# Chimeric Translation for Mitochondrial Peptides: Regular and Expanded Codons

Herve Seligmann, Ganesh Warthi

# Chimeric translation for mitochondrial peptides: regular and expanded codons

Hervé Seligmann[1] and Ganesh Warthi[2,*]

[1]The National Natural History Collections, The Hebrew University of Jerusalem, 91404 Jerusalem, Israel. Email: hselig1@gmail.com

[2]Aix-Marseille University, IRD, VITROME, Institut Hospitalo-Universitaire Méditerranée-Infection, Marseille, France

*Corresponding author: g.warthi6791@gmail.com

Alphabetical author order

**Running head:** Chimeric peptides translated from regular and expanded codons

## Abstract

Frameshifting protein translation occasionally results from insertion of amino acids at isolated mono- or dinucleotide-expanded codons by tRNAs with expanded anticodons. Previous analyses of two different types of human mitochondrial MS proteomic data (Fisher and Waters technologies) detect peptides entirely corresponding to expanded codon translation. Here, these proteomic data are reanalyzed searching for peptides consisting of at least eight consecutive amino acids translated according to regular tricodons, and at least eight adjacent consecutive amino acids translated according to expanded codons. Both datasets include chimerically translated peptides (mono- and dinucleotide expansions, 42 and 37, respectively). The regular tricodon-encoded part of some chimeric peptides corresponds to standard human mitochondrial proteins (mono- and dinucleotide expansions, six (AT6, CytB, ND1, 2xND2, ND5) and one (ND1), respectively). Chimeric translation probably increases the diversity of mitogenome-encoded proteins, putatively producing functional proteins. These might result from translation by tRNAs with expanded anticodons, or from regular tricodon translation of RNAs where transcription/posttranscriptional edition systematically deleted mono- or dinucleotides after each trinucleotide. The pairwise matched combination of adjacent peptide parts translated from regular and expanded codons strengthens the hypothesis that translation of stretches of consecutive expanded codons occurs. Results indicate statistical translation producing distributions of alternative proteins. Genetic engineering should account for potential unexpected, unwanted secondary products.

**Key words:** delRNAs; non-canonical translation; non-canonical transcription; RNA editing; tRNA hopping.

## 1. Introduction

The low stability of codon-anticodon duplexes does not enable mRNA translation without a ribosome stabilizing this interaction long enough to enable peptide

elongation [1]. Ribosomes are very complex molecules [2-4] resulting from complex accretion histories [5-9], meaning that tRNA and tRNA-like structures accreted serially, becoming over time modern rRNA. A history derived from classical comparative analyses partly converges with an alternative, structure-based approach [10-13], from tRNAs [10, 14-20] and (perhaps) tRNA dimers [7, 21, 22]. Similar accretion histories also apparently produced other types of ancient (protein coding) genes [23]. Hence some kind of translation presumably occurred before ribosomes evolved, possibly including direct codon-amino acid interactions [24-31].

## 1.1. Expanded codons and anticodons

Ribosome-free translation seems impossible because codon-anticodon duplexes are too weak to enable peptide elongation. A simple potential solution to ancestral ribosome-free translation by codon-anticodon interactions assumes that modern codons are reduced. Ancestral ribosome-free translation presumably resulted from interactions between codons and anticodons consisting of more than three nucleotides [1, 32-34]. This stabilizes codon-anticodon duplexes long enough to enable ribosome-free peptide elongation (the ribosome also channels the spatial dynamics of aminoacylated tRNA-peptide interactions [35]). In addition, a code based on a specific subset of tetracodons (codons expanded by a fourth, silent nucleotide) would have self-correcting symmetry properties that could be an adequate ancestor of some modern (mitochondrial) genetic codes (the tessera hypothesis [36, 37]).

Natural tRNA-based translation of tetracodons was observed from the onset of molecular biology [38-40]. Modern ribosomes accommodate tRNAs with expanded anticodons during translation [41-43]. Biotechnological applications use tRNAs with expanded anticodons to introduce non-natural amino acids in proteins [44-48]. The antisense sequence of some mitochondrial tRNAs has predicted expanded anticodons [49,50]. The predicted mitochondrial tRNAs with expanded anticodons coevolve with predicted mitochondrial tetracodon-encoded peptides [51-54]. Tetracoding in modern organisms seems to be an adaptation to high temperatures where regular codon-anticodon interactions become relatively unstable: predicted tetracoding increases with lepidosaurian body temperature [55]. This would be a relict of prebiotic ribosome-free translation and genetic code formation, which likely occurred at high temperatures [56-60]; but see [61] for some counter-arguments to life's thermophilic origin hypothesis.

## 1.2. Peptides coded by expanded codons

In addition to tetracoding sequences predicted by alignment methods, peptides corresponding in their entireties to the translation of tetra- and pentacodons have been detected in MS data (produced by the medium accuracy Thermo Fisher (Illkirch, France), and the high accuracy Waters (Milford, MA, USA) technologies) from the human mitochondrial peptidome [62-67]. Detections of complete tetracoded peptides differ from occurrences of isolated tetracodons, notably in mitochondrial genomes

[68-69], which might result from decoding by specific expanded anticodons or as regular codons after deletion of extra nucleotide(s) by post-transcriptional editing [70].

## 1.3. Alternative translations

Translation according to expanded codons by a series of expanded anticodons is not the only known alternative translation. Antitermination, or stop suppression occurs due to decoding by regular near-cognate tRNAs [71-75] or tRNAs with an anticodon matching a stop codon [76-80]. Natural stop-suppressor tRNAs have also been adjusted for genetic code expansion to insert non-natural amino acids in proteins in biotechnological applications [81-83]. Natural stop suppression in mitogenes can be predicted based on alignment analyses [52,84-91], and observed distributions of amino acids inserted at stops [62-67] match genetic code evolution [93-95] and coding symmetries in the genetic code [96].

## 1.4. Alternative transcriptions

Post-transcriptional editing by systematically deleting every fourth, or every fourth and fifth nucleotide after each transcribed nucleotide triplet could produce noncanonical transcripts whose regular transcription matches noncanonical translation of regular transcripts according to tetra- or pentacodons. Mitochondrial transcripts detected in several independent datasets produced by independent sequencing technologies (Sanger and Illumina) match sequences predicted by systematic deletions after each transcribed nucleotide triplet. These noncanonical transcripts (delRNAs), because of their length corresponding to numerous tricodons separated by deleted mono- or dinucleotides, seem more likely produced by noncanonical transcription systematically deleting mono- or dinucleotides than by posttranscriptional edition, though the latter cannot be excluded [97]. These delRNAs have more than expected homopolymers [98] that frequently induce frameshifting [99].

