

Suggestion to Upgrade the Canonical Concept of Translation

Jan Charles Biro*

HOMULUS FND, Los Angeles, CA, USA

Abstract

Background: This research was carried out to provide a summary of a series of bioinformatical observations between 2000 and 2014 concerning the structure of nucleic acids and codons, the interaction between codons and amino-acids and the general concept of translation.

Methods: Public sequence and structure databases and established methods of bioinformatics resources were utilized during these studies.

Results: These studies provided novel insights into the canonical concept of translation and suggest that: 1) codons have structure and the development and function of the 1st, 2nd, 3rd codon residues is different; 2) codon boundaries are physicochemically defined; 3) there is a stereo-chemical compatibility (fitting) between codons and coded amino acids; 4) codon redundancy in the Genetic Code (synonymous codons) provides folding (3D) information to protein syntheses, that is additional to the information requested to determine the sequence of amino-acids (primary or 2D structure); 5) there is a Proteomic Code in the redundant Genetic Code; 6) mRNA-s assist the co-translational folding of their own coded peptides i.e. they function as nucleic acid chaperons; 7) tRNA-s assist even the transfer of folding information from mRNA to proteins (tRNA cycle) in addition to their traditional role as adaptor between codons and amino acids.

Conclusions: Computational (statistical) studies, molecular modeling and theoretical biological considerations suggest the possibility and necessity to upgrade the canonical concept of translation. Traditional laboratory confirmation is requested.

Keywords: Translation; Protein folding; Codon; Chaperon; RNA function

Background

The development of bioinformatics with the complementary computational technology and robust databases (sequences, structures) created an extremely effective and economical method to (re)address fundamental questions of life-sciences. This new approach has the potential to provide novel, creative ideas and even guidelines to the traditional laboratory sciences. One of the most promising areas became the (re)search on the concept of *translation* and finding theoretical solutions for old and conflicting problems.

Shortly, the traditional concept of translation suggests that mRNA is a roughly linear poly-nucleotide (no significant 3D structure) that passes rRNA-s (like magnetic tape passes the sensors of a tape-recorder). There the triple-residue codons in the mRNA are passively *adapted* by tRNA-s to the single-amino-acid residues of the coded protein. The output of the process is a linear poly-amino-acid that spontaneously folds into functional protein.

There are at least four “classical fighting fields” between molecular biologists which are related to the concept of translation: a) the question if direct and specific codon-amino acid interaction is possible; b) the question if the redundancy of the codons is accidental or meaningful; c) the question if all protein-folding information is present already in the sequence of amino acids; d) the question while is there a large discrepancy between the size and function of tRNA-s.

We are systematically addressing these fundamental questions since the premier publication of the *Proteomic-Code* hypotheses in 1981 [1] to explain the nature and origin of specific protein-protein interactions. The first model suggested the *perfect* complementary coding of interacting (co-locating) amino acids in the specifically interacting peptides. It was thoroughly tested in several laboratories with promising but not conclusive results [2]. The refined concept

of Proteomic Code, that suggested *partial* complementary coding of co-locating amino-acids, was released in 2006 [3]. It was followed by the recognition of the significance of Proteomic Code to determine/assist intra-molecular, residue-to-residue interactions, that is, the 3D structure of proteins. Consequently the possibility and even necessity of mRNA-assisted protein-folding (*nucleic acid chaperons*) became obvious for us in 2005 [4]. The existence of a Proteomic Code and the presence of folding (3D) information in nucleic acids (in addition to the information of amino acid sequence of the coded protein) provided a novel explanation for the redundancy of the Genetic Code. It led us in 2013 to the suggestion to complete the canonical Genetic Code [5].

This article is a compact review of approximately 15 years' work and numerous associated peer-reviewed publications. The intention is to provide a comprehensive view of a very large field, regarding the most recent developments which were obtained by the methods of computational and theoretical biology and suggest the upgrade of our canonical (and insufficient) concept of translation. Therefore, the reader is respectfully advised to consult the cited, original publications for details.

Results

Codons have structure

Two alternative hypotheses have been posed to explain the origin of

*Corresponding author: Jan Charles Biro, HOMULUS FND, Los Angeles, CA, USA, Tel: +1-213 627 6134; E-mail: Jan.biro@att.net

Received April 09, 2014; Accepted May 12, 2014; Published May 14, 2014

Citation: Biro JC (2014) Suggestion to Upgrade the Canonical Concept of Translation. J Proteomics Bioinform 7: 112-120. doi:10.4172/0974-276X.1000311

Copyright: © 2014 Biro JC. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

the genetic code. One hypothesis was championed by Woese [6], who argued that there was stereo-chemical matching, i.e. affinity, between amino acids and certain nucleic acid triplet sequences. He proposed that the genetic code developed in a way that was closely connected to the development of the amino acid repertoire, and that this close biochemical connection is fundamental to specific protein–nucleic acid interactions.

The alternative hypothesis was championed by Crick [7], who considered that the basis of the code could be a “frozen accident”, with no underlying chemical rationale. Crick, the first to suggest and promote the idea of an “adaptor” (transfer RNA; tRNA) between nucleic acids and proteins, refused any attempt to propose or model any direct codon-amino acid connection.

However, there is now very strong evidence for a “logical” connection between codons and amino acids, as demonstrated through the construction of *The Common Periodic Table of Codons and Nucleic Acids* [8]. This Table demonstrates the connection between the “RNA World” and “Protein World”, and clearly indicates that codons have structure (Figure 1).

The three nucleotides in codons have different functions that clearly distinguish them from one another. First, they have a preferred order

of reading that defines the 1st, 2nd and 3rd codon positions. The third “wobble” nucleotides have little importance in defining amino acids and several of them are interchangeable. *The Common Periodic Table of Codons and Nucleic Acids* reveals that the 2nd codon nucleotides have a preeminent role in defining the molecular structure and physico-chemical characteristics of the encoded amino acids [8].

