. 8.

Episode 9				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Thr_2	Val_3
C	Thr_1	Pro_3	Thr_4	Pro_4
A	Ser_2	Thr_3		
U	Cys_1	Ser_4		

Fig. 8. E. 9.

Episode 10				
$N_1 \backslash N_2$	G	C	A	U
G	Gly_1	Ala_3	Ile_3	Val_3
C	Ile_1	Pro_3	Ile_2	Ile_4
A	Ser_2	Thr_3		Thr_5
U	Cys_1	Ser_4		

Fig. 8. E. 10.

Episode 11				
N₁\N₂	G	C	A	U
G	Gly₁	Ala₃	Ile₃	Val₃
C	Leu₁	Pro₃	Leu₂	Leu₃
A	Ser₂	Thr₃	Ile₆	Ile₅
U	Cys₁	Ser₄		Leu₄

Fig. 8. E. 11.

Episode 12					
N₁\N₂	G	C	A	U	
G	Gly₁	Ala₃	Asn₂	Val₃	
C	Asn₃	Pro₃	Asn₄	Leu₃	
A	Ser₂	Thr₄	Asn₁	Ile₅	
U	Cys₁	Ser₄	t₂	Leu₆	Leu₅

Fig. 8. E. 12.

Episode 13						
N₁\N₂	G	C	A		U	
G	Gly₁	Ala₃	Asp₃		Val₃	
C	Asp₁	Pro₃	Asp₂		Leu₃	
A	Ser₂	Thr₄	Asn₅	Asn₆	Ile₅	
U	Cys₃	Cys₂	Ser₄	t₄	t₃	Leu₆	Asp₄

Fig. 8. E. 13.

Episode 14						
N₁\N₂	G		C	A		U
G	Gly₁		Ala₃	Asp₃		Val₃
C	Gln₁		Pro₃	Gln₂		Leu₃
A	Ser₅	Ser₆	Thr₄	Gln₄	Asn₆	Ile₅
U	Cys₃	Cys₂	Ser₄	t₄	Gln₅ ⟵ Leu₆ ⟸ Gln₃	

Fig. 8. E. 14.

Episode 15						
N₁\N₂	G		C	A		U
G	Gly₁		Ala₃	Asp₅ Asp₆		Val₃
C	Lys₂		Pro₃	Gln₇	Gln₆	Leu₃
A	Ser₅	Ser₆	Thr₄	Lys₄	Asn₆	Ile₇ Ile₈
U	Cys₃	Cys₂	Ser₄	t₄	Lys₁	Leu₆ Lys₃

Fig. 8. E. 15.

Episode 16							
N₁\N₂	G		C	A			U
G	Gly₁		Ala₃	Glu₄	Asp₆		Val₃
C	Glu₂		Pro₃	Gln₇	Glu₁		Leu₃
A	Ser₅	Ser₆	Thr₄	Lys₄	Asn₆	I₉ I₁₀	Ile₈
U		Cys₂	Ser₄	t₅		Lys₁	Leu₆ Glu₃

Fig. 8. E. 16.

Episode 17								
N₁\N₂	G		C		A			U
G	Gly₁		Ala₃		Glu₄	Asp₆	Val₃	
C	Glu₂		Pro₃		Gln₇	Glu₁	Leu₃	
A	Ser₅	Ser₆	Thr₄		Lys₄	Asn₆	M₁ I₁₀	Ile₈
U	t₆	r	Cys₂	Ser₄	t₅	r	Lys₁ Leu₆	Glu₃

Fig. 8. E. 17.

Episode 18								
N₁\N₂	G		C		A			U
G	Gly₁		Ala₃		Glu₄	Asp₆	Val₃	
C	Glu₂		Pro₃		Gln₇	His₁	Leu₃	
A	Ser₅	Ser₆	Thr₄		Lys₄	Asn₆	M₁ I₁₀	Ile₈
U	Rz₂	r	Cys₂	Ser₄	t₅	r	Lys₁ Leu₆	Glu₃

Fig. 8. E. 18.

Episode 19								
N₁\N₂	G		C		A			U
G	Gly₁		Ala₃		Glu₄	Asp₆	Val₃	
C	Phe₂		Pro₃		Gln₇	His₁	Leu₃	
A	Ser₅	Ser₆	Thr₄		Lys₄	Asn₆	M₁ I₁₀	Ile₈
U	F₁	r	Cys₂	Ser₄	t₅	r	Phe₄ Leu₆	Phe₃

Fig. 8. E. 19.

Episode 20							
N₁\N₂	G		C	A		U	
G	Gly₁		Ala₃	Glu₄	Asp₆	Val₃	
C	Arg₃		Pro₃	Gln₇	His₁	Leu₃	
A	Arg₄	Ser₆	Thr₄	Lys₄	Asn₆	M₁ I₁₀	Ile₈
U	R₁ r	Cys₂	Ser₄	r	Arg₂	Leu₆	Phe₃

Fig. 8. E. 20.

Episode 21							
N₁\N₂	G		C	A		U	
G	Gly₁		Ala₃	Glu₄	Asp₆	Val₃	
C	Arg₃		Pro₃	Gln₇	His₁	Leu₃	
A	Arg₄	Ser₆	Thr₄	Lys₄	Asn₆	M₁ I₁₀	Ile₈
U	Y₁ r	Cys₂	Ser₄	s r s	Tyr₂	Leu₆	Phe₃

Fig. 8. E. 21.

Episode 22							
N₁\N₂	G		C	A		U	
G	Gly₁		Ala₃	Glu₄	Asp₆	Val₃	
C	Arg₃		Pro₃	Gln₇	His₁	Leu₃	
A	Arg₄	Ser₆	Thr₄	Lys₄	Asn₆	M₁ I₁₀	Ile₈
U	W₁ r	Cys₂	Ser₄	s r s	Tyr₂	Leu₆	Phe₃

Fig. 8. E. 22.

Fig. 9.

Fig. 10.