Bioinformatics, 35(4), 2019, 579–583
doi: 10.1093/bioinformatics/bty678
Advance Access Publication Date: 7 August 2018
Original Paper

OXFORD

Genome analysis

Global importance of RNA secondary structures in protein-coding sequences

Markus Fricke[1,2], Ruman Gerst[2], Bashar Ibrahim[1,2], Michael Niepmann[3] and Manja Marz[1,2,4,5,*]

[1]RNA Bioinformatics and High Throughput Analysis, Faculty of Mathematics and Computer Science, Friedrich Schiller University Jena, 07743 Jena, Germany, [2]European Virus Bioinformatics Center (EVBC), 07743 Jena, Germany, [3]Institute of Biochemistry, Faculty of Medicine, Justus-Liebig-University, 35392 Giessen, Germany, [4]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, 04103 Leipzig, Germany and [5]FLI Leibniz Institute for Age Research, 07745 Jena, Germany

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on October 2, 2018; revised on July 4, 2018; editorial decision on July 30, 2018; accepted on August 6, 2018

Abstract

Motivation: The protein-coding sequences of messenger RNAs are the linear template for translation of the gene sequence into protein. Nevertheless, the RNA can also form secondary structures by intramolecular base-pairing.

Results: We show that the nucleotide distribution within codons is biased in all taxa of life on a global scale. Thereby, RNA secondary structures that require base-pairing between the position 1 of a codon with the position 1 of an opposing codon (here named RNA secondary structure class c_1) are under-represented. We conclude that this bias may result from the co-evolution of codon sequence and mRNA secondary structure, suggesting that RNA secondary structures are generally important in protein-coding regions of mRNAs. The above result also implies that codon position 2 has a smaller influence on the amino acid choice than codon position 1.

Contact: evbc@uni-jena.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

RNA secondary structures are involved in the regulation of mRNA function and fate. In many cases, functional RNA secondary structures are located in the 5'- or 3'-untranslated regions (Wilkie *et al.*, 2003), like the secondary structures involved in regulation of histone mRNA translation (Marzluff *et al.*, 2008), the iron response elements (Anderson *et al.*, 2012) or the gamma interferon inhibitor of translation element that is involved in limiting the cellular immune response (Mukhopadhyay *et al.*, 2009). In many viral RNAs, RNA secondary structures in the untranslated regions regulate the viral life cycle (Liu *et al.*, 2009). Functional RNA secondary structures in viral RNAs have also been reported in the protein-coding regions, like for the MS2 bacteriophage coat protein mRNA (Ball, 1973), for cis-replication elements (CREs) involved in the replication of plus strand RNA viruses (Fricke *et al.*, 2015; Liu *et al.*, 2009) or retroviruses (Konecny *et al.*, 2000) and for the RNA encapsidation signals of betacoronaviruses (Fosmire *et al.*, 1992) and Hepatitis B Virus (Junker-Niepmann *et al.*, 1990). However, in the protein-coding regions of cellular mRNAs such RNA secondary structures are known to exert functions only in specific cases, like the selenocysteine insertion sequence involved in the incorporation of the amino acid selenocysteine into proteins (Donovan and Copeland, 2010).

The relationship between codon usage and RNA secondary structures was discussed frequently (Ball, 1973; Carlini et al., 2001; Fitch, 1974; Gu *et al.*, 2014; Konecny *et al.*, 2000; Mao *et al.*, 2014; Oresic *et al.*, 2003; Shabalina *et al.*, 2006). The question arises if mRNA protein-coding regions in general have functional RNA secondary structure elements and if there is a relationship between the evolution of protein sequences and the nucleotide sequences forming RNA secondary structures. Three hypotheses have been considered

(Ball, 1973): (i) codon sequence and secondary structure are independent, i.e. only the codon sequence is under selection pressure solely due to the needs of the amino acid sequence in the functional protein, and the most stable RNA secondary structure is determined only according to the requirements of the codon sequence; (ii) natural selection between synonymous codons which encode the same amino acid could permit the optimization of RNA secondary structures (Fitch, 1974) and (iii) a strong selection pressure for functional RNA secondary structures could drive the choice also of non-synonymous codons in the mRNA, leading to the choice of another amino acid according to the requirements of the RNA secondary structure (Konecny *et al.*, 2000).

