Journal of Theoretical Biology 414 (2017) 76–90

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Natural mitochondrial proteolysis confirms transcription systematically exchanging/deleting nucleotides, peptides coded by expanded codons

Hervé Seligmann

Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes, Faculté de Médecine, URMITE CNRS-IRD 198 UMER 6236, IHU (Institut Hospitalo-Universitaire), Aix-Marseille University, Marseille, France

ARTICLE INFO

Keywords:
Frameshifting transcription
Translational frameshift
Bijective transformation
Digestive enzymes
Proteome
Systematic deletions
DelRNA

ABSTRACT

Protein sequences have higher linguistic complexities than human languages. This indicates undeciphered multilayered, overprinted information/genetic codes. Some superimposed genetic information is revealed by detections of transcripts systematically (a) exchanging nucleotides (nine symmetric, e.g. A < - > C, fourteen asymmetric, e.g. A- > C- > G- > A, swinger RNAs) translated according to tri-, tetra- and pentacodons, and (b) deleting mono-, dinucleotides after each trinucleotide (delRNAs). Here analyses of two independent proteomic datasets considering natural proteolysis confirm independently translation of these non-canonical RNAs, also along tetra- and pentacodons, increasing coverage of putative, cryptically encoded proteins. Analyses assuming endoproteinase GluC and elastase digestions (cleavages after residues D, E, and A, L, I, V, respectively) detect additional peptides colocalizing with detected non-canonical RNAs. Analyses detect fewer peptides matching GluC-, elastase- than trypsin-digestions: artificial trypsin-digestion outweighs natural proteolysis. Results suggest occurrences of complete proteins entirely matching non-canonical, superimposed encoding(s). Protein-coding after bijective transformations could explain genetic code symmetries, such as along Rumer's transformation.

1. Introduction

Comparisons between texts written in human languages and protein sequences reveal greater linguistic complexity in proteins (Popov et al., 1996). This suggests that genes superimpose several messages (Gimona, 2006). These are not 'only' theoretical issues: several frames of the shortest known self-replicating circular RNA virosoid code for peptides (Abou Haidar et al., 2014). Superimposed messages include stops in frames +1 and −1 which terminate early translation after ribosomal slippages (Seligmann and Pollock, 2003, 2004; Seligmann, 2007; Itzovitz and Alon, 2007), circular code codons which apparently prevent ribosomal slippages (Arquès and Michel, 1996; Ahmed et al., 2010; Michel, 2012, 2013; El Soufi and Michel 2014, 2015, 2016), and regulation of stop codon translation according to presence/absence of nearby stem-loop hairpins (Lobanov et al., 2010; Brown et al., 2015).

Superimposed informations include also alternative initiation codons (Kozak, 1984; Strubin et al., 1986; Kumar et al., 1991; Saris et al., 1991; Cao et al., 1995; Garceau et al., 1997; Suzek et al., 2001; Touriol et al., 2003; Seligmann, 2007; Lee et al., 2012; Vanderperre et al., 2013; Van Damme et al., 2014; Wan and Qian, 2014; De Klerk and 't Hoen, 2015; Young et al., 2016), termination codons (Suzeski et al., 1991; Harker et al., 1995; McCaughan et al., 1995; Poole et al., 1995; Martinez et al., 1999; Adachi and Cavalcanti, 2009; Seligmann, 2010a) shown by analyses of ribosomal stalling times, called ribosomal profiling (Brar and Weissman, 2015; Ingolia, 2016), and overlap coding after frameshifts (Fonseca et al., 2014; Mir and Schober, 2014; Saha et al., 2015, 2016).

Additional unknown encrypted information probably regulates non-canonical transcription/edition, such as RDDs (RNA-DNA differences, differences, Li et al., 2011; Chen et al., 2012; Chen and Bundschuh, 2012), systematic nucleotide exchanges (swinger RNAs, Seligmann, 2013a, 2013b), systematic nucleotide deletions (delRNAs, Seligmann, 2015a), and when alterations occur in the genetic code, for example due to suppressor (antitermination) tRNA translational activity (Dabrowski et al., 2015).

