Mathematical Biology and Bioinformatics
2015. V. 10. № 1. P. 116–130. doi: 10.17537/2015.10.116

UDC: 577.217.347

# A search for relict ribonucleotide and amino acid sequences that played a key role in the development of the ribosome and modern protein diversity

**Skoblikow N.E.[1], Zimin A.A.[2]**

[1]*North Caucasian Research Institute of Animal Husbandry, Krasnodar, Russia*
[2]*G.K.Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Moscow region, Russia*

***Abstract.*** The study presents the results of analysis of protein sequence database for prokaryotic microorganisms, which revealed a conservative peptide sequence element of 11 amino acid residues in 20 loci of 16 functionally and phylogenetically differing conservative proteins from representatives of various taxa. This amino acid motif IKAVRELGLER is presumably one of the Last Universal Peptide Ancestors (LUPAs). A fragment of ribosomal RNA (part of the A-site including stems H92, H90 and H93 of the peptidyl transferase center, PTC) translated from one of the potential reading frames is likely to be a source of genetic information for this sequence. We define this m/rRNA fragment with a function of a template for LUPA synthesis as the Last Universal RiboNucleic Ancestor (LURNA). We assume that LURNA and the peptides translated from its sequence were a source of the modern diversity of peptides.

***Key words:*** *ribosome, ribosomal RNA, translation, peptidyl transferase center, universal peptide motif, Last Universal Peptide Ancestor, LUPA, Last Universal RiboNucleic Ancestor, LURNA.*

## INTRODUCTION

The question of monophyletic origin of template-encoded proteins at first thought seems as not simply an unsolvable, but rather a hardly set task. The general point of view presumes the existence of a certain initial variety of proteins [1, 2]. All classifications of modern proteins and specific conservative peptide motifs are based on their differences, rather than similar features [3]. Moreover, even if the source of such diversity (hypothetical "common ancestral gene") ever existed, it is absolutely unclear if it is still preserved and where it can be found. Which of the nucleotide sequences from the extensive available databases can be considered as ancestral?

When trying to find such sequence, it might be worth noticing the RNA molecules involved in the core of the modern translation apparatus – the ribosomal RNAs. They are the most ancient large conservative heterogeneous polymers, common for all cellular forms of life and potentially capable to fulfill the carrier function for hereditary information about protein structure. Traditionally these non-coding molecules were not considered as potential carriers of genetic information. However, recent studies have demonstrated that rRNAs molecules can encode a number of tRNAs [4] and tRNAs and proteins at the same time [5].

The first task in applying such approach is choosing a particular RNA molecule(s) and/or its part as a potential template for translation, since rRNA molecules are heterogenous in structure and function.

All types of rRNA molecules (23S, 16S and 5S in prokaryotes) significantly differ within a genus and even a species, laying a background for phylogenetic differentiation. Notably, these differences concern not only the primary sequence (nucleotide substitutions), but also the secondary and tertiary structure, which is critical for the hypothetical translation of such RNA. Single stems and loops and their groups often originate and disappear. Differences between eukaryotic rRNAs are greatly expressed, especially in higher animals and plants, where these molecules have larger size due to the presence of many additional stems and loops. On the other hand, rRNA molecules of eukaryotic symbionts, mitochondria and plastids, are characterized by a loss of many structural elements, most likely owing to their reductive evolution [6]. Expressed heterogeneity of nucleotide composition hardly allows considering full-length rRNA molecules as the most ancient source of genetic information [5]. Frequently occurring substitutions in rRNAs even in closely related microorganisms, as well as their structural differences in prokaryotes from different taxa, might lead to heterogeneity in the results of hypothetical translation when a ribosomal RNA of a certain microorganism is chosen as a template. Perhaps, only the most conservative sites of rRNA molecules, having no structural variations and low divergence in the nucleotide composition between the members of various, even the most phylogenetically distant taxa, can be considered as an ancient source of hereditary information.

According to a popular point of view [7, 8] the modern 23S rRNAs in all their variety are the products of consequential evolution of a rather small site of the RNA chain into a complex multidomain ribonucleotide structure. In most cases the main part of the V domain of the 23S rRNA – the peptidyl transferase center (PTC) is considered to be the most ancient "core" of the ribosome [7, 8]. This particular part of the ribosome is directly responsible for the realization of its major role in the process of peptide chain formation during translation.

The secondary structure of the PTC consists of a specific Peptidyl Transferase Ring (PTR) and the outgoing stems (enumerated from H74 to H93), which divide the PTR into 5 segments [9]. The primary structure of the PTC is represented not by a continuous linear region of the 23S rRNA molecule, but by a combination of three fragments:

1. 5’-fragment (containing one segment of the PTR and the 5’-chains of stems H74 and H75);
2. small middle fragment (containing the 3’-chain of stem H75 and stem H80) and
3. the largest 3’-fragment (containing four segments of the PTR, the 3’-chain of stem H74 and stems H89, H90, H91, H92 and H93).

We chose the last region as a potential source of translated information. We suggested allocating the most conservative fragment from this part, which lacks structural variations (insertions and deletions) between various microorganisms, sufficient in length for hypothetical translation of peptide(s) consisting of tens of amino acid residues.

A representative selection of microorganisms must contain at least 20 prokaryotes from the main groups (cyanobacteria, proteobacteria, thermophilic gram-negative bacteria, mycoplasma, firmicutes, actinobacteria and archaea of different phyla). Using only one microorganism as a source of rRNA (for instance, *E.coli*, [5]) can bring incorrect results in determination of the invariable in the structure and sequence rRNA fragment. Eukaryotes should not be included in this group because of their obviously late origin. The ribosomal RNAs of mitochondria and chloroplasts also ought to be discarded on account of their reductive evolution [6].

Comparison of PTC-forming regions of the 23S rRNAs from microorganisms of different taxa will allow to define the borders and the length of sequence with structural conservation, and to identify universal invariant nucleotides. Consideration of hypervariable sites requires consideration of not only the occurrence frequency of a nucleotide in a given set of microorganisms, but also the general phylogeny of prokaryotes [10]. Upon determining the consensus sequence of such RNA fragment and defining of potential reading frames, theoretical results of translation can be proposed.

A search for these sequences in proteins will require an application of some restrictions, based on the modern view on the presumed pre-ribosomal coding.

First of all, it is necessary to consider the difference between the amino acids involved in ancient translation systems and the modern amino acid spectrum. For this reason for further search in protein sequences it is necessary to avoid the amino acids which are (according to widespread viewpoints [11]) evolutionarily more recent products of cellular biosynthesis and are detected in minimal quantities in abiogenic synthesis experiments [12]. Among such amino acids are tryptophane, tyrosine, phenylalanine and histidine, which do not have abiogenically synthesized structural homologs. It may be assumed that ornithine was included in ancient peptides as a metabolic predecessor and structural homolog of arginine and lysine [11], while 5-oxoproline / pyroglutamate could be used instead of proline. Notably, both of these amino acids occur in peptides synthesized in non-template synthesis.

Secondly, it is necessary to consider that the ancient code could be different from the standard one (not to mention its lower specificity allowing for encoding amino acid homologs [11]). However, in our objective we will use the standard code.