Note that the human mitogenome, assuming systematic deletion of mono- or dinucleotides after each transcribed nucleotide triplet, includes more palindromes than random sequences with the same length and nucleotide content [100].

A different type of noncanonical transcription exists, consisting of systematic exchanges between nucleotides, producing swinger RNAs that do not resemble their template DNA unless considering the transformation rule that produced the noncanonical RNA. Nine systematic nucleotide exchanges are symmetric (X<->Y, example G<->T [87,88,90], and fourteen are asymmetric (X->Y->Z->X, example C->T->G->C [89]). Empirical genomic coverages by swinger RNAs are replicable across independent datasets [98] and sequencing techniques (Sanger and Illumina, human mitogenomes [63]; 454 and SOLID, Mimivirus [102]). The human mitogenome, assuming systematic nucleotide exchanges, includes more palindromes forming stem-loop hairpins than random sequences with the same length and

nucleotide contents [103]. Systematic nucleotide exchanges conserve error-correcting properties of the genetic code and its embedded circular code regulating ribosomal translation frame [104-107].

1.5. Chimeric RNAs and peptides

Note that swinger DNA has also been reported [108,109], in one case with abrupt switches between the regular and the swinger-transformed part of the mitogenome [110]. Chimeric RNAs, partly corresponding to regular transcription of the human mitogenome, and partly corresponding to swinger-transcription of adjacent parts of the mitogenome, also occur, including abrupt switches between these two parts [111]. These either result from regular transcription of genomic swinger DNA, or from swinger transcription of part of the template mitogenome. Chimeric peptides corresponding to translation of adjacent parts of regular and swinger-transformed RNA exist in mitochondrial proteomic data, including peptides whose regular part corresponds to mitochondrion-encoded proteins, for example CytB [112].

These chimeric molecules are strong evidence for swinger phenomena. First because if they reflected unknown artefacts, these would not have produced the regular parts of the RNA and peptide sequences. Chimeric RNAs and peptides show that unknown phenomena producing variants of known RNAs and proteins exist. In addition, the regular parts of the chimeric RNA/peptide are natural matched positive controls for adjacent noncanonical parts. This strengthens the confidence in the biological reality of these noncanonical phenomena.

In the context of long stretches of translation according to expanded codons, chimeric peptides corresponding in part to regular translation, and in adjacent parts to translation according to expanded codons (Figure 1), would also consist strong evidence for translation of stretches of expanded codons, and indicate that variants of known proteins including parts encoded by expanded codons exist. Hence here we present analyses of two mitochondrial MS datasets (one produced by Thermo Fisher and one by Waters technologies) that explore for chimeric peptides resulting from translation according to adjacent stretches of regular and expanded codons.

These analyses presume that translation produces a distribution of alternative protein products, some might present functional advantages, for example by having functional optima at conditions that differ from those of known canonical proteins (for example temperature). The approach implies caution: genetic engineering should account for potential unwanted byproducts resulting from little known and/or unknown alternative transcriptions and translations: unlike engineered genes, natural genes adapted to avoid or minimize disruptive effects by proteins resulting from rarer noncanonical transcriptions and/or translations.

**2. Materials and Methods**

The data and analytical methods are identical to those previously used for chimeric swinger peptides [112]. The human mitogenome (NC_012920, length 16569 basepairs) was downloaded in its entirety.

2.1. Predicted peptides

The difference with previous analyses consists in the fact that analyses compare observed MS/MS data with hypothetical peptides that result in part from canonical, and from noncanonical translations for consecutive, adjacent stretches of amino acids. The design of these hypothetical chimeric peptides uses two sizes of running windows: one for tetracodons (codons expanded by one nucleotide) and one for pentacodons (codons expanded by two nucleotides). Each running window codes for 30 amino acids according to regular tricodons, 30 consecutive amino acids coded according to noncanonical codons expanded by a mono- or dinucleotide, and another consecutive 30 amino acids coded according to regular tricodons. Hence each hypothetical chimeric peptide consists of 30 canonically coded amino acids at each its 5' and 3' extremities, and 30 noncanonical codons, each expanded by a mono- or dinucleotide. This produces window sizes of (3x30)+(4x30)+(3x30) = 90+120+90=300 nucleotides for tetracodons and 90+150+90=330 nucleotides for pentacodons. Windows move by steps of single nucleotides along the complete genome, producing 2x16569 theoretical chimeric tetra-, and 2x16569 chimeric pentacoded peptides, for + and – mitogenome strands.

For hypothetical chimeric peptides, the length of 30 consecutive amino acids with the same translation modus was chosen because previous experience shows that most detected peptides are shorter than 30 amino acids. Adopting shorter lengths, for example 15 amino acids, for the different parts of the hypothetical chimeric peptide, would not detect chimeric peptides with canonical and/or noncanonical parts longer than 15 amino acids.

For each hypothetical peptide, the relevant proteome analysis software predicts a theoretical mass spectrometry distribution. This distribution is compared with observed MS/MS data.

2.2. Translation of stop codons

Peptides translated from sequences including stop codons are included 19 times in the pool of theoretical peptides, each time inserting at all stop codons the same amino acid (there are 20 amino acids, but leucine and isoleucine have equal masses and hence cannot be distinguished by MS/MS techniques, resulting in 19 alternative peptides where a different amino acid species is inserted at all stops). Hence analyses consider the possibility that each amino acid can translate stop codons. Approximately 2x16569x19=629622 chimeric peptides exist for each tetra- and pentacoded chimeric translations.

2.3. Medium accuracy data searches

For the MS/MS data from [113], consensus searches between observed and predicted MS/MS data were handled with the Sequest (Thermo Fisher Scientific, Illkirch) algorithm with the following mass tolerances: Parent = 1 Da and Fragment = 0.5 Da (monoisotopic masses). Fixed carbamidomethyl (C) and variable Oxydation (M) modifications were activated, as well as the lysine → pyrrolysine modification, and only one missed trypsin cleavage was allowed. False discovery rate was estimated against a reverse decoy database using the Percolator algorithm. No protein grouping was allowed since the database only contained non-redundant entries. Peptides with false discovery rate q < 0.05 and score Xcorr > 1.99 were considered identified. The score Xcorr is a likelihood of match between expected and observed MS/MS data unaffected by peptide length. The false discovery rate q is adapted to populations of detected peptides [114].