Here empirical proteomic analyses aim at strengthening evidence for translations of swinger and delRNAs, also according to alternative, stopless genetic codes (Swart et al., 2016). A general background on these processes is presented below. Note that some symmetries within the genetic code, such as Rumer's transformation that reveals a symmetry within the genetic code (Gonzalez, 2008; Jestin, 2010; Fimmel et al., 2013), can be seen as a type of superimposed information. These phenomena could be explained by swinger RNA and thereof

E-mail address: timonuslepidus@gmail.com.

http://dx.doi.org/10.1016/j.jtbi.2016.11.021
Received 13 April 2016; Received in revised form 11 November 2016; Accepted 22 November 2016
Available online 27 November 2016

translated peptides. Hence the confirmative empirical analyses presented here is to be appreciated in the context of genetic code symmetries and associated coding superimposition theory.

1.1. General background on alternative codings

1.1.1. Alternative genetic codes in mitochondria

Mitochondrial genomes apparently encode more proteins than usually assumed (Breton et al., 2014; Capt et al., 2016). Some of these genes assume translation of stop codons (Faure et al., 2011; Seligmann, 2011a, 2012a, 2012b, 2012c, 2015a; Barthélémy and Seligmann, 2016), according to stopless genetic codes (Swart et al., 2016). Indeed, predictions of mitochondrial peptides translated from +1 and −1 frameshifted protein coding genes coevolve with predicted suppressor tRNAs templated by the strand complementary to mitochondrial tRNAs (antisense tRNAs, Seligmann, 2010b, 2010c, 2010d, 2011b).

It seems that additional superimposed information is expressed by translation according to expanded codons, tetracodons (codons expanded by a 4th silent nucleotide), corresponding to systematic +1 frameshifts during translation. These predicted tetracoded genes coevolve with predictions of tRNAs with expanded anticodons (Seligmann, 2012d, 2013a, 2014a).

Re-analyses of previously published proteomic MS/MS data detected tetra- and pentacoded peptides (data from Gueugneau et al. (2014), reanalyzed by Seligmann (2015b, 2016a)). These observations extrapolate observations of translations of single, isolated expanded codons within otherwise regular proteins (O'Connor, 2003; Rodriguez et al., 2006; Atkins and Bjoerk, 2009; Baranov et al., 2015; Beznoskova et al., 2016), sometimes by tRNAs with expanded anticodons (e.g. Riddle and Carbon 1973; Tuohy et al., 1992; Landweber, 2002; Walker and Frederick, 2006; Dunham et al., 2007; Maehigashi et al., 2014).

Hypotheses about genetic code origins before ribosomal translation consider that tricodon-anticodon interactions without ribosome are too weak for peptide elongation, but not for expanded codons/anticodons (Baranov et al., 2009). Indeed, predicted tetracodons associate with high physiological temperatures in Lepidosauria, in line with the hypothesis that the fourth nucleotide stabilizes codon-anticodon interactions at conditions where tricodons are less efficient than tetracodons (Seligmann and Labra, 2013). Besides these thermodynamic considerations, symmetry and error correction principles predict 64 specific tetracodons, called tesserae, as origins of modern codons, specifically for the vertebrate mitochondrial genetic code (Gonzalez et al., 2012).

1.1.2. Alternative mitochondrial transcriptions: systematic deletions

Reanalyses of previously published human transcriptomic data (data from Garzon et al. (2014), reanalysis by Seligmann (2015b)) detect RNA transcripts matching systematic mono- and dinucleotide deletions after each transcribed nucleotide triplet of the human mitogenome. These delRNAs are covered by corresponding peptides detected in the previously mentioned human proteomic data (Seligmann, 2015b, 2016a; data from Gueugneau et al. (2014)).

Systematic deletions of nucleotides after each translated nucleotide triplet suggest that the same peptide can be obtained (a) by frameshifting translation of regular RNAs, either by systematic ribosomal slippages, or due to tRNAs with expanded codons; or by (b) regular translation of non-canonical RNAs where transcription systematically deletes/ignores mono- or dinucleotides after each transcribed trinucleotide, producing delRNAs.