Thirdly, the fragments which include the stop codons should be eliminated from consideration.

After picking out of the remaining (possibly, rather short) amino acid sequences which meet all the requirements, it is necessary to estimate their presence in taxonomically different prokaryotes (first of all in the selected microorganisms). The longest selected sequence translated from an rRNA site with the lowest variability will be the most perspective result. The rejected sequences also can be subjects for consideration, but not within this work.

It is unlikely that we can find the complete and exact copies of the required sequence. Most likely, some well conserved fragments will be detected, especially those (if the hypothesis is right), which are synthesized in translation from the most conservative rRNA sites.

Multiple alignment of homologous proteins containing elements of the target sequence from various prokaryotes is necessary to confirm their homology with the analyzed sequence and exclude the possibility of their origination as a result of accidental mutational and recombination events. In case of coincidence of the context (confirmation of the presence of the consensus sequence in the same site of homologous proteins), such fragments should be compiled into a list of conservative loci of homologous proteins carrying the target sequence. The following selection criteria may be used for the identified protein fragments:

- good preservation of the sequence in the same protein in different prokaryotes;
- detection of the sequence in the same site of a protein in different prokaryotes;
- presence of the sequence in the same site of a protein in a wide spectrum of phylogenetically distant microbes.

An additional argument speaking for the relict origin of such conservative fragments comes from the functions of the respective proteins. These proteins are presumed to be important, universal and evolutionarily ancient (primarily, nucleic acid polymerases, aminoacyl-tRNA synthetases, ribosomal proteins and translation factors).

The relict nature of the target sequences could have become a basis for their conservation in the structure of different proteins carrying out various functions, providing a background for their diversity. For this reason the identification of the required sequence elements in the functional parts of proteins is by no means obligatory. Quite on the contrary, their detection in the protein loci with minimal specific activity is more probable.

We suggest that detection of even one such relatively extensive peptide motif in a number of phylogenetically and functionally different conservative proteins of taxonomically different prokaryotes, will speak for consideration of the selected rRNA fragment as a hypothetical universal mRNA ("RNA-gene") ancestral for modern cellular organisms (Last Universal RiboNucleic Ancestor – LURNA).

Moreover, it is quite probable that one of the several selected motifs translated from the same nucleotide sequence will be present in proteins with the highest frequency. Still, even it that motif (or its fragment) has the highest frequency of occurrence in many basic proteins of phylogenetically distant prokaryotes, one can hardly speak of the only universal peptide ancestor for most of the modern organisms (Last Universal Peptide Ancestor – LUPA). However, various peptide products translated from different reading frames of a single ancient RNA molecule can likely be considered as a cluster of peptide predecessors having a single genetic source.

## METHODS

Twenty sequences of the 23S rRNA genes from various prokaryotic microorganisms of different taxa were collected from the GenBank database [13] (Table 1).

A consensus sequence of a structurally conservative 23S rRNA fragment was determined after multiple alignment of 23S rRNA sites containing the major part of the PTC. The consensus sequence for the hypervariable regions of the analyzed rRNAs was deduced considering the proximity of the source microorganisms to the phylogenetic trunk [3].

The analysis of the results of hypothetical translation (amino acid sequences) was carried out using the standard code. Only products of hypothetical translation having no stop codons and "evolutionarily late" biogenic amino acids (W, Y, F, H) were included into the list of established amino acid sequences subject to a search.

The search for the amino acid sequence in prokaryotes of different taxa was carried out using Protein BLAST [14] with *blastp* algorithm [15].

**Table 1.** List of microorganisms selected as sources of the 23S rRNA genes

| # | Specific name | Taxonomic group | Gene ID | Gene length |
|---|---|---|---|---|
| 1 | *Synechococcus sp. RCC307* | Cyanobacteria | NC_009482.1 | 2866 |
| 2 | *Aquifex aeolicus VF5* | Aquificae | NC_000918.1 | 2956 |
| 3 | *Thermus thermophilus JL-18* | Deinococcus-Thermus | NC_017587.1 | 2910 |
| 4 | *Ehrlichia ruminantium str. Welgevonden* | Alphaproteobacteria | NC_006832.1 | 2765 |
| 5 | *Escherichia coli str. K-12 substr. W3110* | Gammaproteobacteria | NC_007779.1 | 2904 |
| 6 | *Treponema pallidum str. Fribourg-Blanc* | Spirochaetes | NC_021179.1 | 2951 |
| 7 | *Chlamydia trachomatis A2497* | Chlamydiae | NC_017437.1 | 2940 |
| 8 | *Mycoplasma pneumoniae M129-B7* | Mollicutes | NC_020076.1 | 2894 |
| 9 | *Clostridium cf. saccharolyticum K10* | Firmicutes | NC_021047.1 | 2884 |
| 10 | *Corynebacterium diphtheriae C7* | Actinobacteria | NC_016801.1 | 3083 |
| 11 | *Thermoplasma volcanium GSS1* | Euryarchaeota | NC_002689.2 | 2907 |
| 12 | *Pyrococcus furiosus COM1* | Euryarchaeota | NC_018092.1 | 2872 |
| 13 | *Methanococcus maripaludis X1* | Euryarchaeota | NC_015847.1 | 2957 |
| 14 | *Methanosarcina acetivorans C2A* | Euryarchaeota | NC_003552.1 | 2999 |
| 15 | *Pyrolobus fumarii 1A* | Crenarchaeota | NC_015931.1 | 3116 |
| 16 | *Desulfurococcus mucosus DSM 2162* | Crenarchaeota | NC_014961.1 | 3080 |
| 17 | *Sulfolobus solfataricus 98/2* | Crenarchaeota | NC_017274.1 | 3038 |
| 18 | *Pyrobaculum neutrophilum V24Sta* | Crenarchaeota | NC_010525.1 | 3647 |
| 19 | *Korarchaeum cryptofilum OPF8* | Korarchaeota | NC_010482.1 | 3527 |
| 20 | *Nitrososphaera gargensis Ga9.2* | Thaumarchaeota | NC_018719.1 | 2963 |

Multiple alignment of homologous proteins from various prokaryotes with recognized sequence fragments was done with ClustalX [16].

# RESULTS

## Establishing the structural variations in the fragment of the V domain of 23S rRNA from prokaryotes of different taxons

The comparison of the 23S rRNA genes from different prokaryotes testified a single nucleotide insertion in stem H88 of *Clostridium saccharolyticum*, and another insertion in *Ehrlichia ruminantium* downstream of stem H73.

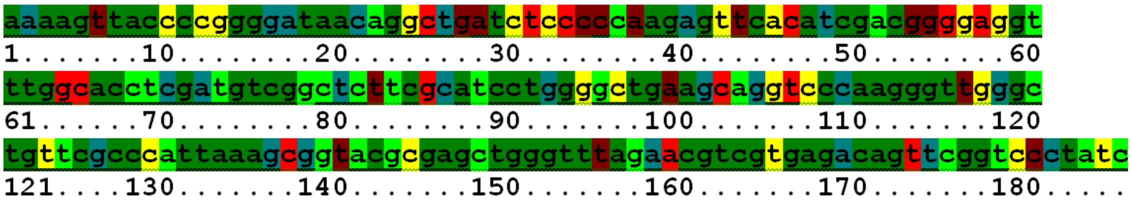
Thus, only one fragment 211 nt long corresponding to positions C2416-C2626 on the map of the *E.coli* 23S rRNA gene, remained structurally invariable within the set of the selected microorganisms. Interestingly, this longest structurally invariant region of the 23S rRNA includes 4 out of the 5 segments of the peptidyl transferase ring, PTR. It also contains 9 out of the 11 modified bases of the PTC. The maximal possible size of such hypothetical peptide would be 70 amino acid residues.