Previous analyses show that searches of the trypsinized proteome produced by [113] detect much more peptides when analyzing data separately for peptides cleaved at K, and separately for those cleaved at R, and when searches assume cleavage at the amino extremity of these amino acids rather than the carboxyl extremity at which trypsin cleaves proteic chains. These searches mainly detect peptides ending at the carboxyl extremity of K or R, corresponding to the experimental trypsin cleavage by assuming one missed cleavage. This increased efficiency of searches remains unexplained, but is not due to artifacts, because noncanonical peptides detected by all searches (assuming cleavage at any amino acid and at any (carboxyl and amino) extremities) map with comparable rates on corresponding, detected noncanonical RNAs, and correspond in their overwhelming majorities to trypsinized peptides as expected by the experimental parameters (addition of trypsin) [66].

2.4. High accuracy data searches

The same pool of theoretical chimeric peptides was used for PLGS searches of another human mitoproteome dataset [115], extracted by a higher accuracy method (Waters, Milford, MA). Mass peak estimates are more accurate (5 ppm for data extracted by [115], versus 0.5 Da for data by [113]). Precise comparison of accuracies between these techniques is unfeasible: Sequest (Thermo Fisher Scientific, Illkirch) uses fixed cutoffs; PLGS adapts cutoffs to masses of detected peptide: 0.5 Da in the latter sample would occur for peptides with mass $5 \times 10^6 \times 0.5$ Da.

The twelve samples from [115] were processed using ProteinLynx Global Server version 3.0.1 (Waters, Saint-Quentin En Yvelines, France). Processing parameters were 250 counts for the low energy threshold, 100 counts for the elevated energy threshold and 750 counts for the intensity threshold. Hits are considered significant according to standard criteria, with PLGS peptide score 6.49. This score is compared to a decoy database to estimate FDR, as done for Xcorr from the dataset produced by [113], and peptides with q<0.05 are retained.

Each sample was searched separately for peptides 38 times, each search assuming cleavage at a different extremity (carboxyl or amino) of each amino acid species (merging L and I, 2x19=38).

2.5. Minimal size of detected chimeric peptides

Detected peptides are further filtered so as to retain only peptides with at least eight consecutive amino acids coded according to regular codons, and at least eight consecutive amino acids coded according to noncanonical expanded codons. This size was determined so that each regular- and noncanonically-encoded parts of the chimeric peptide have an approximate maximal e value 0.0014 ($629622x1/19^8$).

**3. Results and discussion**

3.1. Chimeric peptides with a tetra- or pentacoded part

Tables 1 and 2 present 28 chimeric peptides detected in the MS/MS data published by [113] and 14 chimeric peptides detected in the MS/MS data published by [115], with at least eight amino acids coded by tricodons and eight adjacent amino acids coded by tetracodons. Tables 1 and 2 also present 19 chimeric peptides detected in the MS/MS data published by [113] and 18 chimeric peptides detected in the MS/MS data published by [115], with at least eight amino acids coded by tricodons and eight adjacent amino acids coded by pentacodons.

In about half the cases, the noncanonical part of the peptide is on the 5' extremity of the peptide, not for tetra- and not for pentacoded parts, nor for any dataset. The noncanonical part can either be at the 5'- or the 3'-encoded extremity of chimeric peptides, with no apparent bias for one of these extremities in the current results.

3.2. Chimeric peptides integrated in regular mitogenome-encoded proteins

Most peptides include amino acids inserted at stop codons, in each regular-encoded and noncanonical parts. The regular-encoded part of a total of seven chimeric peptides corresponds to one among the 13 classical mitogenome-encoded proteins. Six among these have adjacent tetracoded parts, and one an adjacent pentacoded part. The proteins are: AT6, CytB, ND1, ND2 (two different peptides), ND5 (tetracoded) and ND1 (pentacoded). These small numbers do not enable to test whether biases exist in terms of which proteins tend to include more or less noncanonical parts, nor in relation to their position on the mitogenome, as these genes are scattered across the whole mitochondrial operon. The hypothesis that noncanonical peptides result from mitochondrial polymorphisms and heteroplasmy [116-121] is unlikely: their exact correspondence to sequences predicted by translation of expanded codons excludes this option. As a group, they result from regular mitogenomic DNA variability.

3.3. Noncanonical transcription or translation?

Above results confirm that tetra- and pentacoded of amino acid stretches occur, conjugated with regular encoded stretches of amino acids. The alternative, that chimeric peptides originate from regular translation of chimeric RNAs produced in part by regular transcription, and in part by noncanonical transcription systematically deleting mono- or dinucleotides after each transcribed trinucleotide, cannot be excluded. The data at hand don't enable to test between these two alternatives potentially producing identical peptides. We tentatively presume that both mechanisms are at work because expanded codons and anticodons have been previously reported, and because noncanonical transcripts corresponding to transcription systematically deleting mono- and dinucleotides also exist.

## 3.4. Adaptive diversity

Amino acid stretches encoded by noncanonical codons (or resulting from noncanonical transcription) might be integrated in regular mitogenome-encoded proteins, as suggested by their association with stretches of tricodon-encoded amino acids clearly corresponding to regular membrane-bound mitochondrial proteins. The possibility that these chimeric peptides are part of functional proteins cannot be excluded. The existence of chimeric peptides suggests that natural protein diversity can be increased by mixing types of decoding processes, such as regular tricodons and noncanonical codons expanded by one or two nucleotides. This diversity might have unknown adaptive/functional components, including widening ranges of functionally optimal conditions at which some metabolic activities might occur.

Results also stress that natural translation of expanded codons is not extremely rare. The hypothesis that expanded codons (but not systematic deletions) are adaptive at high temperatures could be tested by comparing abundances of detected peptides coded by expanded codons at different temperatures, expecting more translation according to expanded codons at higher temperatures. Other analyses searching for peptides corresponding to codons expanded by more nucleotides (>2) will also contribute to our understanding of these noncanonical transcriptions and translations that increase the coding potential of sequences.