2 Materials and methods

2.1 Prerequisites

For a given nucleotide sequence $X = (x_1, \ldots, x_n)$ of length n, we labeled each nucleotide position i in a coding region with the corresponding codon index $y_i \in \{1, 2, 3\}$, calculated as $y_i = (i - 1) \mod 3 + 1$. We defined for each base-pair a base-pair index $y_i : y_j$. We further group the base-pair indexes into base pair/secondary structure index classes c_1={1:1, 2:3, 3:2}, c_2={2:2, 1:3, 3:1} and c_3={3:3, 1:2, 2:1}.

We used a pairwise *t*-test with Benjamini and Hochberg *P*-value adjustment as significance test. The *P*-values are encoded as following: ≤ 0.001 ***, ≤ 0.01 ** and ≤ 0.05 *.

2.2 Data

We downloaded all genomes and annotations of the NCBI RefSeq genome database (May 2016, ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq) and (in case of multiple isoforms) randomly selected one isoform of each protein-coding gene sequence. We discarded all sequences with *length* $\mod 3 \neq 0$. We analyzed mitochondrial and chloroplast mRNAs separately. Due to runtime reasons, we reduced the dataset. We randomly selected 200.000 CDSs of each group (if available), with a balanced number of CDS per species per group. The mitochondrial and chloroplast mRNAs are downloaded from the NCBI nucleotide database with the following search term: `biomol_genomic[PROP] AND refseq[filter] AND mitochondrion[filter]` or with the filter `chloroplast[filter]` (July 2017, www.ncbi.nlm.nih.gov/nuccore). The mitochondrial mRNAs are separated by the taxa plants, protists, fungi and animals.

2.3 RNA secondary structure prediction of CDS

We performed an RNA secondary structure prediction with `RNAfold` (Lorenz *et al.*, 2011) (using the default parameters) for all species and each CDS. We used a sliding window approach with a window size of 50 nt and a step size of 5 nt to be fast (for runtime reasons) and to move the edge of the window over all possible frames. For the predicted RNA secondary structure of each window, we counted the frequency of the base pair/RNA secondary structure classes. For the viral CDS, we additionally used window sizes of 25, 50, 100, 200 and 400 nt to exemplary exclude a window specific bias. We are aware of the fact that `RNAfold` produces false positive secondary structure predictions, but we assume false positives to be equally distributed over all three classes.

2.4 Sequence shuffling

The di-nucleotide shuffling of viral CDS was performed with `uShuffle` (Jiang *et al.*, 2008). For the mono-, codon-, and di-codon-nucleotide shuffling, we used a simple in-house script (available on request). The codon and di-codon-nucleotide shuffling will persevere the codon frequency of the underlying mRNA sequences.

3 Results and discussion

We provide a global-scale analysis of the crosstalk between amino acid coding requirements and RNA secondary structure formation in protein-coding sequences. Using `RNAfold` (Lorenz *et al.*, 2011), we analyzed the occurrence of three different base pair/secondary structure classes, based on the arrangement of amino acid codons within RNA secondary structures. Each arrangement of valid base-pairs (A:U, U:A, G:C, C:G, U:G, G:U) between the codon positions 1, 2 or 3 of one codon and the codon positions 1, 2 or 3 of an opposing codon in RNA secondary structures can be grouped into the following classes: c_1={1:1, 2:3, 3:2}, c_2={2:2, 1:3, 3:1} and c_3={3:3, 1:2, 2:1} (Fig. 1). Thereby, these sequence stretches do not need to have a length in multiples of three nucleotides to belong to the respective class, as illustrated in an artificial RNA structure (Fig. 1B). Consequently, these three base pair/RNA secondary structure classes cover virtually all possible RNA secondary structures. The frequency of these base pair/secondary structure classes was counted on a global scale in the protein-coding regions of mRNAs from viruses, bacteria, archaea, plants, fungi, invertebrates, non-mammalian vertebrates and mammalia.