## Establishing the nucleotide variability of the structurally stable fragment of 23S rRNA

Comparison of 23S rRNA sequences from various microorganisms allowed establishing the sites with different variability. As expected, the lowest variability in the nucleotide composition was registered at sites enclosing the PTR segments, while the highest variability was noted at sites most distant from the PTR (regions of stems H89, H91 and H73). The variability of the flanking sites of the chosen sequence (upstream of the 3’-H74 chain and inside the 3’-H73 chain) was so high, that we paid no regard to determination of their consensus nucleotides and the corresponding amino acids. Thus, the 23S rRNA fragment subject to hypothetical translation, decreased to the length of 185 nucleotides (corresponding to A2432-C2616 on the *E.coli* 23S rRNA gene site), with a maximal length of estimated peptide product of 62 amino acid residues.

Of course, the revealed variations in the RNA sequence might lead to differing results of translation, and each possible variant can be favored by different arguments. The proximity of a source microorganism deemed to be a source of rRNA to the phylogenetic trunk, according to the scheme by T. Cavalier-Smith [10], was applied as a basic parameter for establishing the final version of the sequence regarding the hypervariable positions.

The final consensus sequence of the rRNA fragment appeared as follows (Fig. 1).



```
aaaagttaccccggggataacaggctgatctcccccaagagttcacatcgacggggaggt
1.......10.........20.........30.........40.........50.........60
ttggcacctcgatgtcggctcttcgcatcctggggctgaagcaggtcccaagggttgggc
61......70.........80.........90........100........110........120
tgttcgcccattaaagcggtacgcgagctgggtttagaacgtcgtgagacagttcggtccctatc
121....130........140........150........160........170........180.....
```

**Fig. 1**. Consensus sequence of the 23S rRNA fragment selected as a template for hypothetical translation. Highlighted by green color – invariant nucleotides (20 out of 20 in the set), cyan – having 1 substitution, light green – 2-4 substitutions, yellow – 5-7 substitutions, red – 8-9 substitutions, dark red – 10 and more substitutions in comparison to the prevailing nucleotide.

## Translation from hypothetical reading frames of the rRNA fragment

Analyzing the potential outcome of the hypothetical translation (using the standard code) from the three possible reading frames (RF), we received the following variants of amino acid sequences (Fig. 2):

```
RF-1:    kgtpgitg-spprvhidgevwhldvgsshpgaevgpkgwavrplkryaswv-nvvrqfgpyp
RF-2:    krysgdnrlispkssyrrgglaprcrlfaswg-srsqglgcspikavrelglerretvrsls
RF-3:    -kvlqg-qadlpqefistgrfgtsmsahrilglk-vprvglfah-sgtragfrts-dssvpi
```

**Fig. 2**. Results of hypothetical translation from the three potential reading frames (RF) of a fragment of the V domain of 23S rRNA. Stop codons and "evolutionarily late" amino acids (W, Y, F, H) are highlighted by red color. Underlined are the products of translation from the invariant sites of 23S rRNA.

As seen from Fig. 2, only 3 products (all three originate from the same RF) had an uninterrupted size of more than 9 amino acid residues (12, 11 and 29, respectively):

1. sgdnrlispkss
2. rrgglaprcrl
3. srsqglgcspikavrelglerretvrsls.

Obviously, we chose the longest sequence, rejecting the first 4 amino acids which are a product of hypothetical translation from the hypervariable 23S rRNA site, to increase the specificity of search. Finally, we chose the sequence GLGCSPIKAVRELGLERRETVRSLS 25 amino acids long for studying. This sequence is a product of hypothetical translation corresponding to G2543-C2616 (74 nt) on the *E.coli* 23S rRNA gene map and enclosing stems H91, H92, H90, H93, H73 and two segments of the PTR.

## Search for similarity of the most extended amino acid sequence GLGCSPIKAVRELGLERRETVRSLS with prokaryotic proteins

BLASTp was used to search for the GLGCSPIKAVRELGLERRETVRSLS amino acid sequence in proteins of various microorganisms, primarily in:

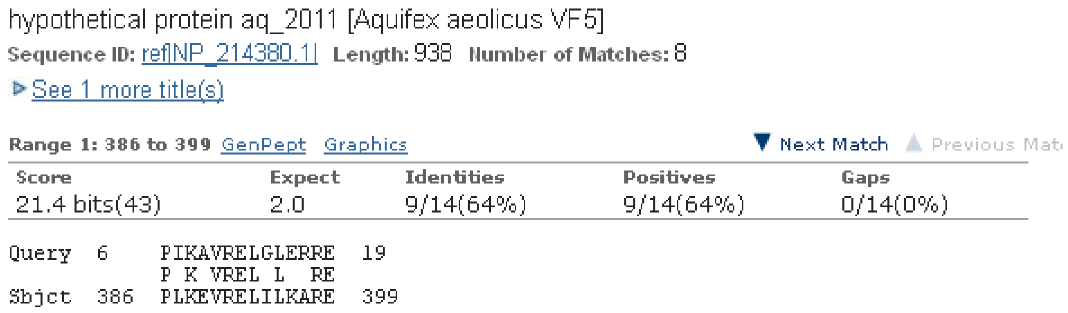
1. *Synechococcus spp.,*
2. *Aquifex aeolicus,*
3. *Thermus thermophilus,*
4. *Chlamydophila pneumoniae,*
5. *Treponema pallidum,*
6. *Escherichia coli,*
7. *Mycoplasma pneumoniae,*
8. *Bacillus subtilis,*
9. *Corynebacterium spp.,*
10. *Haloarcula marismortui,*
11. *Pyrobaculum neutrophilum,*
12. *Nitrososphaera gargensis.*

As we expected, the complete 25 amino acid sequence was not located in any protein in its unchanged form; only fragments of different length and structural integrity were found.

Sequence fragments found in products defined as "hypothetical protein", "probable conserved protein", "unnamed protein product", "putative …", "reading frame …" etc (Fig. 3.) were excluded from the analysis.