Codon boundaries are physico-chemically defined

The selection of wobble bases is not random. Codon usage frequency tables clearly indicate that synonymous codons are not equally utilized (as would be expected if selection was random). The 1st and 3rd codon positions in exons (but not in introns) contain more G or C bases than the 2nd positions. There are three hydrogen bonds between C and G (dG=-1524 kcal/1000 bases) but only two between A and T (dG=-365 kcal/1000 bases). Consequently, the GC-rich 1st and 3rd codon residues contribute more to the thermodynamic force between complementary nucleic acid sequences (including the codon–anticodon interactions, folds, loops) than the 2nd codon residues. This is a statistically derived conclusion that is valid for large numbers of interactions and not for every interacting codon. This difference between thermodynamic potential of codon residues could be interpreted as a virtual, physico-chemical definition of codon boundaries. Such a definition is useful

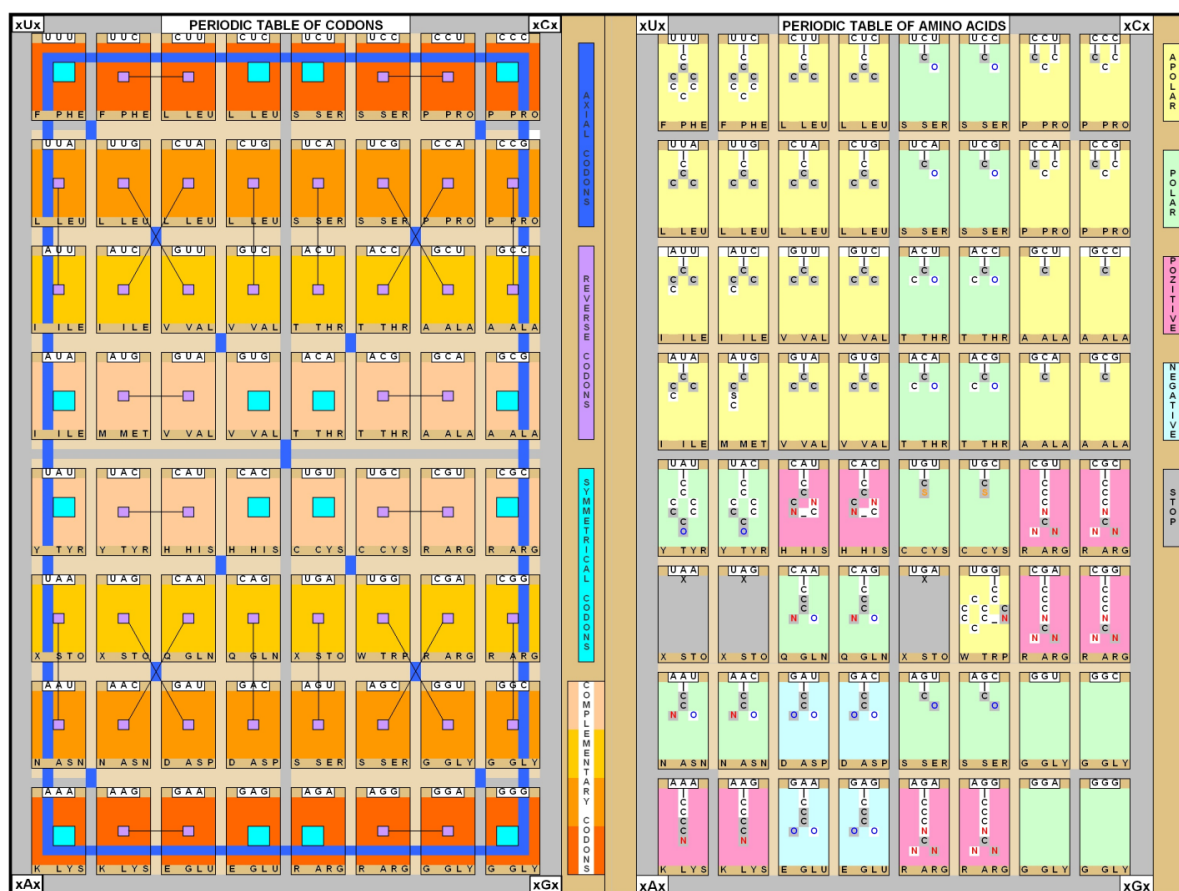


Figure 1: The Common Periodic Table of Codons and Amino Acids. Codons were sorted according to the order, symmetry and complementarity of their bases (left). The corresponding order of amino acids reveals periodicity of the physicochemical properties (polarity, charge, and molecular structure) of the encoded amino acids (right). Note that the periodic tables distinguish four separate fields, each corresponding to the four bases at the central codon positions. The frames of amino acid residues are rooted to the codons (boxes). The names of amino acids are indicated by one and three letters. The common structural features of amino acids in the same field are emphasized by letters (atoms) in gray background, that are C-C-C; C-; C-C-N(or O) and C-N(or O), corresponding to xUx, xCx, xAx, xGx codon-fields (for more details see ref [8]).

in terms of gaining a better understanding of how the correct codon reading is achieved and translation is protected against frame-shifts [9] (Figure 2).

Stereo-chemical “fitting” between codons and amino acids

The existence of connection between codons and the physico-chemical properties of encoded amino acids is well supported by *The Common Periodic Table of Codons and Nucleic Acids* [8]. This connection suggests co-evolution and possible specific spatial compatibility between codons and amino acids. This question was studied in relation to well-known examples of highly specific protein-nucleic acid interactions provided by restriction endonucleases (RE) and their nucleic acid cut sites (RS) [10]. Such studies confirmed that codons and encoded amino acids preferentially co-locate with one another in these structures, suggesting a stereo-chemical connection between nucleic acids and proteins (Figure 3).

Messenger RNA-s carry protein-folding (3D) information

According to research mRNA does not have any structure, as the recent concept of translation is stated as being analogous to a tape recording: mRNA freely passes through ribosomes, where it is read and translated into protein, on the codon to amino acid basis, with the help of tRNA-s as adaptors. Some mRNA structure is not necessary for this process, and some laboratories regard possible mRNA structures only as an unnecessary complication that slows down translation and reduces protein yield. However, thermodynamic studies oppose this view and determine that mRNA does have 3D structure [11,12]. During this study it was demonstrated that the FE (folding energy, dG) associated with coding sequences is significant and negative (-407 kcal/1000 bases, mean value), indicating that these sequences can form structures. However, the FE only has a small free component, less than 10% of the total. The contributions of the 1st and 3rd codon bases to the FE are larger than the contribution of the 2nd (central) bases. It is possible to achieve an approximately 4-fold change in free FE by altering the wobble bases in synonymous codons (Figure 4).