As a result, we found base-pairs/RNA secondary structures of class c_1 are significantly under-represented globally in the mRNA protein-coding regions in all taxa of life (Fig. 2).

In class c_1, two opposing nucleotides which each are in position 1 of two opposing codons would be required to pair. However, the nucleotide in codon position 1 largely defines the exact nature of an amino acid and by that its function in the protein. Thus, this arrangement would require two specific amino acids in a protein to co-evolve to allow RNA secondary structure formation, which appears highly unlikely. Conversely, the fact of an under-represented class c_1 is interpreted as a selection pressure to allow RNA secondary structures in mRNA coding regions. At the same time, pairing of the nucleotides in position 2 of a codon with a

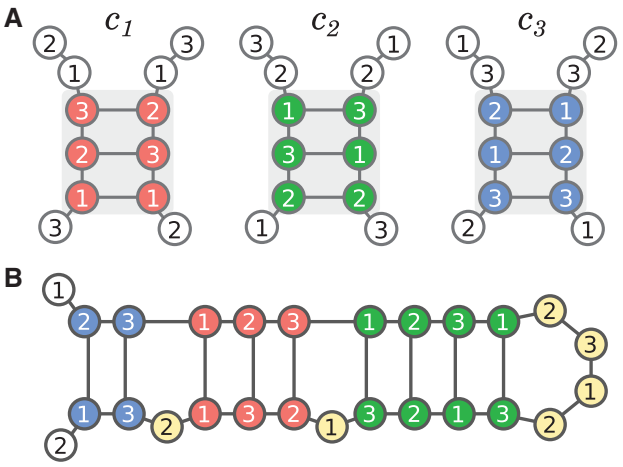

Fig. 1. (**A**) Three base pair/RNA secondary structure classes that represent all possible base-pairs between positions 1, 2 and 3 of one codon with positions 1, 2 and 3 of an opposing codon in RNA secondary structures. The classes are based on the following groups c_1={1:1, 2:3, 3:2}, c_2={1:3, 2:2, 3:1} and c_3={1:2, 2:1, 3:3}. (**B**) An artificial example RNA secondary structure with the different classes corresponding to A. Shifts between the different classes are possible, introduced through loops and bulges