Proteins falling into the following groups were also excluded from the analysis:

- transcriptional regulators,
- tRNA modification enzymes,
- transposases,
- CRISPR-associated proteins,
- ABC transporters,
- RNases,
- proteases.



hypothetical protein aq_2011 [Aquifex aeolicus VF5]

Sequence ID: ref|NP_214380.1| Length: 938 Number of Matches: 8

▶ See 1 more title(s)

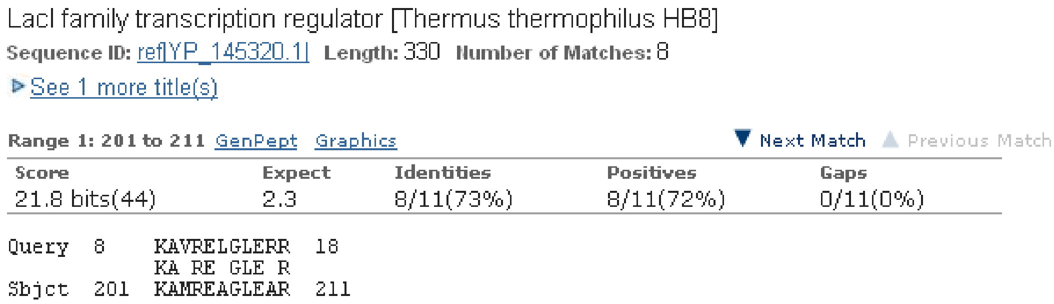
Range 1: 386 to 399 GenPept Graphics ▼ Next Match ▲ Previous Mat

| Score | Expect | Identities | Positives | Gaps |
|---|---|---|---|---|
| 21.4 bits(43) | 2.0 | 9/14(64%) | 9/14(64%) | 0/14(0%) |

```
Query  6    PIKAVRELGLERRE  19
            P K VREL L  RE
Sbjct  386  PLKEVRELILKARE  399
```

**Fig. 3**. Example of partial recognition of the target amino acid sequence in the primary structure of a hypothetical protein in bacterium *Aquifex aeolicus*.

Though in these proteins we found sequences having high homology with the target sequence (Fig. 4), an adequate comparison of these multiple representatives from various protein families is a task for further studies.



LacI family transcription regulator [Thermus thermophilus HB8]

Sequence ID: ref|YP_145320.1| Length: 330 Number of Matches: 8

▶ See 1 more title(s)

Range 1: 201 to 211 GenPept Graphics ▼ Next Match ▲ Previous Match

| Score | Expect | Identities | Positives | Gaps |
|---|---|---|---|---|
| 21.8 bits(44) | 2.3 | 8/11(73%) | 8/11(72%) | 0/11(0%) |

```
Query  8    KAVRELGLERR  18
            KA RE GLE R
Sbjct  201  KAMREAGLEAR  211
```

**Fig. 4**. Example of partial recognition of the target amino acid sequence in the primary structure of a transcription regulator of bacterium *Thermus thermophilus*.

Only proteins meeting the following criteria were selected:

- occurrence of at least 5 invariant amino acid residues from the target 25 amino acid sequence;
- lack of deletions and insertions;
- detailed information about the protein which excludes any identification error and allows for a comparative analysis with homologs and unrelated proteins.

Upon revealing such proteins and making a set, the presence of the target sequence in the groups of homologous proteins was estimated. Frequently, the sequence fragment was found in a particular protein of a certain microorganism, but was virtually absent in the equivalent loci of protein homologues from other microorganisms. We regarded such result either as a randomly developed combination of amino acids due to spontaneous mutational and recombination events, or as a rapid loss of the initial sequence in a large number of other phylogenetically distant (and, perhaps, evolutionarily more recent) taxa. In any case, such fragment of primary protein sequence was not included in the subsequent analysis.

We established the consensus sequence of amino acids characteristic for the analyzed protein site by multiple alignment (Table 2, Table 3). The fragment was selected for further analysis if it was found in a similar locus of a protein in at least five prokaryotes from different taxa – typically, in two groups of gram-negative bacteria (including Proteobacteria), in Firmicutes, Actinobacteria and Archea. When there was no similarity between the bacterial and archaeal proteins, the fragment was picked out if found in similar loci of no less than five bacteria from various taxa, and such instances were adequately noted.

A list was made, containing protein fragments with the highest contents of amino acid residues identical and homologous to the target sequence. Due to the large amount of data, we considered it rational to select only the loci which contain amino acids coinciding with the main part of the sequence composed of 15 amino acid residues: SPIKAVRELGLERRE.

**Table 2.** Example of partial detection of the target amino acid sequence in the primary structure of phosphoenolpyruvate synthase (PEPs) in a set of phylogenetically distant prokaryotes. Amino acids in protein sequences coinciding with the target sequence are in red

| Microorganism | Protein ID | Protein length | Position number | Fragment of alignment |
|---|---|---|---|---|
| *Synechococcus sp. PCC 7502* | AFY72263.1 | 813 | 435 | GRTCHAAIIARELGIPAIVGCGDAS |
| *Aquifex aeolicus VF5* | NP_214468.1 | 856 | 427 | GRTSHAAIVARELGIPAVVGTGNAT |
| *Mucilaginibacter paludis DSM18603* | EHQ24539.1 | 819 | 429 | GRTCHAAIVARELGVPAIVGCGNAT |
| *Escherichia coli str. K-12 W3110* | BAA15471.1 | 792 | 417 | GRTCHAAIIARELGIPAVVGCGDAT |
| *Lactobacillus oryzae JCM 18671* | GAK48720.1 | 800 | 393 | GMTCHAAIVSREMQIPCIVGTKSQH |
| *Corynebacterium glycinophilum* | AHW65035.1 | 812 | 425 | GRTCHAAIIARELGIPAIVGTGDAT |
| *Haloarcula marismortui* | YP_136393.1 | 786 | 423 | GMTSHAAIVSRELGVPAVVGAEDAT |
| *Pyrococcus yayanosii CH1* | YP_004624420.1 | 791 | 412 | GRTSHAAIVSRELGIPCVVGTKVAT |
| Consensus | | | | G  C        RELG+          + |
| RF-2 | | | | GLGCSPIKAVRELGLERRETVRSLS |

**Table 3.** Example of partial detection of the target amino acid sequence in the primary structure of alanyl-tRNA synthetase (alaRS) in a set of phylogenetically distant prokaryotes. Amino acids in protein sequences coinciding with the target sequence are in red

| Microorganism | Protein ID | Protein length | Position number | Fragment of alignment |
|---|---|---|---|---|
| *Synechococcus PCC 6301* | BAD78854.1 | 892 | 231 | LTALEKQNIDTGMGLERMAQVLQGV |
| *Aquifex aeolicus VF5* | NP_213887.1 | 867 | 210 | LTPLPHPNIDTGMGLERIASVLQGK |
| *Deinococcus proteolyticus* | ADY25918.1 | 892 | 244 | LAPLPFKNIDTGMGLERVASVVQDV |
| *Ehrlichia ruminantium* | CAH57866.1 | 887 | 221 | LSVLPRKCIDTGMGLERIAAVMQGV |
| *Escherichia coli str. K-12 W3110* | BAA16559.1 | 876 | 228 | MEPLPKPSVDTGMGLERIAAVLQHV |
| *Clostridium perfringens F262* | EIA16630.1 | 879 | 216 | YNELAQKNIDTGMGLERIATIMQGV |
| *Streptomyces coelicolor A3(2)* | NP_625781.1 | 890 | 219 | LGELPSKNIDTGLGLERLAMILQGV |
| *Haloarcula marismortui* | WP_011224298.1 | 927 | 246 | YSPMDTYIVDTGYGLERWTWMSQGT |
| Consensus | | | | L   P K +      GLER    V |
| RF-2 | | | | LGCSPIKAV-RELGLERRETVRSLS |

In the result of primary analysis we revealed and picked out 20 protein fragments containing sequence elements, which presumably derived from the same ancestral sequence translated from 23S rRNA (Table 4).