## 3.5. Unwanted effects of genetic engineering

Experiments and analyses exploring for which genome regions undergo the different noncanonical transcriptions and translations in which cell types and under which conditions would deepen our understanding of cell metabolism, implying likely biomedical applications. Results also stress that genetic engineering should explore potential effects of proteins produced from noncanonical transcripts and/or by noncanonical translations, to avoid undesirable effects from discarded noncanonical processes such as swinger and del-transcriptions, and translation of stop codons and according to expanded codons.

## Acknowledgement

## References

[1] Baranov PV, Venin M, Provan G 2009 Codon size reduction as the origin of trhe triplet genetic code. PLoS One 4, e5708.

[2] Ramakrishnan V 2014 The ribosome emerges from a black box. Cell 159, 979-984.

[3] Noeske J, Wasserman MR, Terry DS, ARB, Blanchard SC, Cate JH 2015 high-resolution structure of the *Escherichia coli* ribosome. Nat Struct Mol Biol 22, 336-341.

[4] Yusupova G, Yusupov M 2017 Crystal structure of eukaryotic ribosome and its complexes with inhibitors. Philos trans R Soc Lond B Biol Sci 372, pii:20160184.

[5] Harish A, Caetano-Anollés G 2012 Ribosomal history reveals origins of modern protein synthesis. PLoS One 7, e32776.

[6] Caetano-Anollés G, Caetano-Anollés D 2017 Commentary: History of the ribosome and the origin of translation. Front Mol Bioswci 3, 87.

[7] Farias ST, Rêgo TG, José MV 2014 Origin and evolution of the peptidyl transferase center from proto-tRNAs. FEBS Open Bio 4, 175-178.

[8] Caetano-Anollés D, Caetano-Anollés G 2016a Piecemeal buildup of the genetic code, ribosomes, and genomes from primordial tRNA building blocks. Life (Basel) 6, e43.

[9] Caetano-Anollés G, Caetano-Anollés D 2015 Computing the origin and evolution of the ribosome from its structure - Uncovering processes of macromolecular accretion benefiting synthetic biology. Comput Struct Biotechnol J 13, 427-447.

[10] Farias ST, Rêgo TG, José MV 2019 Origin of the 16S ribosomal molecule from ancestor tRNAs. Sci 1, 8.

[11] Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV, Williams LD 2015 History of the ribosome and the origin of translation. Proc Natl Acad Sci U S A 112, 15396-15401.

[12] Gulen B, Petrov AS, Okafor CD, Vander Wood D, O'Neill EB, Hud NV, Williams LD 2016 Ribosomal small subunit domains radiate from a central core. Sci Rep 6, 20885.

[13] Lanier KA, Athavale SS, Petrov AS, Wartell R, Williams LD 2016. Imprint of ancient evolution on rRNA folding. Biochemistry 55, 4603–4613.

[14] Bloch DP, McArthur B, Guimarães RC, Smith J, Staves MP 1989 tRNA-rRNA sequence matches from inter- and intraspecies comparisons suggest common origins for the two RNAs. Braz J Med Biol Res 22, 931-944.

[15] Bloch DP, McArthur B, Mirrop S 1985 RNA-rRNA sequence homologies: evidence for an ancient modular format shared by tRNAs and rRNAs. Biosystems 17, 209-225.

[16] Bloch DP, McArthur B, Widdowson R, Spector D, Guimaraes RC, Smith J 1983 tRNA-rRNA sequence homologies: evidence for a common evolutionary origin? J Mol Evol 19, 420-428.

[17] Bloch D, McArthur B, Widdowson R, Spector D, Guimaraes RC, Smith J 1984 tRNA-rRNA sequence homologies: a model for the origin of a common ancestral molecule, and prospects for its reconstruction. Orig Life 14, 571-578.

[18] Caetano-Anollés G 2002 Tracing the evolution of RNA structure in ribosomes. Nucleic Acids Res 30, 2575-2587.

[19] Caetano-Anollés G, Sun FJ 2014 The natural history of transfer RNA and its interactions with the ribosome. Front Genet 5, 127.

[20] Barthélémy RM, Seligmann H 2016 Cryptic tRNAs in chaetognath mitochondrial genomes. Comput Biol Chem 62, 119-132.

[21] Agmon IC 2016 Could a proto-ribosome emerge spontaneously in the prebiotic world? Molecules 21, e1701.

[22] Guimarães CR 2017 Self-referential encoding on modules of anticodon pairs-roots of the biological flow system. Life (Basel) 7, e16.

[23] Demongeot J, Seligmann H 2019 More pieces of ancient than recent theoretical minimal proto-tRNA-like RNA rings in genes coding for tRNA synthetases. J Mol Evol, in press.

[24] Pelc SR 1965 Correlation between coding-triplets. Nature 207, 597-599.

[25] Pelc SR, Welton MG 1966 Stereochemical relationship between coding triplets and amino-acids. Nature 209, 868-870.

[26] Welton MG, Pelc SR 1966 Specificity of the stereochemical relationship between ribonucleic acid triplets and amino-acids. Nature 209, 870-872.

[27] Yarus M, Widmann JJ, Knight R 2009 RNA-amino acid binding: a stereochemical era for the genetic code. J Mol Evol 69, 406-429.

[28] de Ruiter A, Zagrovic B 2015 Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. Nucleic Acids Res 43, 708-718.

[29] Bartonek L, Zagrovic B 2017 2017 mRNA/protein sequence complementarity and its determinants: the impact of affinity scales. PLoS Comput Biol 13, e1005648.

[30] Yarus M 2017 The genetic code and RNA-amino acid affinities. Life (Basel) 7, e13.

[31] Chatterjee S, Yadav S 2019 The origin of prebiotic information system in the peptide/RNA world: a simulation model of the evolution of translation and the genetic code. Life (Basel) 9, e25.

[32] Landweber LF 2002 Custom codons come in threes, fours and fives. Chem Biol 9, 143.

[33] Di Giulio M, Moracci M, Cobucci-Ponzano B 2014 RNA editing and modifications of RNAs might have favoured the evolution of the triplet genetic code from an ennuplet code. J Theor Biol 359, 1-5.

[34] Di Giulio M 2015 A model for the origin of the first mRNAs. J Mol Evol 81, 10-17.