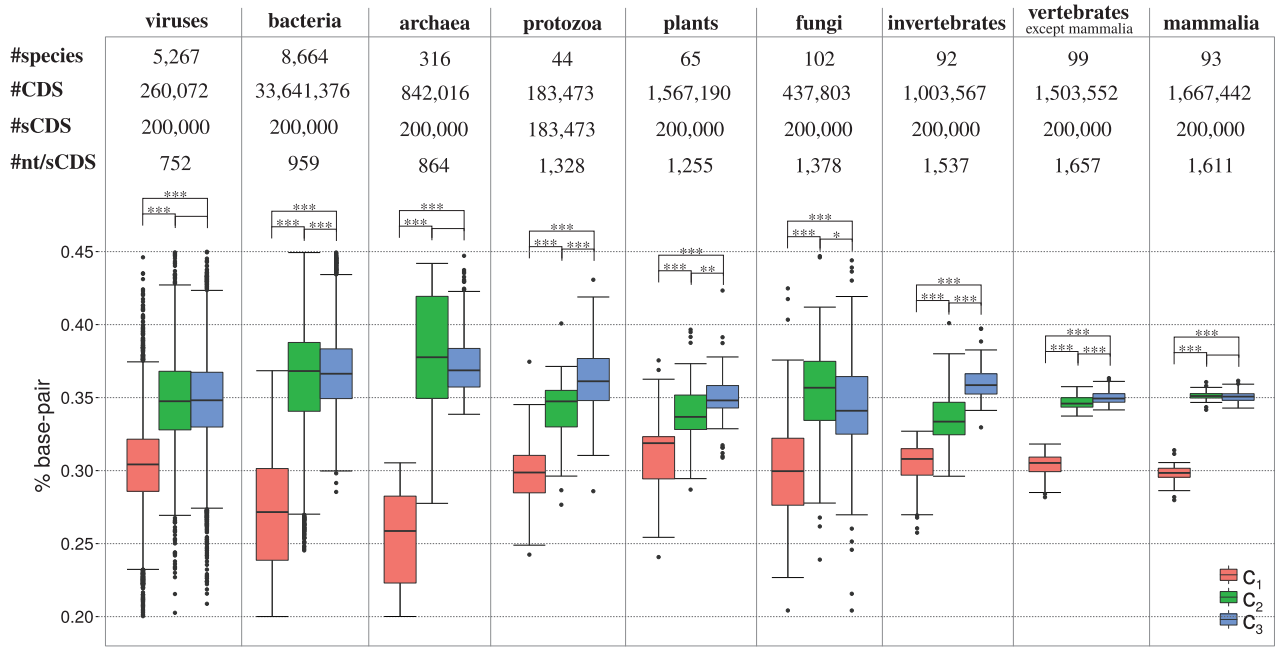

Fig. 2. Distribution of class c_1, c_2 and c_3 within all selected coding sequences (sCDS, see methods for selection procedure) except mitochondria and chloroplasts for each taxonomic group. #species – number of different species; #CDS – number of coding sequence (CDS); #sCDS – number of selected CDS; #nt/sCDS – number of nucleotides per selected CDS

nucleotide in position 3 of another codon in class c_1 may be highly likely due to the degenerate nature of triplett position 3.

To ensure the under-representation of class c_1 is not an artifact of the RNA folding algorithm, we performed several tests. The under-representation of class c_1 is not triggered by a single specific base-pair since all base-pairs of class c_1 are under-represented (Supplementary Fig. S1). The bias is not affected by the folding range of our method (see Supplementary Material and Supplementary Fig. S2) and occurs not by chance (Supplementary Fig. S4). We can also find the same effect if we use other folding algorithms, such as CONTRAfold (Do *et al.*, 2006), a secondary structure prediction method based on conditional log-linear models (data not shown). Furthermore, if we consider the complete ensemble of RNA secondary structures, we find the same bias (see Supplementary Material and Supplementary Fig. S3). The effect disappears after artificial shuffling of the underlying RNA sequences with a word size of one or two, but remains with a shuffling of word size three or six (conserving the underlying codons and di-codons, Supplementary Fig. S4), showing the effect being not specific to the nucleotide distribution (like GC content) but specific for the underlying codon distribution. Moreover, if we build artificial codons but keep the nucleotide distribution per codon position constant, the effect remains the same (Supplementary Fig. S5). This demonstrates that the global nucleotide distribution per codon position but not the codon itself is important.

In addition, we generally also find a less pronounced but visible under-representation of the pairing of codon position 2 with position 2 of an opposing codon in class c_2 (Supplementary Fig. S1). This illustrates also codon position 2 to be important, likely because Uracil in position 2 defines the hydrophobic character of an amino acid. A change of the hydrophobic versus hydrophilic nature of amino acids could result in general misfolding of proteins. However, our results shown here imply that codon position 2 has a smaller influence on the amino acid choice than codon position 1.

Although changes in codon position 2 are generally believed to be more important for protein structure and function than changes in codon position 1 (Berleant *et al.*, 2009; Pohlmeyer, 2008), our results challenge this general assumption. In this context, it is important to note that in previous studies in which the physicochemical properties of amino acids like size, hydropathy and charge were considered, charged amino acids were usually classified by simply having a charged side chain or not. However, the polarity of charge and the effects of changing this polarity were not taken into account. As a consequence, changes in amino acid charge from positive to negative or *vice versa* were usually assigned only low penalties, other than changes in amino acid hydropathy and size (Berleant *et al.*, 2009; Goodarzi *et al.*, 2007; Miyata *et al.*, 1979; Nemzer, 2017).