Notably, only 16 proteins were the sources of these fragments, i.e. some proteins contained several fragments.

Significant similarity was found between the analyzed peptide and the selected proteins, all of which can be described as important, universal and evolutionarily ancient for all cellular forms of life: 4 aminoacyl-tRNA synthetases, class II (phenylalanyl (β-subunit – PheRS_B), glycyl (α-subunit – GlyRS_A), alanyl (AlaRS), histidyl (HisRS)), 2 aminoacyl-tRNA synthetases, class I (methionyl (MetRS), glutamyl (GluRS)), 3 ribosomal proteins (L7/L12, L18, L30), DNA-polymerase III (α-subunit – DnaE), NAD-dependent DNA ligase A (LigA) and 5 metabolic enzymes (carbamoyl phosphate synthase (large subunit – CPS_L), phosphoenolpyruvate synthase (PEPs), acetyl-CoA carboxylase (subunit A, biotin carboxylase – Acc_C), dihydrolipoamide dehydrogenase (DLD), heterodisulfide reductase subunit B (HdrB)).

**Table 4.** List of protein loci with elements of sequence presumably derived from the same ancestral translation product from 23S rRNA*

| # | Protein | Protein length | Start position | Consensus sequence of fragment |
|---|---|---|---|---|
| 1 | CPS_L | 1028-1088 | 390-397 | **S**L.**KA**L**R**.**L**...**R**.. |
| 2 | RpL7/12* | 121-131 | 66-73 | ..**IKAVRE**...LGL. |
| 3 | RpL18* | 116-136 | 94-118 | ..A.**A**A**RE**A**GLE**... |
| 4 | PEPs | 786-856 | 393-435 | ..A.VA**RELG**I.... |
| 5 | Acc_C | 444-600 | 13-24 | ..**I**R**A**.**RELGL**.... |
| 6 | DLD | 449-499 | 255-274 | ...P..**R**NI**GLE**K.. |
| 7 | PheRS_B | 584-836 | 114-125 | ..**I**..L.**ELGLE**.K. |
| 8 | MetRS | 497-717 | 53-60 | PI.VKA.**ELG**I..E. |
| 9 | dnaE* | 1128-1179 | 529-602 | ..M..**V**.**ELGL**.K.D |
| 10 | HdrB | 265-331 | 19-35 | .T..LLK**ELG**I**E**..**E** |
| 11 | HisRS | 403-466 | 144-154 | ....IL**R**.**LGL**.... |
| 12 | GluRS* | 468-491 | 54-62 | ..L..LK.**LGL**D..**E** |
| 13 | GlyRS_A* | 285-303 | 88-95 | ..L..L**R**.**LG**ID..**E** |
| 14 | GlyRS_A* | 285-303 | 33-42 | .**P**A..L**R**.**LG**.**E**... |
| 15 | GlyRS_A* | 285-303 | 148-157 | .**P**V.V...Y**GLER**.. |
| 16 | RpL30* | 59-62 | 10-12 | ..K.**AV**K.**LGL**.**R**.. |
| 17 | HisRS* | 403-466 | 293-331 | .TPAV...I**GLER**.. |
| 18 | AlaRS | 867-927 | 40-61 | ..PLK...**LGLE**... |
| 19 | AlaRS | 867-927 | 210-246 | P.K.I...M**GLER**.. |
| 20 | LigA | 647-691 | 475-489 | ..L.....**LGLER**.. |
|  | RF-2 |  |  | SPIKAVRELGLERRE |

* Asterisks denote proteins selected only in bacteria, but not in archaea. Amino acids coinciding in the majority of protein homologue sequences and the target sequence are highlighted by red color. Non-homologous amino acids are marked with dots. Yellow color outlines possible insertions in the "ancestral" sequence, while possible translocated fragments of the "ancestral" sequence are highlighted by green color.

As it is seen from Table 4, the selected fragments most often contained the amino acids from the middle part of the "ancestral" sequence, primarily from the 11 amino acid peptide **IKAVRELGLER**.

To sum up, the result of our research is the *identification of a certain amino acid sequence which is partially preserved in the same sites of structurally and functionally differing proteins in phylogenetically distant microorganisms*.

## DISCUSSION

Apparently, the obtained results might have ambiguous interpretation concerning both the argumentation for the fragmentary origin of genes of the selected proteins from an unspecified m/rRNA ancestor and the question of limiting the reading frame search with only one relatively short amino acid sequence.

First of all, it should be noted that the defined protein fragments are possibly not phylogenetically related descendants of the ancestral amino acid sequence. It is very likely that the motifs (A/V)RE(L/A)GL identified in the ribosomal protein L18, phosphoenolpyruvate synthase and biotin carboxylase have various origin, being products of convergent evolution. A similar logic is applicable for motifs ELG(L/I) and (L/I/M)GLER. It is particularly important, considering that arginine and leucine encoded by six triplets occur twice in the motifs. To clarify this observation it is necessary to analyze, at least, the nucleotide sequences of gene fragments for the selected proteins.

For instance, the different gene fragments of glutamyl-tRNA synthetase, glycyl-tRNA synthetase α-subunit (3 fragments), phenylalanyl-tRNA synthetase β-subunit, carbamoyl phosphate synthase, ribosomal protein L30, biotin carboxylase, DNA-polymerase III α-subunit and subunit B of heterodisulfide reductase from *Aquifex aeolicus* possess a certain similarity. Multiple mutual nucleotide coincidences were found even in case of single nucleotide mismatches with the target sequence deriving from the 23 rRNA (Fig. 5).

```
Translation:                 /S//P//I//K//A//V//R//E//L//G//L//E//R//R/
LURNA                    124 tcgcccattaaagcggtacgcGAGCTGGGTTTAGAAcgtcgt
GluRS [A.aeolicus]       169 atgctgatagaggatttaaagtggctcggtatagattgggac
GlyRS_A [A.aeolicus] 271 atttaccttgaaagccttgaaagactaggcataaatcccctg
GlyRS_A [A.aeolicus] 103 aaccctgccactttcctgaaagttcttggtaaaaagccgtgg
GlyRS_A [A.aeolicus] 451 gacgagatatccgttgagatcacttacggactggaaaggata
PheRS_B [A.aeolicus] 346 ggactccttctctccgctcaggaactcggacttgaagagaaa
CPS_L [A.aeolicus]      1176 agcactcctaaaggctgtaaggagtttagaactcgacaggta
RpL30 [A.aeolicus]        52 agacatatccaggcggtaaagtccttaggtcttaaaaaaaga
Acc_C [A.aeolicus]        46 agggctataagggcgtgtagggaactgggattgaaaactgtg
LigA [A.aeolicus]       1732 tacgacatggttcagctcgaagaactcggtctcctgaagatg
HdrB [A.aeolicus]        108 tccaccagaatagtagccaaagaacttggtcttgaacttgac
Consensus                    a.c..yat.r..g...t.aaggaactyggtct.raa..gr.r
Translation:                 /.//.//I//.//.//.//K//E//L//G//L//E//.//./
```

**Fig. 5**. Multiple alignment of 10 fragments from 8 genes of different proteins from *Aquifex aeolicus*. Nucleotides coinciding with the sequence, presumably derived from 23S rRNA, are highlighted by red color. Blue color marks nucleotides coinciding in the majority of aligned fragments, but not coinciding with the target sequence.