These observations suggest the importance (non-randomness) of wobble bases.

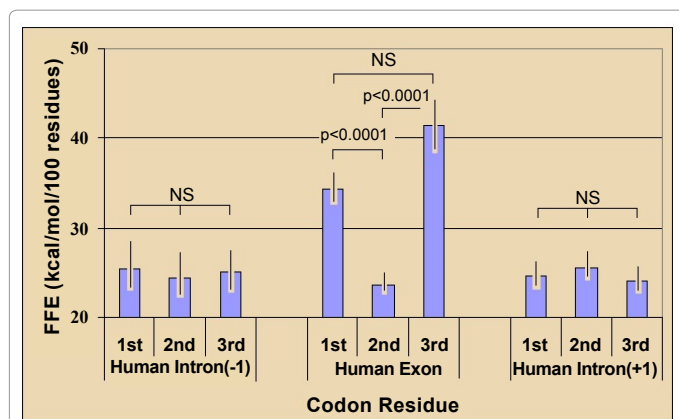


Figure 2: Free folding energies in different codon residues. Free folding energies (FFE) were determined in phase-selected sub-sequences of 81 genes (random selection from human sequence database). The original nucleic acids contained intact three-letter codons (1st+2nd+3rd). Sub-sequences were constructed by periodic removal of one letter from the codon and maintaining the other two (1st+2nd, 1st+3rd, 2nd+3rd), or removing two letters and maintaining only one (1st, 2nd, 3rd). Distinctions were made between exons and the preceding (-1) and following (+1) sequences (introns). The dG values were determined using *mfold* and the FFE was calculated. Each bar represents the mean ± SEM, n = 81 [9].

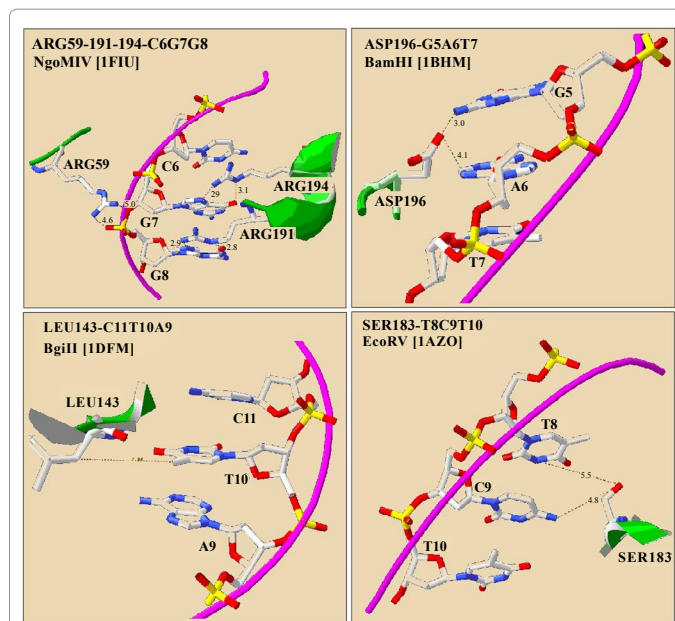


Figure 3: Examples of codon-amino acid co-locations from restriction enzyme (RE) and specific cut site (RS) complexes [10].

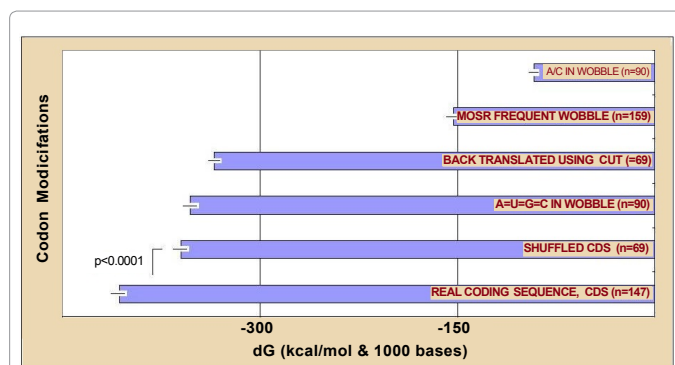


Figure 4: Effect of wobble bases on the dG of CDS. The TFE of mRNA is indicated in native sequences (CDS), after residue randomization (shuffle) and the indicated manipulation of the wobble bases (see the text for details). Each column represents the mean ± S.E.M.; n is indicated in the columns [12].

Proteins are assumed to contain all necessary information for unambiguous folding (*Anfinsen’s principle*, [13]). However, *ab initio* structure prediction is often unsuccessful, as the amino acid sequence itself is not sufficient to guide among endless folding possibilities, [14]. It seems logical to attempt to find the “missing” information in nucleic acids, specifically in redundant (synonymous) codons and their wobble bases.

Messenger-RNA energy dot plots (EDP) and protein residue contact maps (RCM) were comparable (Figure 5). The structure of mRNA is conserved if the protein structure is conserved, even if sequence similarity is low. These observations led us to propose that similarity may exist between nucleic acid and protein folding [15].