However, our results suggest that codon position 1 is globally more important than codon position 2. Therefore, we inspected the genetic code to find out which changes in amino acid properties can occur after changes in codon position 1 or position 2. We found that charge reversal (i.e. the change of a positive to a negative charge or *vice versa*) only occurs upon single nucleotide exchanges in codon position 1 (G to A, G to C or *vice versa*) but never upon single nucleotide exchanges in codon position 2. We conclude that the selection pressure against charge reversal—which may destroy salt bridges in proteins that are important for the protein's structure and function, like the His-Asp salt bridge that mediates the Bohr effect in hemoglobin (Russu *et al.*, 1980)—may have been largely underestimated in previous studies.

Globally, structured mRNAs have a more pronounced base-pair bias than less structured mRNAs. Compared to the low structured mRNAs (Fig. 3A), we found a slight increase of cytosin and guanin content at codon positions 1 and 2 in the structured mRNAs but a strong increase at codon position 3 (Fig. 3B). Based on the higher G/C content, the G:C base-pairing was significantly increased. There seems to be a consistent trend over all organisms that the increase in G/C content among the base-pairs of one class occurs to a large extent in those base-pairs which are involved in a pairing with a codon position 3 (Supplementary Fig. S1). The largest differences

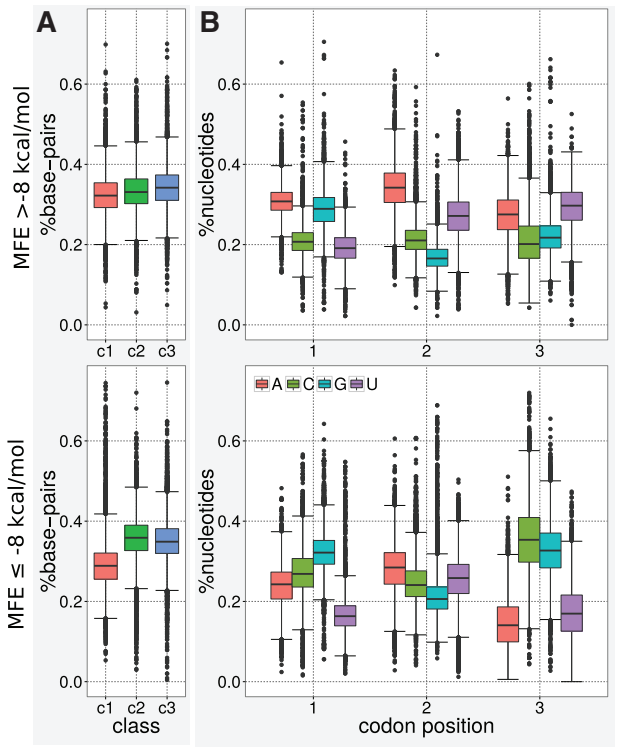

Fig. 3. (**A**) Class distribution and (**B**) nucleotide distribution of each codon positions of all human mRNAs with less stable (mean MFE>−8 kcal/mol over all folding windows, #7104 mRNAs, TOP) and rather stable (mean MFE ≤ −8 kcal/mol over all folding windows, #15022 mRNAs, BOTTOM) RNA secondary structures. The nucleotide distributions in codon positions 2 and 3 of the less stable mRNAs are similar, since base-pairings of 2:3 and 3:2 are likely. In contrast, mRNAs with more stable RNA secondary structure show an increased GC content at codon position 1 and in particular at position 3. This increases the G:C and C:G base-pairings 1:3, 3:1 and 3:3 (Supplementary Fig. S6) and consequently increases the frequency of classes c_2 and c_3