[35] Seligmann H, Warthi G 2017 Genetic code optimization for cotranslational protein folding: codon directional asymmetry correlates with antiparallel betasheets, tRNA synthetase classes. Comput Struct Biotechnol J 15, 412-424.

[36] Gonzalez DL, Giannerini S, Rosa R. 2012 On the origin of the mitochondrial genetic code: towards a unified mathematical framework for the management of genetic information. Available from Nature Precedings. See http://dx.doi.org/10.1038/npre.2012.7136.1.

[37] Gonzalez DL, Giannerini S, Rosa R. 2016 The non-power model of the genetic code: a paradigm for interpreting genomic information. Phil Trans Roy Soc A Math Phys Eng Sci 374, 20150062.

[38] Riddle DL, Carbon J 1973 Frameshift suppression – nucleotide addition in anticodon of glycine transfer-RNA. Nature – New Biol 242, 230-234.

[39] Tuohy TM, Thompson S, Gesteland RF, Atkins JF 1992 Seven, eight and nine-membered anticodon loop mutants of tRNA(2Arg) which cause + 1 frameshifting. Tolerance of DHU arm and other secondary mutations. J Mol Biol 228, 1042–1054.

[40] Phelps SS, Gaudin C, Yoshizawa S, Benitez C, Fourmy D, Joseph S 2006 Translocation of a tRNA with an extended anticodon through the ribosome. J Mol Biol 360, 610-622.

[41] Walker SE, Fredrick K 2006 Recognition and positioning of mRNA in the ribosome by tRNAs with expanded anticodons. J Mol Biol 360, 599-609.

[42] Dunham CM, Selmer M, Phelps SS, Kelley AC, Suzuki T, Joseph S, Ramakrishnan V 2007 Structures of tRNAs with an expanded anticodon loop in the decoding center of the 30S ribosomal subunit. RNA 13, 817-823.

[43] Maehigashi T, Dunkle JA, Miles SJ, Dunham CM 2014 Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops. Proc Natl Acad Sci U S A 111, 12740-12745.

[44] Moore B, Persson BC, Nelson CC, Gesteland RF, Atkins JF 2000 Quadruplet codons: Implications for code expansion and the specification of translation step size. J Mol Biol 298, 195-209.

[45] Magliery TJ, Anderson JC, Schultz PG 2001 Expanding the genetic code: Selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in *Escherichia coli*. J Mol Biol 307, 755-769.

[46] Hendrickson TL, de Crecy-Lagard V, Schimmel P 2004 Incorporation of nonnatural amino acids into proteins. Ann Rev Biochem 73, 147-176.

[47] Wang NX, Shang X, Cerny R, Niu W, Guo J 2016 Systematic evolution and study of UAGN decoding tRNAs in a genomically recoded bacteria. Sci Rep 6, 21898.

[48] Arranz-Gibertt P, Vanderschurent K, Isaacs FJ 2018 Next-generation genetic code expansion. Curr Op Chem Biol 46, 203-211.

[49] Seligmann H 2010a Undetected antisense tRNAs in mitochondrial genomes? Biol Direct 5, 39.

[50] Seligmann H 2011a Pathogenic mutations in antisense mitochondrial tRNAs. J Theor Biol 269, 287-296.

[51] Seligmann H 2012b Overlapping genetic codes for overlapping frameshifted genes in Testudines, and *Lepidochelys olivacea* as special case. Comput Biol Chem 41, 18-34.

[52] Seligmann H 2012c An overlapping genetic code for frameshifted overlapping genes in Drosophila mitochondria: antisense antitermination tRNAs UAR insert serine. J Theor Biol 298, 51-76.

[53] Seligmann H 2013a Pocketknife tRNA hypothesis: anticodons in mammal mitochondrial tRNA side-arm loops translate proteins? Biosystems 113, 165-176.

[54] Seligmann H 2014a Putative anticodons in mitochondrial tRNA sidearm loops: Pocketknife tRNAs? J Theor Biol 340, 155-163.

[55] Seligmann H, Labra A 2013 Tetracoding increases with body temperature in Lepidosauria. Biosystems 114, 155-163.

[56] Di Giulio M 2000 The late stage of genetic code structuring took place at a high temperature. Gene 261, 189-95.

[57] Di Giulio M 2003a The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. J Theor Biol 221, 425-36.

[58] Di Giulio M 2003b The universal ancestor and the ancestor of bacteria were hyperthermophiles. J Mol Evol 57, 721-30.

[59] Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A 2013 Experimental evidence for the thermophilicity of ancestral life. Proc Natl Acad Sci U S A 110, 11067-11072.

[60] Sleep NH 2018 Geological and geochemical constraints on the origin and evolution of life. Astrobiology 18, 1199-1219.

[61] Gutfraind A, Kempf A 2008 Error-reducing structure of the genetic code indicates code origin in non-thermophile organisms. Orig Life Evol Biosph 38, 75-85.

[62] Seligmann H 2015a Codon expansion and systematic transcriptional deletions produce tetra-, pentacoded mitochondrial peptides. J Theor Biol 387, 154-165.

[63] Seligmann H 2016a Translation of mitochondrial swinger RNAs according to tri-, tetra- and pentacodons. Biosystems 140, 38-48.

[64] Seligmann H 2016b Natural chymotrypsin-like-cleaved human mitochondrial peptides confirm tetra-, pentacodon, non-canonical RNA translations. Biosystems 147, 78-93

[65] Seligmann H 2016c Unbiased mitoproteome analyses confirm non-canonical RNA, expanded codon translations. Comput Struct Biotechnol J 14, 391-403.

[66] Seligmann H 2017a Natural mitochondrial proteolysis confirms transcription systematically exchanging/deleting nucleotides, peptides coded by expanded codons. J Theor Biol 414, 76-90.

[67] Seligmann H 2017b Reviewing evidence for systematic transcriptional deletions, nucleotide exchanges, and expanded codons, and peptide clusters in human mitochondria. Biosystems 160, 10-24.

[68] Beckenbach AT, Robson SK, Crozier RH 2005 Single nucleotide +1 frameshifts in an apparently functional mitochondrial cytochrome b gene in ants of the genus *Polyrhachis*. J Mol Evol 60, 141-152.

[69] Russell RD, Beckenbach AT 2008 Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. J Mol Evol 67, 682-695.