There is a “Proteomic Code”

Nucleic acids contain much excess information owing to codon redundancy. The paradigm insists that this excess information is used to provide protection (security backup) against mutations, i.e. alterations

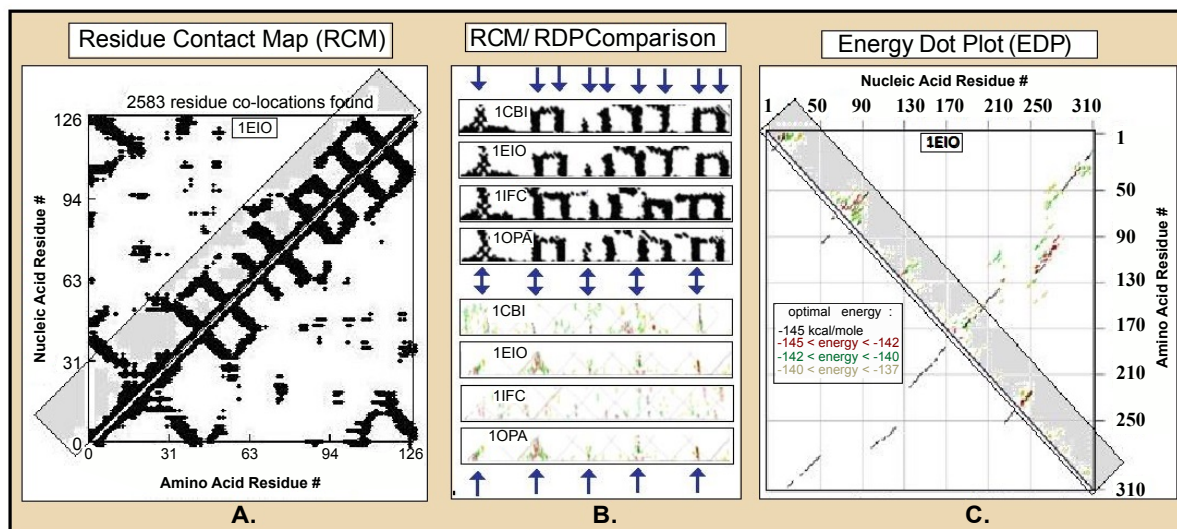


Figure 5: Comparison of protein and mRNA structures. 2D projections of proteins and corresponding mRNAs of four sequences were obtained using SeqX (RCM, A. [8]) and *mfold* tool (energy dot plots, C). The central, axial segments of these projections (grey areas) were compared (B). The sites of structural similarity are indicated (blue arrows) [15].

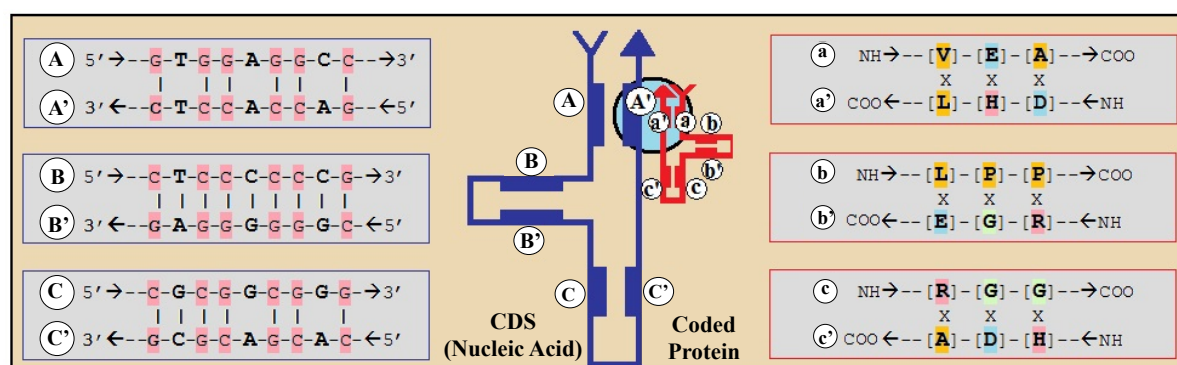


Figure 6: The Concepts of Proteomic Code and Nucleic Acid Assisted Protein Folding. The 3D structure of an encoded protein (red) is established and maintained by segments with specifically interacting domains that contain numerous amino acid co-locations (a-a', b-b', c-c'). Co-locating amino acids (X between their one letter names) are preferentially encoded by partially complementary codons, where the 1st and 3rd codon residues (pink letters connected by |) are complementary to one another (A-T or G-C) but the 2nd codon residues may be, but are not necessarily, complementary to each other. This rule is called the **PROTEOMIC CODE**. The complementary sites in nucleic acids define segments in the CDS (Nucleic Acid, blue, A-A', B-B', C-C'), which provide a 3D nucleic acid structure similar to the structure of the encoded protein. Codon amino acid interactions transfer the spatial information in CDS to proteins during translation. This process is called **NUCLEIC ACID ASSISTED PROTEIN FOLDING** [2,16,17,19,20].

in wobble bases should not affect the correct sequence of amino acids in encoded proteins. However, we might expect redundancy even of the 1st or 2nd codon residues of the same reason, but this is certainly not the case. It is especially strange if we consider that some proteins are not able to fold correctly, as the amino acid sequence alone is often insufficient to provide correct and sufficient protein folding information. Therefore, research concerning connections between nucleic acid and protein structures and interactions was initiated.

We demonstrated that *co-locating amino acids are preferentially encoded by partially complementary codons, where the 1st and 3rd codon residues are complementary to each other in reverse orientation, but the 2nd codon residues may but not necessarily do complement one another.* This connection between codon co-locations (partial complementarity) and amino acid co-locations (interactions) allows the possibility of transfer of spatial (folding) information from nucleic acids to proteins.

This is called the 'Proteomic Code' and is missing from the redundant, universal Genetic Code of Nirenberg [16].

In 1981 we proposed the idea that specifically interacting peptides are encoded by complementary codons [1]. This was the "first generation" of the Proteomic Code. Several scientists found the idea useful during the design and production of interacting peptides with specific high affinity (see Biro [2] for review). However, it became apparent that not all peptides encoded by complementary codons do interact with each other. Fortunately a modification of the original concept, where complementarity of the 2nd codon residues is permitted but not obligatory (the "second generation" of the Proteomic Code, [2,17]), solves this problem (Figure 6).