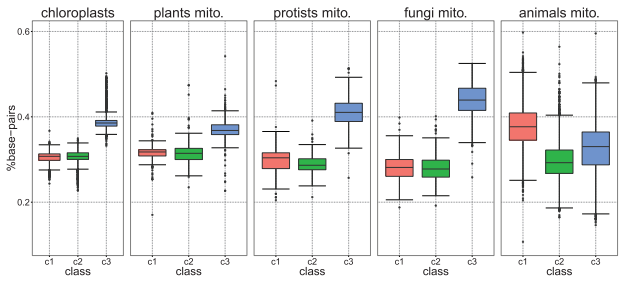

Fig. 4. Base pair/RNA secondary structure class distribution in mitochondria and chloroplasts

were found in 1:3, 3:1, 2:3, 3:2 and 3:3 pairings, whereas no big changes were found in 1:1, 2:2, 1:2 and 2:1 pairings (Supplementary Fig. S6). This shows that an increase in G/C content in more structured mRNAs occurred mainly in the wobble position 3, leaving the amino acid sequence largely unchanged. In turn, this suggests that the co-evolution of amino acid sequence and RNA secondary structures may have been mainly driven by the amino acid sequence. Conversely, it appears difficult to estimate at this global scale if the RNA secondary structure influences the amino acid sequence by the choice of non-synonymous codons. Even though the pronounced base-pair bias in all taxonomic groups supports the presence of structured mRNAs *in vivo*, up to now we are aware of only a few functional RNA structures in mRNA coding regions (Ball, 1973; Fosmire *et al.*, 1992; Fricke *et al.*, 2015; Konecny *et al.*, 2000; Liu *et al.*, 2009; Nassal *et al.*, 1990).

Interestingly, in nearly all mitochondrial (except animal) and chloroplast mRNAs the RNA secondary structure class c_2 and class c_1 is under-represented, whereas in animals only class c_2 but not class c_1 is under-represented (Fig. 4). In class c_2 RNA secondary structures, pairing of two opposing nucleotides in position 2 of each codon would be required. However, variation in this position would likely change the hydrophobic versus hydrophilic nature of the amino acids. Thus, the under-representation of class c_2 in plastid genomes may be due to the fact that plastid genomes largely encode those highly hydrophobic membrane proteins which are involved in plastid redox metabolism to allow for short regulatory gene expression circuits (Allen, 2015). This effect appears particularly pronounced in animal mitochondria which have lost all protein-coding genes that do not encode respiratory chain complex subunits (Lang *et al.*, 1999). Here, only class c_2 but not class c_1 is under-represented (Fig. 4). This suggests that in these animal mitochondrial sequences exclusively coding for those subunits of the mitochondrial respiratory chain complexes that are located in the membrane (Soto *et al.*, 2012; Wirth *et al.*, 2016), hydropathy has become the most important parameter for codon selection.

Some viral RNA genomes contain known functional RNA secondary structures. For a proof of principle, we therefore analyzed the 3'-region of the Hepatitis C Virus RNA genome which contains RNA secondary structure elements known to be involved in genome replication (Niepmann *et al.*, 2018). The results show that in two regions containing functional signals (J7880 and J8880) class c_1 is strongly under-represented, whereas in the coding sequence (CDS) region containing the so-called CRE and encoding the hydrophobic transmembrane region of the protein, class c_2 is strongly under-represented, with high values in class c_1 (Supplementary Fig. S7). This demonstrates that under-representation of RNA secondary structure classes c_1 or c_2 can actually correlate with functional RNA secondary structures.

As a control, we also considered if the codon frequencies as found in nature contribute to the under-representation of RNA secondary structure class c_1. Therefore, we calculated the probability that two codon positions can base-pair by just using the codons taken from the table of the genetic code, while taking into account the actual frequency of each codon as found in each taxonomic group in nature. Without taking the underlying codon frequency into account, the possible base-pair distribution of classes c_1, c_2 and c_3 would be identical (33.3% each). In contrast, when we weight each possible base pair between the codons with the underlying codon frequency, we found a slight under-representation of class c_1 (see Table 1, left part). This demonstrates that even the codon frequencies are sufficient to show an under-representation of RNA secondary structure class c_1, underlining the global importance of RNA secondary structures in CDSs. Importantly, in the naturally occurring CDSs with the order of codons as required by the amino acids in the proteins, the under-representation of RNA secondary structure class c_1 comes with an additional strong bias (Table 1, right part).