[70] Bar-Yaacov D, Avital G, Levin L, Richards AL, Hachen N, Rebolledo Jaramillo B, Nekrutenko A, Zarivach R, Mishmar D 2013 RNA-DNA differences in human mitochondria restore ancestral form of 16S ribosomal RNA. Genome Res 23, 1789-1796.

[71] O'Connor M 1998 tRNA imbalance promotes -1 frameshifting via near-cognate decoding. J Mol Biol 279, 727-736.

[72] Seligmann H 2010b Do anticodons of misacylated tRNAs preferentially mismatch codons coding for the misloaded amino acid? BMC Mol Biol 11, 41.

[73] Seligmann H 2011b Error compensation of tRNA misacylation by codon-anticodon mismatch prevents translational amino acid misinsertion. Comput Biol Chem 35, 81-95.

[74] Seligmann H 2012d Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. Curr Genomics 13, 37-54.

[75] O'Donoghue P, Prat L, Heinemann IU, Ling J, Odoi K, Liu WR, Söll D 2012 Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNAPyl for genetic code expansion. FEBS Lett 586, 3931-3937.

[76] Beier H, Grimm M 2001 Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res 29, 4767-4782.

[77] Wang R, Xiong J, Wang W, Miao W, Liang A 2016a High frequency of+1 programmed ribosomal frameshifting in *Euplotes octocarinatus*. Sci Rep 6, 21139.

[78] Wang R, Zhang Z, Du J, Fu Y, Liang A 2016b Large-scale mass spectrometry-based analysis of *Euplotes octocarinatus* supports the high frequency of +1 programmed ribosomal frameshift. Sci Rep 6, 33020.

[79] Massey SE 2017 The identities of stop codon reassignments support ancestral tRNA stop codon decoding activity as a facilitator of gene duplication and evolution of novel function. Gene 619, 37-43.

[80] Tharp JM, Ehnbom A, Liu WR 2018 tRNAPyl: Structure, function, and applications. RNA Biol 15, 441-452.

[81] Gan Q, Fan C 2017 Increasing the fidelity of noncanonical amino acid incorporation in cell-free protein synthesis. Biochim Biophys Acta Gen Subj 1861B, 3047-3052.

[82] Brabham R, Fascione MA 2017 Pyrrolysine amber stop-codon suppression: development and applications. Chembiochem 18, 1973-1983.

[83] Tharp JM, Liu WR 2018 Using amber and ochre nonsense codons to code two different noncanonical amino acids in one protein gene. Methods Mol Biol 1728:147-154.

[84] Seligmann H 2011c Two genetic codes, one genome: frameshifted primate mitochondrial genes code for additional proteins in presence of antisense antitermination tRNAs. Biosystems 105, 271-285.

[85] Faure E, Delaye L, Tribolo S, Levasseur A, Seligmann H, Barthélémy RM 2011 Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene. Biol Direct 6, 56.

[86] Seligmann H 2012a Putative mitochondrial polypeptides coded by expanded quadruplet codons, decoded by antisense tRNAs with unusual anticodons. Biosystems 110, 84-106.

[87] Seligmann H 2012e Overlapping genes coded in the 3'-to-5'-direction in mitochondrial genes and 3'-to-5' polymerization of non-complementary RNA by an 'invertase'. J Theor Biol 315, 38-52.

[88] Seligmann H 2013b Polymerization of non-complementary RNA: systematic symmetric nucleotide exchanges mainly involving uracil produce mitochondrial RNA transcripts coding for cryptic overlapping genes. Biosystems 111, 156-174.

[89] Seligmann H 2013c Systematic asymmetric nucleotide exchanges produce human mitochondrial RNAs cryptically encoding for overlapping protein coding genes. J Theor Biol 324, 1-20.

[90] Seligmann H 2013d Triplex DNA:RNA, 3'-to-5' inverted RNA and protein coding in mitochondrial genomes. J Comput Biol 20, 660-671.

[91] Seligmann H 2018c Directed mutations recode mitochondrial genes: from regular to stoplessgenetic codes. Chapter in: Mitochondrial DNA, Seligmann H and Warthi G eds, InTechOpen.

[92] Seligmann H 2015b Phylogeny of genetic codes and punctuation codes within genetic codes. Biosystems 129, 36-43.

[93] Massey SE, Garey JR 2007 A comparative genomics analysis of codon reassignments reveals a link with mitochondrial proteome size and a mechanism of genetic code change via suppressor tRNAs. J Mol Evol 64, 399-410.

[94] Seligmann H 2015b Phylogeny of genetic codes and punctuation codes within genetic codes. Biosystems 129, 36-43.

[95] Seligmann H 2018a Alignment-based and alignment-free methods converge with experimental data on amino acids coded by stop codons at split between nuclear and mitochondrial genetic codes. Biosystems 167, 33-46.

[96] Seligmann H 2018b Bijective codon transformations show genetic code symmetries centered on cytosine's coding properties. Theory Biosci 137, 17-31.

[97] Warthi G, Seligmann H 2019 Transcripts with systematic nucleotide deletion of 1-12 nucleotide in human mitochondrion suggest potential non-canonical transcription. PLoS One 14, e0217356.

[98] Warthi G, Seligmann H 2018 Swinger RNAs in the human mitochondrial transcriptome. Chapter in: Mitochondrial DNA, Seligmann H and Warthi G eds, InTechOpen.

[99] El Houmami N, Seligmann H 2017 Evolution of nucleotide punctuation marks: from structural to linear signals. Front Genet 8, 36.

[100] Seligmann H 2019 Localized context-dependent effects of the "ambush" hypothesis: more off-frame stop codons downstream of shifty codons. DNA Cell Biol, in press.

[101] Seligmann H 2016e Systematically frameshifting by deletion of every 4th or 4th and 5th nucleotides during mitochondrial transcription: RNA self-hybridization regulates delRNA expression. Biosystems 142-143, 43-51.

[102] Seligmann H, Raoult D 2018 Stem-loop RNA hairpins in giant viruses: invading rRNA-like repeats and a template free RNA. Front Microbiol 9, 101.

[103] Seligmann H 2016d Swinger RNA self-hybridization and mitochondrial non-canonical swinger transcription, transcription systematically exchanging nucleotides. J Theor Biol 399, 84-91.