Messenger RNA-s are chaperons

The novel discovery that co-locating amino acids (in protein

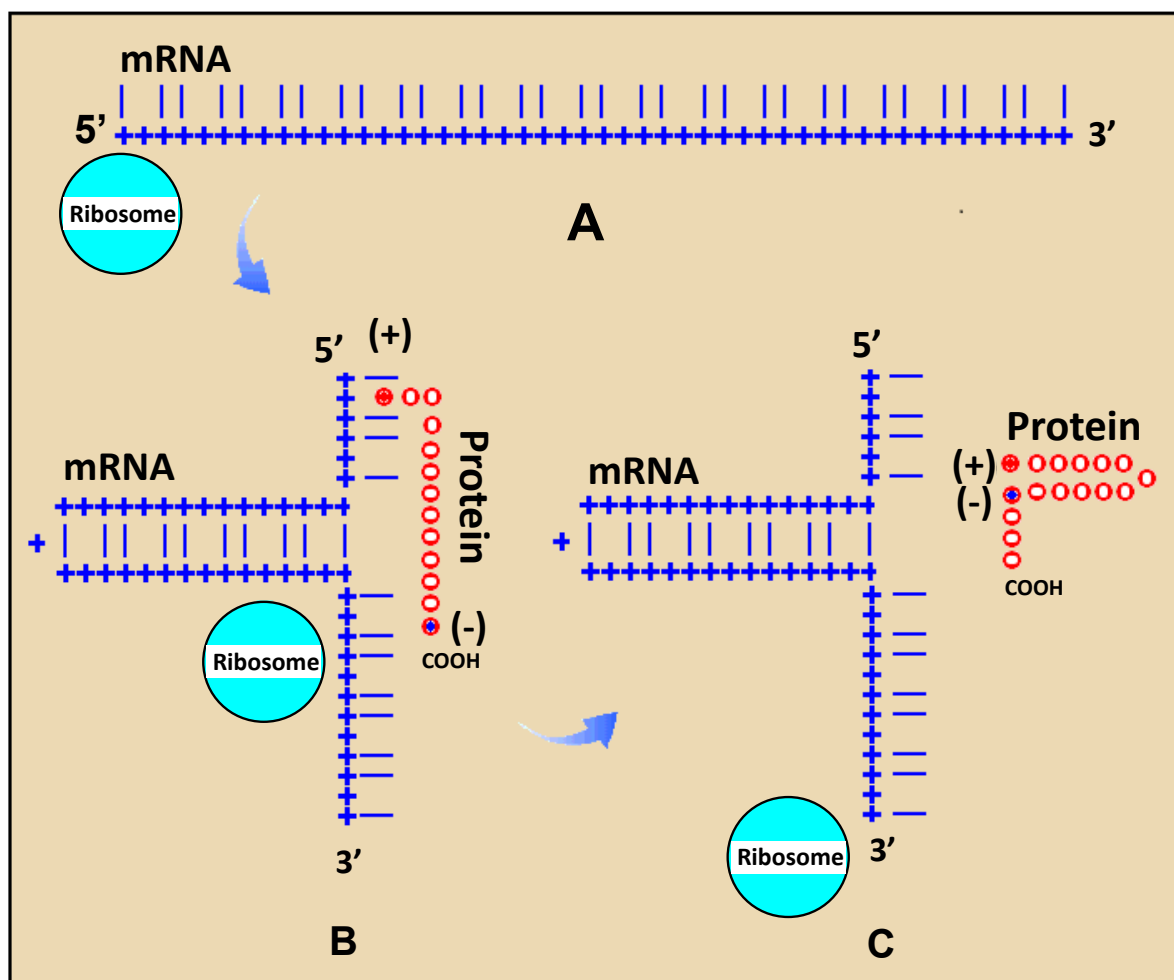


Figure 7: RNA assisted (co-translational) protein loop formation. Translation begins with the attachment of the 5' end of a mRNA to the ribosome (A). Ribonucleotides are indicated by blue + and the 1st and 3rd bases in the codons by blue lines, while the 2nd base positions are left empty. A positively charged amino acid [(+) and red dots], for example arginine, remains attached to its codon. The mRNA forms a loop because the 1st and 3rd bases are locally complementary to each other in reverse orientation (B). The growing protein is indicated by red circles (o). When translation proceeds to an amino acid with especially high affinity to the mRNA-attached arginine, for example a negatively charged Glu or Asp [(-) and blue dot], the charge attraction removes the Arg from its mRNA binding site and the entire protein is released from the mRNA and completes a protein loop (C). The protein continues to grow toward the direction of its carboxy terminal (COOH) [15].

structures) are preferentially coded by co-locating (complementary) codons (in the coding nucleic acid structures, mRNA) led to the theory of mRNA chaperons. It means that mRNAs contain even folding information (in addition to the well-known sequence information) and able to guide/assist the folding of their own coded peptides as molecular chaperons [15-17]. The possible mechanism of this nucleic acid assisted protein folding is modeled and illustrated in Figures 7 and 8. This initial model didn't explain the involvement and role of tRNA-s.

Transfer RNAs are “active” adaptors

Transfer-RNA was proposed by Crick as a necessity to “adapt” the codon to the encoded amino acid (a codon is three times longer than an amino acid). However, it became apparent that tRNA is approximately 20-times larger than necessary for this role. The “adaptor” became a clumsy “barrier” between the nucleic acid and protein worlds [6,18]. Therefore, radical revision (research) of the function of tRNA was necessary to understand the transfer of spatial information from nucleic acids to peptides, and to make sense of the size and frequency

of tRNA-s [20]. Thermodynamic studies, and the literature, indicate that tRNA has several possible configurations (in addition to the canonical cloverleaf form). Furthermore, side-by-side interactions between tRNA-s are thermodynamically favored. Consequently, we concluded and suggested that there is a tRNA cycle involving unfolding, interaction and refolding of tRNA-s, and that this cycle brings codon-anticodon sites into the proximity of the corresponding amino acids [19,20]. Some “dedicated” amino acids remain in contact with their codons after polymerization of amino acids and release of the newly synthesized peptide. This temporary contact is necessary for nucleic acid-assisted protein folding, and this direct codon-amino acid contact is established by tRNA-s (Figure 9).