It was shown that RNA secondary structure is under positive selection in coding regions of bacteria (Katz and Burge, 2003), but the conclusions are controversial. On the one hand, stable structures in mRNAs can slow down ribosomes (Chen *et al.*, 2013; Tholstrup *et al.*, 2012; Wen *et al.*, 2008), arguing against a high number of very stable RNA structures in coding regions. On the other hand, a recent study of yeast mRNAs found a positive correlation of mRNA secondary structure with protein abundance, without affecting mRNA half-life (Zur and Tuller, 2012). For this finding, Mao *et al.* proposed a possible explanation that appears counter-intuitive at

Table 1. Class distribution over all possible base-pairs between the 64 codons of the genetic code, weighted with the natural codon frequency of both codons (left). Class distribution based on the RNA secondary structure as depicted in Figure 2 (right)

	Codon frequency based			mRNA structure based		
	c1	c2	c3	c1	c2	c3
Archaea	31.7	33.8	34.4	25.2	37.8	37.3
Bacteria	32.1	33.7	34.1	27.0	36.3	36.7
Fungi	32.5	33.7	33.7	30.1	35.2	34.1
Invertebrates	32.8	33.5	33.5	30.3	33.5	36.0
Plants	32.8	33.5	33.5	31.2	33.9	34.9
Protozoa	32.7	33.5	33.6	29.6	34.0	36.1
Mammals	32.7	33.6	33.5	29.8	35.1	35.0
Vertebrates	32.8	33.6	33.5	30.3	34.6	35.0
Viruses	32.4	33.8	33.6	30.2	34.8	34.8

first sight. By computational methods they found that moderate RNA secondary structures may increase ribosomal density and therewith translation rate (Mao *et al.*, 2014). Their interpretation is that the first ribosomes that encounter the coding region are slowed down by RNA secondary structures. Then, all following ribosomes catch up closely to create a line of migrating ribosomes which will disable reformation of RNA secondary structures and thus translate the mRNA with high efficiency.

Taken together, synonymous codons in mRNA protein-coding regions are globally selected in all taxa of life in a way that the need to form RNA secondary structures in which a codon position 1 is required to pair with the codon position 1 of another codon is reduced. In turn, this means that mRNA CDSs evolved to allow the formation of RNA secondary structures using synonymous codons, underlining the general importance of such RNA secondary structures for the function of mRNAs in cells. However, this does not exclude that in specific cases conserved functional RNA structures may also drive co-evolution of amino acid sequences using non-synonymous codons.

Funding

This work was supported by the German Research Foundation (DFG) [SFB 1021], AquaDiva [INST 275/367-1] and FungiNet [INST 275/365-1].

Conflict of Interest: none declared.

References

Allen,J.F. (2015) Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci. USA*, **112**, 10231–10238.

Anderson,C.P. *et al.* (2012) Mammalian iron metabolism and its control by iron regulatory proteins. *Biochim. Biophys. Acta*, **1823**, 1468–1483.

Ball,L.A. (1973) Secondary structure and coding potential of the coat protein gene of bacteriophage MS2. *Nature New Biol.*, **242**, 44–45.

Berleant,D. *et al.* (2009) The genetic code–more than just a table. *Cell Biochem. Biophys.*, **55**, 107–116.

Carlini,D.B. *et al.* (2001) The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. *Genetics*, **159**, 623–633.

Chen,C. *et al.* (2013) Dynamics of translation by single ribosomes through mRNA secondary structures. *Nat. Struct. Mol. Biol.*, **20**, 582–588.

Do,C.B. *et al.* (2006) Contrafold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.

Donovan,J. and Copeland,P.R. (2010) Threading the needle: getting selenocysteine into proteins. *Antioxid. Redox Signal.*, **12**, 881–892.