The human mitogenome, assuming systematic mono- and dinucleotide deletion after every nucleotide triplet, includes more palindromes than corresponding randomized sequences. Results on palindromes indicate biological roles for sequences after systematic deletions, independently of previously mentioned transcriptomic and proteomic data. Palindromes seem very ancient mechanisms of RNA activity regulation (Seligmann and Raoult, 2016). Palindromes apparently downregulate delRNA occurrence (Seligmann, 2016b).

1.1.3. Alternative mitochondrial transcriptions: systematic nucleotide exchanges

Besides systematic deletions, systematic exchanges during transcription occur. Specific nucleotides are systematically exchanged by other specific nucleotides (nine symmetric exchanges, e.g. A < - > C (Seligmann, 2012e, 2013b, 2013c); and fourteen asymmetric exchanges, e.g. A- > C- > G- > T- > A (Seligmann, 2013d)). In mathematical contexts, systematic nucleotide exchanges are called bijective transformations (Fimmel et al., 2015a, 2015b, Gumbel et al., 2015), corresponding RNAs are 'swinger' RNAs.

Properties of swinger RNAs can be predicted from bijective transformations of the natural circular genetic code (Michel and Seligmann, 2014). The natural circular code is a set of twenty codons over-represented in coding versus non-coding (frameshifted) frames of genes (Arquès and Michel, 1996) with several strong mathematical properties enabling detection of coding frames (Ahmed et al., 2007, 2010). This is putatively due to interactions of the mRNA with rRNA (El Soufi and Michel, 2014, 2015) and tRNA (Michel, 2012, 2013). Observed abundances of mitochondrial swinger RNAs are proportional to relative invariance of the natural circular code after corresponding swinger/bijective transformations.

Swinger transcription seems due to regular RNA polymerases, because chimeric RNAs, part regular, part swinger-transformed have been detected. These are part regular and part swinger-transformed, where both parts, accounting for the swinger transformation, correspond to contiguous mitogenome regions (Seligmann, 2015c). Peptides matching translations of such contiguous regular and swinger RNA also exist, including peptides whose 'regular' part corresponds to classical mitochondrion-encoded proteins (Seligmann, 2016d).

Abundances and lengths of human mitochondrial swinger RNAs detected in GenBank's EST database converge with abundances and lengths of swinger RNA reads detected among reads produced by massive new generation Illumina sequencing techniques (data from Garzon et al., 2014, reanalyzed by Seligmann, 2016a). Hence swinger RNA occurrence was confirmed by two independent datasets sequenced each by a different sequencing technique. Note that at least one type of swinger DNA has also been detected in GenBank databases (Seligmann, 2014b, 2014c, 2015e).

Some swinger transformations correspond to symmetries in the genetic code (Jestin, 2010; Fimmel et al., 2013). Detections of peptides corresponding to these transformations might explain genetic code symmetries, in relation to multilayered superimposed protein-coding (Popov et al., 1996). Occurrences of peptides corresponding to these transformations stress the need to better understand these genetic code symmetries, as superimposed protein coding seems to be embedded within the genetic code's structure. These concepts are also in line with multilayered templating and encoding assumed for rRNA-like ancestral protogenomes (Root-Bernstein and Root-Bernstein, 2015, 2016).

1.1.4. Swinger transformations and the structure of the 'regular' mitogenome

The human mitogenome includes also swinger repeats: short sequences repeating other sequences, if a bijective transformation is applied to the repeated region. These swinger repeats are more numerous than observed for randomized sequences with equal length and nucleotide contents (Seligmann, 2015d). This result strengthens evidence for the existence of swinger polymerization(s), independently of observations of swinger RNAs.

The human mitogenome after swinger transformations includes more palindromes potentially forming secondary structures than corresponding randomly shuffled sequences. Non-random occurrence of swinger palindromes indicates biological roles for swinger-transformed sequences, independently of occurrences of swinger RNA/DNA and corresponding peptides. Palindromes upregulate swinger RNAs (Seligmann, 2016e), downregulate delRNAs (Seligmann, 2016b).