General Discussion and Conclusions

Nucleic acids and proteins are two very different classes of biological macromolecules and it is tempting to treat them separately. Nucleic acids are mainly related to the preservation and passage of

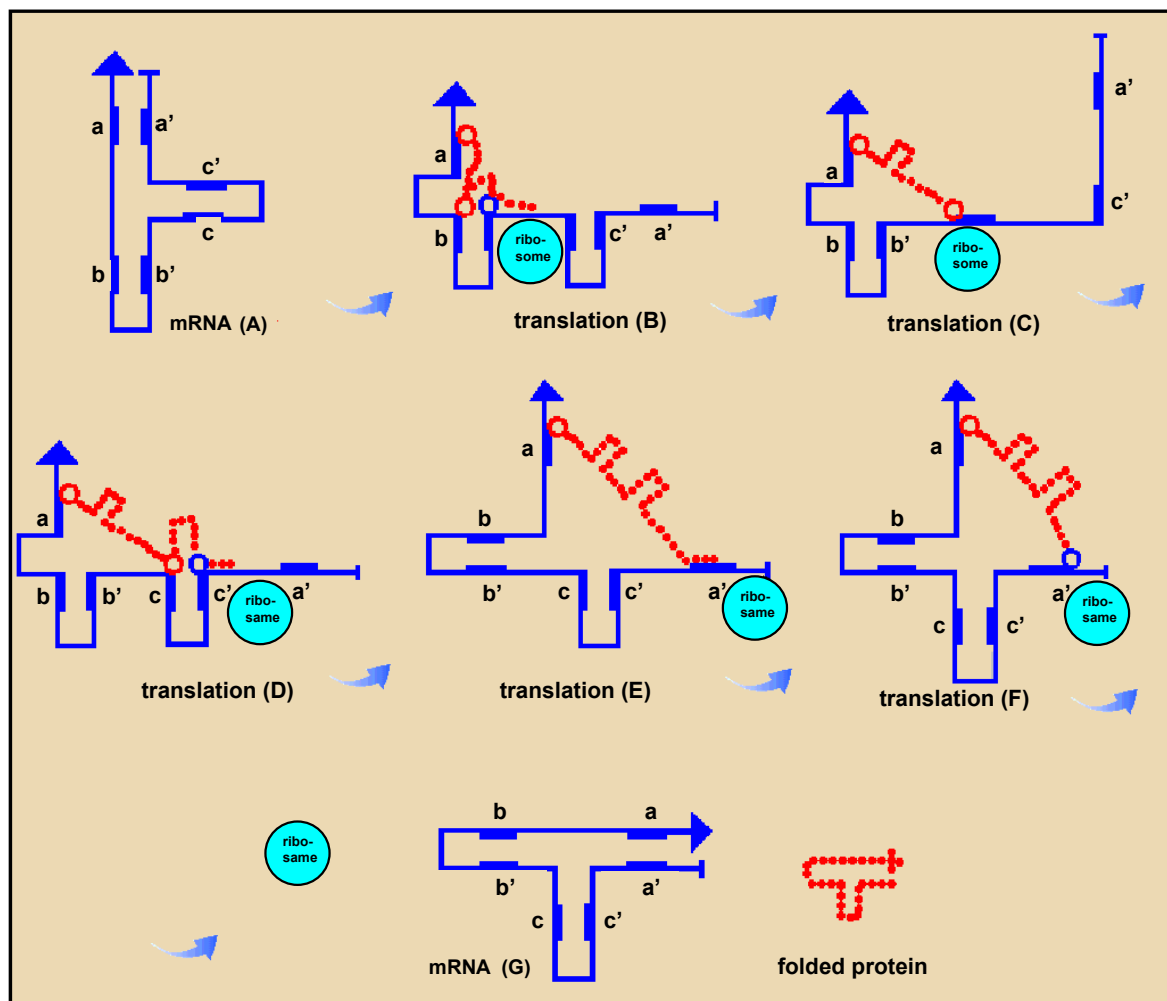


Figure 8: RNA-assisted (co-translational) protein folding. There are three reverse and complementary regions in a mRNA (blue line, A): a-a', b-b', c-c', which fold the mRNA into a T-like shape. During the translation process the mRNA unfolds on the surface of the ribosome, but subsequently refolds, accompanied by its translated and lengthening peptide (red dotted line, B-F). The result of translation is a temporary ribonucleo-peptide complex, which dissociates into two T-shape-like structures: the original mRNA and the properly folded protein product (G). The red circles indicate the specific, temporary attachment points between the RNA and protein (for example a basic amino acid) while the blue circles indicate amino acids with exceptionally high affinity for the attachment points (for example acidic amino acids); these capture the amino acids at the attachment point and dissociate the ribonucleo-protein complex. Transfer-RNAs are of course important participants in translation, but they are not included in this scenario [15].

biological information (genomics, genetics and inheritance). Proteins, on the other hand, are mostly known as functional, regulatory and structural molecules. Some scientists (represented by Carl Woese) are speaking about two different “worlds”, there the protein “world” developed from the ancient and recently extinct RNA “world” [6,18]. Francis Crick, - who became very influential as the founding father of molecular biology, - strongly refused any specific connection (stereochemical fitting) between the proteins and nucleic acids, with the only “permitted” exception of the tRNA adaptors [7].

We don't know whether life was possible without any proteins and only with nucleic acids as functional molecules. However we do know that today every single life-form contains both kinds of macromolecules. It is possible to fold DNA without any protein, *in vitro*, but the natural DNA folding and structure formation (*in vivo*) always occurs in the presence of proteins. It is possible to synthesize proteins without nucleic acids (*in vitro*) but folding is often erroneous. The natural protein syntheses and folding (*in vivo*) always occurs in

the presence of nucleic acids. Consequently the obligatory existence of nucleic-acid/protein co-location in living organism makes it possible and very likely that the two molecular forms co-developed during the evolution and they specifically and intimately interact with each other.

Consequently it is not really surprising to find specific, functional connection between codons and coded amino acids (*The Periodic Table of Codons and Nucleic Acids*) and see (literally) that codons and coded amino acids are frequently co-locating (even if the extent of these co-locations is not known).