[104] Michel CJ, Seligmann H 2014 Bijective transformation circular codes and nucleotide exchanging RNA transcription. Biosystems 118, 39-50.

[105] Fimmel E, Danielli A, Strüngmann L 2013 On dichotomic classes and bijections of the genetic code. J Theor Biol 336, 221-230.

[106] Fimmel E, Giannerini S, Gonzalez DL, Strüngmann L 2015a Circular codes, symmetries and transformations. J Math Biol 70, 1623-44.

[107] Fimmel E, Giannerini S, Gonzalez DL, Strüngmann L 2015b Dinucleotide circular codes and bijective transformations. J Theor Biol 386, 159-65.

[108] Seligmann H 2014b Species radiation by DNA replication that systematically exchanges nucleotides? J Theor Biol 363, 216-222.

[109] Seligmann H 2014c Mitochondrial swinger replication: DNA replication systematically exchanging nucleotides and short 16S ribosomal DNA swinger inserts. Biosystems 125, 22-31.

[110] Seligmann H 2015c Sharp switches between regular and swinger mitochondrial replication: 16S rDNA systematically exchanging nucleotides A<->T+C<->G in the mitogenome of *Kamimuria wangi*. Mitochondrial DNA A DNA Mapp Seq Anal 27, 2440-2446.

[111] Seligmann H 2015d Swinger RNAs with sharp switches between regular transcription and transcription systematically exchanging ribonucleotides: Case studies. Biosystems 135, 1-8.

[112] Seligmann H 2016f Chimeric mitochondrial peptides from contiguous regular and swinger RNA. Comput Struct Biotechnol J 14, 283-297.

[113] Gueugneau M., Coudy-Gandilhon C., Gourbeyre O., Chambon C., Combaret L., Polge C 2014 Proteomics of muscle chronological ageing in post-menopausal women. BMC Genomics 15, 1165.

[114] Käll L, Storey JD, MacCoss MJ, Noble WS 2008 Posterior error probabilities and false discovery rates: two sides of the same coin. J Proteome Res 7,40–44.

[115] Alberio T, Bondi H, Colombo F, Alloggio I, Pieroni L, Urbani A 2014 Mitochondrial proteomics investigation of a cellular model of impaired dopamine homeostasis, an early step in Parkinson's disease pathogenesis. Mol Biol Syst 10, 1332-1344.

[116] Wallace DC 1994 Mitochondrial DNA sequence variation in human evolution and disease. Proc Natl Acad Sci U S A 91, 8739–8746.

[117] He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N 2010 Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature 464, 610–614.

[118] Avital G, Buchshtav M, Zhidkov I, Feder J, Dadon S, Rubin E, Glass D, Spector TD, Mishmar D 2012 Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. Hum Mol Genet 21, 4214–4224.

[119] Ramos A, Santos C, Mateiu L, Mdel M, Alvarez L, Azevedo L, Amorim A, Aluja MP 2013 Frequency and pattern of heteroplasmy in the complete human mitochondrial genome. PLoS One 8, e74636.

[120] Ross JM, Stewart JB, Hagstrom E, Brene S, Mourier A, Coppotelli G, Freyer C, Lagouge M, Hoffer BJ, Olson L, Larsson NG 2013 Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. Nature 501, 412–415.

[121] Ye K, Lu J, Ma F, Keinan A, Gu Z 2014 Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. Proc Natl Acad Sci U S A 111, 10654–10659.

Figure 1. RNA sequence and its chimeric translation according to regular tricodons and tetra- and pentacodons. Sequences corresponding to 90 codons (two groups of 30 regular tricodons, each at the 5' and 3' extremity of a group of 30 noncanonical codons expanded by mono- or dinucleotides (tetra- and pentacodons)) form running windows of 90+120+90=300 nucleotides (tetracodons) and 90+150+90=330 nucleotides (pentacodons). Hence for each of the 16569 positions along the human mitogenome, chimeric peptides are translated from 30 regular, 30 noncanonical and 30 regular codons. These hypothetical peptides (lengths truncated in Figure 1 for presentation purposes) are compared with actual MS mitoproteomic data.

Tetracodons

1 ATGACCCCAATACGCAAAACTAACCCCCTAATA|AAATTAATTAACCACTCATTCATCGACCTCCCCACC|CCATCCAACATCTCC

  M  T  P  M  R  K  T  N  P  L  M |K  *  *  H  H  H  D  S  H |P  S  N  I  S

2 TGACCCCAATACGCAAAACTAACCCCCTAATAA|AATTAATTAACCACTCATTCATCGACCTCCCCACCC|CATCCAACATCTCCG

  W  P  Q  Y  A  K  L  T  P  *  *  |N  N  N  T  I  I  T  P  T  |H  P  T  S  P

Pentacodons

1 ATGACCCCAATACGCAAAACTAACCCCCTAATA|AAATTAATTAACCACTCATTCATCGACCTCCCCACCCCATCCAAC|ATCTCC

  M  T  P  M  R  K  T  N  P  L  M |K  N  T  S  H  T  P  P  P  |I  S

2 TGACCCCAATACGCAAAACTAACCCCCTAATAA|AATTAATTAACCACTCATTCATCGACCTCCCCACCCCATCCAACA|TCTCCG

  W  P  Q  Y  A  K  L  T  P  *  *  |N  I  P  H  I  P  P  P  Q  |S  P

Table 1. Chimeric peptides transcribed in part according to regular, and in part according to expanded codons (tetra- and pentacodons) from the human mitogenome, detected in MS data from Guegneau et al (2014). Columns are: 1. Regular tricodon translation frame (positive strand, 1-3; negative strand, 4-6), tetra- and pentacoded parts indicated by T and P; 2. Position of regular tricoded part on translated human mitogenome; 3. S, amino acid inserted at stop codons; 4. Detected peptide sequence, minor letters indicate translated stops, "|" separates regular tricoded from other part, underlined parts are tetra- and pentacoded. Ambiguous limits between tricoded and other part are also indicated when occurring, ambiguous part is considered parsimoniously as tricoded. 5. Xcorr between expected and observed MS; 6. PSM, counts observed MS matching expected MS; 7. q, false discovery rate; 8. PEP, posterior error probability, peptide specific; 9. Position-specific amino acid modifications; 10. Positions in regular mitogenome-encoded proteins matching regular tricoded part of detected chimeric peptide.