The obtained results allow to define the identified amino acid sequence IxhhxELGLE (or (I/L)xhh(K/R)ELG(L/I)E), where "h" denotes hydrophobic amino acids) as a universal ancestral peptide motif. This motif presumably belongs to a group of yet unspecified universal peptide motifs consolidated by a common origin of their encoding sequence from a single ancestral RNA, which is homologous to a modern ribosomal RNA.

Further detailed and large-scale comparative studies are required for nucleotide sequences from fragments of genes encoding homologous proteins in other microorganisms. Identification of a microorganism with the largest number of unchanged forms of these universal peptide motifs in the basic proteins with early functions can become an interesting outcome of such search, giving an argument in favor of its highest relatedness to the Last Universal Cellular Ancestor – LUCA [17]. Our preliminary data point at the *Aquifex* bacteria as probable candidates for this role.

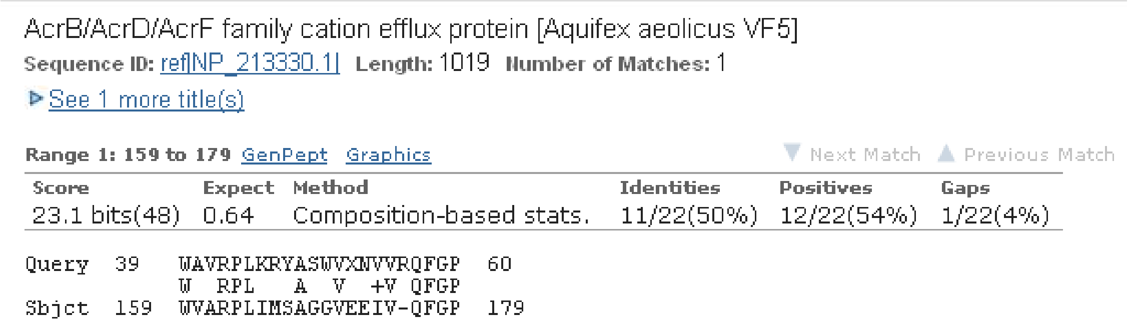
Such motifs, having a diverse amino acid composition, but a single genetic source, may be defined as a group of last common ancestors of template-coded peptides (Last Universal Peptide Ancestors – LUPAs), which lay the structural basis for the origination and evolution of the modern protein diversity. Perhaps, the elucidated motif will not appear to be the most widespread. Still, the importance of the obtained results was to demonstrate a possibility of existence of hypothetically m/rRNA-encoded peptide motif in the heterological proteins from phylogenetically very different microorganisms.

Our approach allows partial creation of a universal phylogenetic "protein bush", a structure having not a single, but several trunks, due to production of at least 6 (taking into account the complementary RNA chain) variants of amino acid sequences from a single RNA template. The search for amino acid sequences obtained as a result of hypothetical translation from the RNA strand, which is complementary to the rRNA [5], appears to be a productive approach, considering that simultaneous performance of the template and ribozyme functions required simultaneous presence of several identical RNA molecules apparently synthesized from the initial complementary strand.

It is necessary to take into account that genes of many modern proteins can descend from ancient RNA-genes, which originated earlier than the studied m/rRNA and coexisted with it. Therefore, the rRNA fragment that we chose should be regarded not as the first, but rather as the last common predecessor for a part (perhaps, very considerable or even prevailing) of mRNAs which are ancestral for modern protein-coding genes. Moreover, it can turn out that the chosen rRNA fragment, which we consider as the most probable candidate for the role of the last universal mRNA-ancestor (LURNA), might happen to be not the only one.

It should be noted that the estimated sequence is variable, i.e. it presumes certain variations in the nucleotide composition, because of the impossibility to determine the exact nucleotide sequence of the initial rRNA fragment. A search using other reading frames with removal of restrictions for the presence of "late" amino acids (which allowed to considerably increase the length of the target sequences) delivered interesting results to confirm this assumption (Figs. 6-8). However, we have not verified the presence of these sequences in homologous proteins of other microorganisms (with a view to avoid accidental matches).

Attention should be paid also to the neglected smaller fragments of rRNA which take part in the formation of the peptidyl transferase center. The highly conservative fragment containing the 3'-chains of stems H74 and H89 not included into the study, as well as the segment of the PTR directly involved in the transpeptidation stage, are particularly interesting (especially since the latter site is considered to be the most ancient element of the 23S rRNA [18]).



AcrB/AcrD/AcrF family cation efflux protein [Aquifex aeolicus VF5]

Sequence ID: ref|NP_213330.1| Length: 1019 Number of Matches: 1

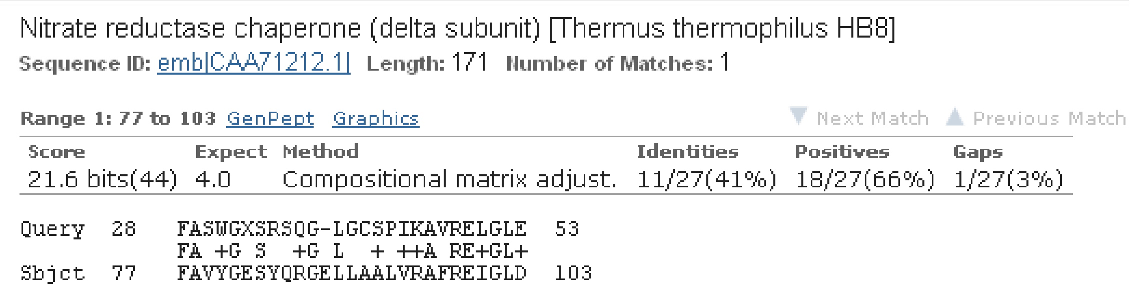
See 1 more title(s)

Range 1: 159 to 179 GenPept Graphics

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 23.1 bits(48) | 0.64 | Composition-based stats. | 11/22(50%) | 12/22(54%) | 1/22(4%) |

```
Query  39   WAVRPLKRYASWVXNVVRQFGP  60
            W  RPL   A  V  +V QFGP
Sbjct  159  WVARPLIMSAGGVEEIV-QFGP  179
```

**Fig. 6**. Example of partial detection of the target amino acid sequence, the product of translation from reading frame RF-1 in the primary structure of a protein of *Aquifex aeolicus*.