It is more and more obvious, that not all protein folds correctly on its own (spontaneously) which indicates the lack of complete folding information in only the amino acid sequence of the folding protein. This contradicts the *Anfinsen theorem*. However, at the same time there is a large excess of molecular information in the mRNAs (codon redundancy, synonymous codons) without any well-established biological function or purpose. The information balance between mRNA and coded protein may be described by equation:

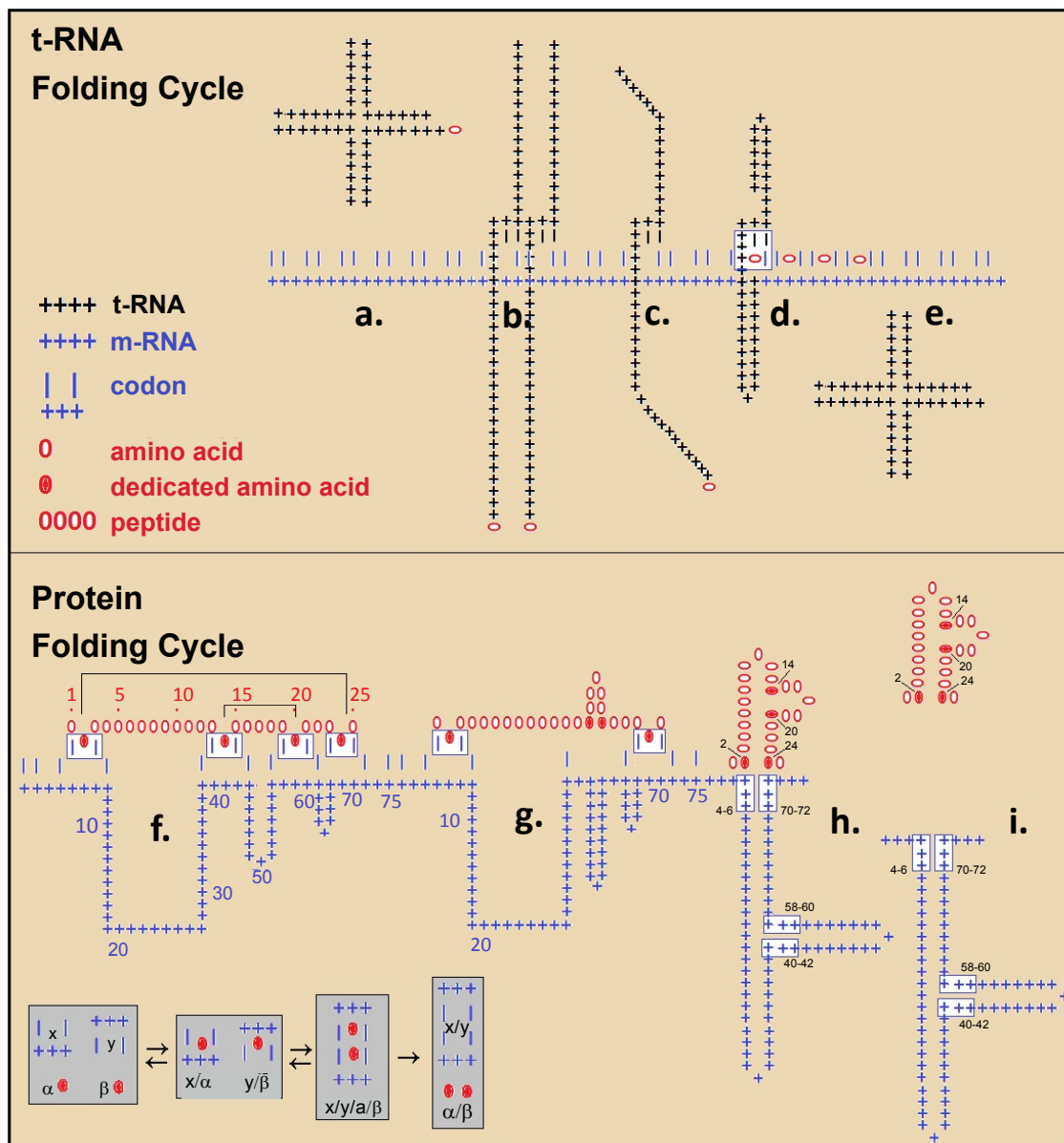


Figure 9: Concept of RNA assisted protein folding. The model comprises tRNA (upper part) and protein (lower part) folding cycles. During the tRNA cycle, the aminoacyl-tRNA (clover-leaf form, (a)) unfolds, interacts with its codon, and the previously attached tRNA (b) refolds to a configuration that brings the amino acid tail into close proximity with the codon-anticodon site (c, d), loses the amino acid, refolds to its original cloverleaf configuration (e) and is recycled. The protein folding cycle begins when the peptide synthetase forms peptide bonds between individual amino acids. Some “dedicated” amino acids remain attached to their codons, but most are displaced. The difference in length between the peptide and mRNA creates mRNA folds (f) and the interaction between complementary codons creates peptide folds (g), one after the other (h). The growing peptide-mRNA complex dissociates after “pairing” the last “dedicated” amino acid pair with its corresponding codon pair (i) and the mRNA is recycled. The numbers indicate the positions of the dedicated amino acids and their codons in a 25 amino acid-long peptide and its 75 nucleotide-long mRNA.

The inserted gray boxes depict the rules of the Proteomic Code [2]: co-locating amino acids (α and β) are encoded by codons (x and y) which are complementary to each other at the 1st and 3rd nucleotide positions; they form different complexes with each other (x/α , y/β , $x/y/\alpha/\beta$, x/y , α/β) [19].

Sequence information in mRNA (+) excess from codon redundancy in mRNA = Sequence information in Protein (+) folding information in Protein

We suggest that

(+) excess from codon redundancy in mRNA = (+) folding information in Protein

The size and large number of tRNA-s is also an un-explained dilemma, which suggests the existence of much more function than it is requested from a passive “adaptor”. The mRNA-assisted, co-translational protein folding, that we suggest here, might request a more active tRNA, than we know today and might explain the tRNA dilemma. We suggest a “tRNA cycle” to model this novel, active function.