Fitch,W.M. (1974) The large extent of putative secondary nucleic acid structure in random nucleotide sequences or amino acid derived messenger-RNA. *J. Mol. Evol.*, **3**, 279–291.

Fosmire,J.A. *et al.* (1992) Identification and characterization of a coronavirus packaging signal. *J. Virol.*, **66**, 3522–3530.

Fricke,M. *et al.* (2015) Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. *RNA*, **21**, 1219–1232.

Goodarzi,H. *et al.* (2007) Solvent accessibility, residue charge and residue volume, the three ingredients of a robust amino acid substitution matrix. *J. Theor. Biol.*, **245**, 715–725.

Gu,W. *et al.* (2014) The impact of RNA structure on coding sequence evolution in both bacteria and eukaryotes. *BMC Evol. Biol.*, **14**, 87.

Jiang,M. *et al.* (2008) `uShuffle`: a useful tool for shuffling biological sequences while preserving the k-let counts. *BMC Bioinformatics*, **9**, 192.

Junker-Niepmann,M. *et al.* (1990) A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.*, **9**, 3389–3396.

Katz,L. and Burge,C.B. (2003) Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res.*, **13**, 2042–2051.

Konecny,J. *et al.* (2000) Concurrent neutral evolution of mRNA secondary structures and encoded proteins. *J. Mol. Evol.*, **50**, 238–242.

Lang,B.F. *et al.* (1999) Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.*, **33**, 351–397.

Liu,Y. *et al.* (2009) Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochim. Biophys. Acta*, **1789**, 495–517.

Lorenz,R. *et al.* (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.

Mao,Y. *et al.* (2014) Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **42**, 4813–4822.

Marzluff,W.F. *et al.* (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.*, **9**, 843–854.

Miyata,T. *et al.* (1979) Two types of amino acid substitutions in protein evolution. *J. Mol. Evol.*, **12**, 219–236.

Mukhopadhyay,R. *et al.* (2009) The GAIT system: a gatekeeper of inflammatory gene expression. *Trends Biochem. Sci.*, **34**, 324–331.

Nassal,M. *et al.* (1990) Translational inactivation of RNA function: discrimination against a subset of genomic transcripts during HBV nucleocapsid assembly. *Cell*, **63**, 1357–1363.

Nemzer,L.R. (2017) Shannon information entropy in the canonical genetic code. *J. Theor. Biol.*, **415**, 158–170.

Niepmann,M. *et al.* (2018) Signals Involved in Regulation of Hepatitis C Virus RNA Genome Translation and Replication. *Front Microbiol.*, **9**, 395.

Oresic,M. *et al.* (2003) Tracing specific synonymous codon-secondary structure correlations through evolution. *J. Mol. Evol.*, **56**, 473–484.

Pohlmeyer,R. (2008) The genetic code revisited. *J. Theor. Biol.*, **253**, 623–624.

Roseman,M.A. (1988) Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J. Mol. Biol.*, **200**, 513–522.

Russu,I.M. *et al.* (1980) Role of the beta 146 histidyl residue in the alkaline Bohr effect of hemoglobin. *Biochemistry*, **19**, 1043–1052.

Shabalina,S.A. *et al.* (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.*, **34**, 2428–2437.

Soto,I.C. *et al.* (2012) Biogenesis and assembly of eukaryotic cytochrome c oxidase catalytic core. *Biochim. Biophys. Acta*, **1817**, 883–897.

Tholstrup,J. *et al.* (2012) mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res.*, **40**, 303–313.

Wen,J.-D. *et al.* (2008) Following translation by single ribosomes one codon at a time. *Nature*, **452**, 598–603.

Wilkie,G.S. *et al.* (2003) Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. *Trends Biochem. Sci.*, **28**, 182–188.

Wirth,C. *et al.* (2016) Structure and function of mitochondrial complex I. *Biochim. Biophys. Acta*, **1857**, 902–914.

Zur,H. and Tuller,T. (2012) Strong association between mRNA folding strength and protein abundance in *S. cerevisiae*. *EMBO Rep.*, **13**, 272–277.