Nitrate reductase chaperone (delta subunit) [Thermus thermophilus HB8]

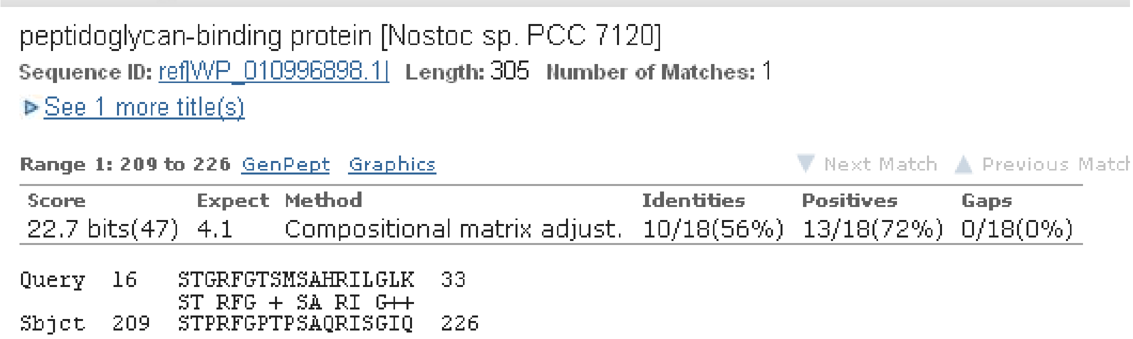
Sequence ID: emb|CAA71212.1| Length: 171 Number of Matches: 1

Range 1: 77 to 103 GenPept Graphics

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 21.6 bits(44) | 4.0 | Compositional matrix adjust. | 11/27(41%) | 18/27(66%) | 1/27(3%) |

```
Query  28   FASWGXSRSQG-LGCSPIKAVRELGLE  53
            FA +G S  +G L  + ++A RE+GL+
Sbjct  77   FAVYGESYQRGELLAALVRAFREIGLD  103
```

**Fig. 7**. Example of partial detection of the target amino acid sequence, the product of translation from reading frame RF-2 in the primary structure of a protein of *Thermus thermophilus*.



peptidoglycan-binding protein [Nostoc sp. PCC 7120]
Sequence ID: ref|WP_010996898.1| Length: 305 Number of Matches: 1
▶ See 1 more title(s)

Range 1: 209 to 226 GenPept Graphics ▼ Next Match ▲ Previous Match

| Score | Expect | Method | Identities | Positives | Gaps |
|---|---|---|---|---|---|
| 22.7 bits(47) | 4.1 | Compositional matrix adjust. | 10/18(56%) | 13/18(72%) | 0/18(0%) |

```
Query  16   STGRFGTSMSAHRILGLK  33
            ST RFG + SA RI G++
Sbjct  209  STPRFGPTPSAQRISGIQ  226
```

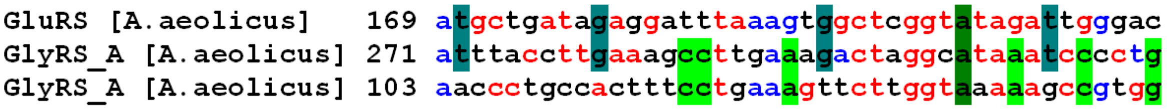
**Fig. 8**. Example of partial detection of the target amino acid sequence, the product of translation from reading frame RF-3 in the primary structure of a protein of cyanobacterium *Nostoc*.

Remarkably, less conservative sites of the 23S rRNA are perhaps the derivatives of one or two initial rather small rRNA molecules (roughly corresponding to the A- and P-sites), which formed the early structure of the PTC [7, 8, 19]. Such view on the origin of large molecules of modern ribosomal 23S RNAs suggests the lack of necessity for consideration of other sites in 23S rRNA as potential candidates for the role of LURNA.

Concerning the 16S rRNA molecule, it can be assumed that the consideration of the 3’-domain fragments containing the most conservative stems H44 and H45 may bring interesting results [20]. However, a higher structural variability of this region (not mentioning the variability of sequence) in different microorganisms in comparison to the PTC-containing fragments of the 23S rRNA, will be an evident problem.

The spectrum of proteins containing the elements of our target sequence is a strong argument in favor of the relict nature of this motif. It was found in several ribosomal proteins (in the most conserved form) and in enzymes promoting the complication of polynucleotide strands: DNA-polymerase III α-subunit and NAD-dependent DNA ligase A. But the highest distribution of this motif was registered in the enzymes playing a major role in genetic code realization in the modern translation apparatus – aminoacyl-tRNA synthetases of both classes. Multiple presence of the target sequence elements in glycyl-tRNA synthetase α-subunit (provided that two of the three fragments are homologous), as well as expressed similarity of one of these fragments with a site in glutamyl-tRNA synthetase of a different class, appear to be especially interesting. Comparative analysis of gene fragments encoding these sites reveals prominent similarity allowing to make an assumption within our hypothesis concerning the origin of the ancestral parts of these protein genes from a single nucleotide sequence (Fig. 9).

The spectrum of proteins containing the determined motif and carrying out metabolic functions is also of specific interest. Three of them (biotin carboxylase, phosphoenolpyruvate synthase and dihydrolipoamide dehydrogenase) play key roles in the metabolism of carbohydrates involving acetyl-CoA. The presence of the motif in heterodisulfide reductase is also important, in respect of the role of metal sulfides at the earliest stages of biogenesis [21, 22].



```
GluRS  [A.aeolicus]    169 atgctgatagaggatttaaagtggctcggtatagattgggac
GlyRS_A [A.aeolicus]   271 atttaccttgaaagccttgaaagactaggcataaatccctg
GlyRS_A [A.aeolicus]   103 aaccctgccactttcctgaaagttcttggtaaaaagccgtgg
```

**Fig. 9**. Alignment of a fragment of the glutamyl-tRNA synthetase gene and two fragments of the gene encoding glycyl-tRNA synthetase α-subunit of *Aquifex aeolicus*. Nucleotides matching the sequence presumably derived from 23S rRNA are in red. Nucleotides matching between two of the three loci are highlighted by blue-green and light green. Deep green highlights the nucleotide matching between all the three loci.

The presence of the motif in the sequence of carbamoyl phosphate synthase is very indicative. The product of this enzyme carbamoyl phosphate is a key intermediate in the ornithine cycle and is essentially required for the biosynthesis of pyrimidine nucleotides. Carbamoyl phosphate is a phosphorilated derivative of formamide, which is considered as an essential molecule in the origin of metabolism and the systems of genetic heredity [23, 24].

Despite the presence of the motif in a number of proteins realizing such diverse functions, its functional role remains vague. The preliminary analysis could not detect its presence in the loci of proteins with specific activity. Probably, it was due to the chemical nature of the motif itself, which contains many hydrophobic amino acid residues. However, there are evidences speaking for an increase in the activity of many ribozymes in the presence of hydrophobic components [25]. It is quite possible that the involvement of one or several such peptides in the early stages of evolution of the "maternal" proto-ribosome promoted an increase of its functionality and predetermined their conservation in the structure of the ribosomal proteins. Interestingly, one of the possible scenarios of the origin of the translation system attributes a key role to a hypothetical peptide serving as some kind of a cofactor for the ribozyme ancestor of the ribosome. It is expected that such peptide has a non-specific capability to stimulate and/or stabilize the ribozymes, carries a pair of negatively charged amino acids and forms a complex with divalent cations [1]. Localization of the two residues of glutamic acid in the determined motif meets this requirement.