Consequently there are many well-known observations that are

warning for the flows in the recent concept of translation and are indicating that the canonical model became obsolete. Fortunately the development of *in silico* biology provides a very elegant solution for many of these problems that we have. However the theory (based on bioinformatical experiments) is, at this time, far ahead of our relevant knowledge from classical biochemical laboratory observations and it is colliding with many, strong paradigms.

We suggest that we face these difficulties and seriously consider the needs (and ways) for upgrading the canonical concept of translation.

Addendum: “Roadmap”

Science is often paradigm-driven. It is expectable, that any effort to upgrade one of the “holy grails” of molecular biology will be challenging because of collisions with familiar and comforting paradigms [5]. However computational and theoretical biology strengthened his position, as scientific discipline, and became a very powerful instrument and source of developments in life sciences.

Major contributions to the modern concept of translation are expected from the following sources:

1. *In Silico* technologies: Bioinformatics, computational/theoretical biology, system biology

a. Protein v. s. nucleic acid sequence comparisons (alignments) regarding the 2D distribution of physicochemical properties. Amino acid size-, charge- and hydrophathy indexes and matrices were successfully used to characterize specific protein to protein interactions [21]. Similar methods should be usable even for prediction of specific nucleic-acid to protein interactions.

b. Large-scale comparison of the 3D structures of peptides and mRNAs to detect and characterize the presence and conservation of protein folding information already on RNA level.

c. The development of residue-based protein folding methods, say *protFOLD*, in analogy to the very successful *mFOLD* method which is widely used for modeling nucleic acid structures [22].

2. Laboratory technologies:

a. Physico-chemical definition of codons (64!), as fundamental functional units of nucleic acids (instead of the 4 nucleotides).

b. Studies on “synonymous proteins” i. e. proteins which have the same secondary structure (sequence) but their 3D structure (folding) is different. Differences in the coding mRNA sequences (synonymous codons) of the respective proteins are expected [23].

c. Refocusing on the molecular biology of tRNA to explain its unusually large size and abundance.

It is not to forget that molecular biology is a very young biological discipline where “upgrade” is only the sign of natural and healthy development.

Competing Interests

The author states that he has no competing interests.

Acknowledgement

This work was supported by grants from the Homulus Foundation, Los Angeles, CA, USA. We used public databases and software during this study. We

would like to thank the thousands of scientists who have created and maintained these resources, supported by generous grants from nations and individuals. JCB has been a “scientist in the US National Interest” since 2006. Jan Charles Biro wishes to thank the trust and the support of this great Nation.

The author also wishes to thank the continuous attention and advice of George L Gabor MIKLOS PhD (Founder of Atomic Oncology, Sydney, Australia). The pioneering works and views of Prof. Carl Woese (1928-2012) were very useful and indeed essential for parts of our work. We respectfully recognize and acknowledge this.

References

1. Biró J (1981) Comparative analysis of specificity in protein-protein interactions. Part II.: The complementary coding of some proteins as the possible source of specificity in protein-protein interactions. *Med Hypotheses* 7: 981-993.
2. Biro JC (2007) The Proteomic Code: a molecular recognition code for proteins. *Theor Biol Med Model* 4: 45.
3. Biro JC (2006) A novel intra-molecular protein-protein interaction code based on partial complementary coding of co-locating amino acids. *Med Hypotheses* 66: 137-142.
4. Biro JC (2005) Nucleic acid chaperons: a theory of an RNA-assisted protein folding. *Theor Biol Med Model* 2: 35.
5. Biro JC (2013) Suggestion to Complete the Canonical Genetic Code: The Proteomic Code and Nucleic Acid Assisted Protein Folding. *J Theor Comput Sci* 1: 103.
6. Woese CR (1967) *The Genetic Code: The Molecular Basis for Gene Expression*. New York: Harper & Row.
7. Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38: 367-379.
8. Biro JC, Benyó B, Sansom C, Szilávecz A, Fördös G, et al. (2003) A common periodic table of codons and amino acids. *Biochem Biophys Res Commun* 306: 408-415.
9. Biro JC (2006) Indications that “codon boundaries” are physico-chemically defined and that protein-folding information is contained in the redundant exon bases. *Theor Biol Med Model* 3: 28.
10. Biro JC, Biro JM (2004) Frequent occurrence of recognition site-like sequences in the restriction endonucleases. *BMC Bioinformatics* 5: 30.
11. Biro JC, Biro JM (2004) Frequent occurrence of recognition site-like sequences in the restriction endonucleases. *BMC Bioinformatics* 5: 30.
12. Biro JC (2008) Correlation between nucleotide composition and folding energy of coding sequences with special attention to wobble bases. *Theor Biol Med Model* 5: 14.
13. Anfinsen CB, Redfield RR, Choate WL, Page J, Carroll WR (1954) Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J Biol Chem* 207: 201-210.
14. Levinthal C (1969) How to fold graciously in Mossbauer spectroscopy in biological systems? In: *Mössbauer Spectroscopy in Biological Systems Proceedings*, University of Illinois Press.
15. Biro JC (2005) Nucleic acid chaperons: a theory of an RNA-assisted protein folding. *Theor Biol Med Model* 2: 35.
16. Biro JC (2006) A novel intra-molecular protein-protein interaction code based on partial complementary coding of co-locating amino acids. *Med Hypotheses* 66: 137-142.
17. Biro JC (2006) Protein folding information in nucleic acids which is not present in the genetic code. *Ann N Y Acad Sci* 1091: 399-411.
18. Woese CR (2001) Translation: in retrospect and prospect. *RNA* 7: 1055-1067.
19. Biro JC, Biro JM (2013) The concept of RNA-assisted protein folding: representation of amino acid kinetics at the tRNA level. *J Theor Biol* 317: 168-174.
20. Biro JC (2013) Coding nucleic acids are chaperons for protein folding: a novel theory of protein folding. *Gene* 515: 249-257.
21. Biro JC (2006) Amino acid size, charge, hydrophathy indices and matrices for protein structure analysis. *Theor Biol Med Model* 3: 15.

22. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406-3415.

23. Hu S, Wang M, Cai G, He M (2013) Genetic code-guided protein synthesis and folding in *Escherichia coli*. *J Biol Chem* 288: 30855-30861.