We believe, that the primary function (more precisely, property) of the initial short peptide was the structural function. Further on, the increase of the initial template in length, most possibly due to various modifications (first of all, duplication) of the initial gene, promoted this peptide to act as a specific universal basis, "core", "backbone" or "handle" for a set of many other, more specialized peptide domains, parts of a novel longer molecule. Such "skeletal" role of this presumed ancestral peptide could become a basis for its conservation in the structure of wide spectrum of proteins with the most diverse functions.

## CONCLUSIONS

The results of our research allowed making two main conclusions.

Different sites of primary structure of many conservative proteins, even those structurally and functionally unrelated, consistently display similarity with the elements of the amino acid sequence IKAVRELGLER, which apparently originated from an ancient peptide with unknown functions. This peptide can be ascribed to the group of Last Universal Peptide Ancestors, LUPAs.

This peptide is presumably a product of translation from one of the reading frames of the RNA molecule which contributed to the formation of the main part of the core (and probably the most ancient) functional site of the ribosome – the peptidyl transferase center, PTC. Fulfilling a *double ancestral function*: structural for the PTC, and template for a number of ancient peptides, this RNA sequence (containing subsequence ATTAAAGCGGTACGCGAGCTGGGTTTAGAACGT in its middle part) is perhaps the Last Universal RiboNucleic Ancestor – LURNA.

We are grateful to Nafisa Nazipova for perusal and critical reading of the manuscript and for her valuable comments. We would like to thank Konstantin Shavkunov for his thorough work on translation of the manuscript into English. This study was partially supported by RFBR (grant № 13-04-00991).

## REFERENCES

1. Koonin E.V. *The Logic of Chance: The Nature and Origin of Biological Evolution*. Upper Saddle River: FT Press, 2011. 528 p.

2. Caetano-Anolles G., Wang M., Caetano-Anolles D., Mittenthal J.E. The origin, evolution and structure of the protein world. *Biochem. J.* 2009. V.417. P. 621–637. doi: 10.1042/BJ20082063
3. Sobolevsky Y., Guimarães R.C., Trifonov E.N. Towards functional repertoire of the earliest proteins. *J. Biomol. Struct. Dyn.* 2013. V.31. № 11. P.1293–300. doi: 10.1080/07391102.2012.735623
4. de Farias S.T., do Rêgo T.G., José M.V. Evolution of transfer RNA and the origin of the translation system. *Front. Genet.* 2014. V. 5. P.303. doi: 10.3389/fgene.2014.00303
5. Root-Bernstein M., Root-Bernstein R. The ribosome as a missing link in the evolution of life. *Journal of Theoretical Biology*. 2015. V. 367. № 21. P. 130–158 doi: 10.1016/j.jtbi.2014.11.025
6. Greber B.J., Boehringer D., Leitner A., Bieri P., Voigts-Hoffmann F., Erzberger J.P., Leibundgut M., Aebersold R., Ban N. Reductive evolution of the mitochondrial 16S rRNA. *Nature*. 2014. V. 505. P. 515–519 doi: 10.1038/nature12890
7. Bokov K., Steinberg S.V. A hierarchical model for evolution of 23S ribosomal RNA. *Nature*. 2009. V. 457. P. 977-980. doi: 10.1038/nature07749
8. Hsiao Ch., Mohan S., Kalahar B.K., Williams L.D. Peeling the onion: ribosomes are ancient molecular fossils. *Mol. Biol. Evol.* 2009. V. 26 № 11. P. 2415-25. doi: 10.1093/molbev/msp163
9. Polacek N., Mankin AS. The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. *Crit. Rev. Biochem. Mol. Biol*. 2005. V. 40. № 5. P. 285–311. doi: 10.1080/10409230500326334
10. Cavalier-Smith T. Rooting the tree of life by transition analyses. *Biology Direct*. 2006. V. 1. P. 19. doi: 10.1186/1745-6150-1-19
11. Copley S.D., Smith E., Morowitz H.J. A mechanism for the association of amino acids with their codons and the origin of the genetic code. *Proc Natl Acad Sci USA*. 2005. V. 102. № 12. P. 4442–4447. doi: 10.1073/pnas.0501049102
12. Miller S.L., Urey H.C. Organic compound synthesis on the primitive Earth. *Science.* 1959. V. 3370. №130. P.245–251.
13. *GENBANK Entrez Nucleotide Database*. URL: http://www.ncbi.nlm.nih.gov/nucleotide/ (accessed 15.03.2015).
14. *Basic Local Alighnment Search Tool*. URL: http://blast.ncbi.nlm.nih.gov/Blast.cgi (accessed. 15.03.2015).
15. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. Basic local alignment search tool. *J. Mol. Biol.* 1990. V. 215. № 3. P. 403–410. doi: 10.1016/S0022-2836(05)80360-2
16. Jeanmougin F., Thompson J.D., Gouy M., Higgins D.G., Gibson T.J. Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 1998. V. 23. P. 403–405. doi: 10.1016/S0968-0004(98)01285-7
17. Woese C. The universal ancestor. *PNAS*. 1998. V. 95. №. 12. P. 6854–6859. doi: 10.1073/pnas.95.12.6854
18. Petrov A.S., Bernier C.R., Hsiao Ch., Norris A.M., Kovacs N.A., Waterbury C.C., Stepanov V.G., Harvey S.C., Fox G.E., Wartell R.M., Hud N.V., Williams L.D. Evolution of the ribosome at atomic resolution. *PNAS*. 2014. V. 111. № 28. P. 10251–10256. doi: 10.1073/pnas.1407205111
19. Tamura K. Ribosome evolution: Emergence of peptide synthesis machinery. *Journal of Biosciences*. 2011. V. 36. № 5. P. 921–928. doi: 10.1007/s12038-011-9158-2
20. Harish A., Caetano-Anollés G. Ribosomal History Reveals Origins of Modern Protein Synthesis. *PLoS One*. 2012. V.7. № 3. e32776. doi: 10.1371/journal.pone.0032776
21. Pascal R. Boiteau L., Forterre P., Gargaud M., Lazcano A., Lopez-Garcia P., Maurel M-C., Moreira D., Pereto J., Prieur D., Reisse J. Prebiotic Chemistry – Biochemistry -

Emergence of Life (4.4-2 Ga). *Earth, Moon and Planets*. 2007. V. 98. P. 153–203. doi: 10.1007/s11038-006-9089-3

22. Mulkidjanian A.Y., Galperin M.Y. On the origin of life in the zinc world. 2. Validation of the hypothesis on the photosynthesizing zinc sulfide edifices as cradles of life on *Earth. Biol. Direct.* 2009. V. 4. P. 27. doi: 10.1186/1745-6150-4-27
23. Saladino R., Botta G., Pino S., Costanzo G., Di Mauro E. Genetics first or metabolism first? The formamide clue. *Chem. Soc. Rev.* 2012. V. 41. № 16. P. 5526–5565. doi: 10.1039/c2cs35066a.
24. Pino S., Sponer J.E., Costanzo G., Saladino R., Di Mauro E. From Formamide to RNA, the Path Is Tenuous but Continuous. *Life*. 2015. V. 5. P. 372–384. doi: 10.3390/life5010372
25. Müller U.F., Bartel D.P. Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA*. 2008. V. 14. № 3. P. 552–562. doi: 10.1261/rna.494508

Received March 30, 2015.
Published April 08, 2015.