| T | Pos | S | Peptide | Xcorr | PSM | q | PEP | Modifications | Gene |
|---|-----|---|---------|-------|-----|---|-----|---------------|------|
| 4 | 359-373 | A | KGGaYISGA|aaSGENSVNVIKEaa | 3.65 | 193 | 0 | 0.473 | | |
| 5 | 2140-2155 | D | KALENFGKGAAGDGRAHRdVIF|MdPLSCGSQNVMIISS | 3.26 | 15 | 0 | 0.533 | K8(Lys->PyrLys); C28(Carbamidomethyl); M34(Oxidation) | |
| 6 | 1648-1661 | D | KSMQWAILGLFVVG|SGLFNILdEV | 2.54 | 7 | 0 | 0.181 | M3(Oxidation) | |
| 4 | 527-537 | D | RdDMSAWL|ddRMIQPdFTS | 3.59 | 299 | 0 | 1 | | |
| 3 | 1774-1783 | F | RESKNMPISHlfH|ITLLNLYFYL | 2.21 | 31 | 0 | 0.449 | | nd2 285-294 |
| 3 | 4102-4113 | G | KNFGATPNKSNN|QQLgTPNLLPIPHLPPVTY | 2.28 | 1 | 0 | 0.907 | | |
| 5 | 4348-4356 | G | RWCgGWWg|M|GGLGSWESLGS | 3.87 | 221 | 0 | 0.942 | C3(Carbamidomethyl) | |
| 5 | 5417-5427 | G | KRGgGGLVE|IFL|DSCEVLATSLYICL | 3.13 | 1 | 0 | 0.513 | C15(Carbamidomethyl); C25(Carbamidomethyl) | |
| 2 | 378-397 | X | RQNTTSHSLKLKGPGGASY|P|LiAVCIMTRQLPLCQLM | 2.48 | 1 | 0 | 0.219 | K10(Lys->PyrLys); C25(Carbamidomethyl); C34(Carbamidomethyl) | |
| 4 | 936-945 | X | KWSiLEFGEGLCWi|G|CGGNVVSNE | 2.36 | 10 | 0 | 0.744 | C12(Carbamidomethyl); C16(Carbamidomethyl) | |
| 1 | 1762-1780 | X | KTMASSSPPSiPPSPSLT|S|LYiPITHSSTLPI | 3.99 | 1 | 0 | 0.696 | | |
| 6 | 2565-2578 | | KPMITVPAHKGMA|M|LVMMLVLCNS | 2.13 | 27 | 0 | 1 | K1(Lys->PyrLys); C22(Carbamidomethyl) | |
| 4 | 1702-1712 | M | KSTAASTIDPA|mG|SNGLGAmWAE | 4.18 | 286 | 0 | 0.532 | | |
| 4 | 3436-3445 | N | RPPLnQMRAG|EGGGnIKVSFL | 3.84 | 81 | 0 | 0.735 | | |
| 1 | 2883-2890 | N | RLITTQQW|QnMTQKnYLPNnDD | 2.72 | 3 | 0 | 0.226 | K14(Lys->PyrLys) | at6 41-48 |
| 6 | 1138-1148 | P | KCVGQDMpI|W|ISGLFSApGW | 4.41 | 542 | 0 | 1 | C2(Carbamidomethyl); M7(Oxidation) | |
| 1 | 3854-3862 | R | rHNYNKLH|L|LHNNQIVMLP | 3.45 | 17 | 0 | 0.551 | M17(Oxidation) | |
| 1 | 4324-4353 | R | PLLGLLLAAAGKSAQLGLHPWLPSAMEG|PT|QLAYPSAYVr | 2.08 | 2 | 0 | 1 | | nd5 223-241 |
| 4 | 564-575 | T | KGVSVGtVMLDSLG|I|WtItQAPtSEP | 4.14 | 2 | 0 | 1 | M9(Oxidation) | |
| 6 | 1058-1076 | V | KVGGEWSMFDSLYFDI|C|SLLvLWMMDPEHMNSMAL | 2.05 | 1 | 0 | 0.909 | C17(Carbamidomethyl); M33(Oxidation) | |
| 5 | 2583-2595 | W | RYwDAwQVK|MVGWLVwMSEAGV | 2.62 | 4 | 0 | 0.527 | | |
| 3 | 5404-5433 | Y | KPLPATAV|SNQPSTITHQLQLQSHPSPTyMPTNLPTLy | 2.65 | 11 | 0 | 0.493 | | |
| 4 | 2233-2247 | Y | RRAWTKYVDEMNM|VG|GWSyyWGKLSQyW | 2.18 | 2 | 0 | 0.684 | K6(Lys->PyrLys); M11(Oxidation); K23(Lys->PyrLys) | |
| 6 | 4397-4404 | S | RSVSIsNA|MHWSDMSEGWHGSFsKDsLYLSLIYGY | 2.6 | 1 | 0 | 0.288 | M9(Oxidation) | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2 | 378-397 | P | RQNTTSHSLKLKGPGGASY|P|LpAVCIMTRQLPLCQLM | 2.36 | 1 | 0 | 0.319 | K10(Lys->PyrLys); C25(Carbamidomethyl); M27(Oxidation); C34(Carbamidomethyl) |
| 3 | 1909-1916 | | KLVTLAPMTAH|L|LLPPPGK | 2 | 1 | 0 | 1 | K1(Lys->PyrLys); M8(Oxidation); K19(Lys->PyrLys) |
| 5 | 4085-4096 | K | kAPIIYSIKV|TL|FNNSWL | 2.93 | 65 | 0 | 0.34 | K1(Lys->PyrLys) |
| 5 | 3393-3403 | K | kLYCVWM|M|APKMEETPA | 3.51 | 228 | 0 | 0.64 | K1(Lys->PyrLys); C4(Carbamidomethyl); M7(Oxidation); M8(Oxidation) |
| P | | | | | | | | |
| 2 | 1441-1466 | A | KMSAETDSMALT|LISaTMaIEPIPENPKFSVPPITPHP | 2.29 | 1 | 0 | 0.338 | K1(Lys->PyrLys); M18(Oxidation) |
| 4 | 3294-3302 | D | RSSKLQYGd|FPAVMNNSVRKEGWdWSS | 2.92 | 69 | 0.046 | 0.257 | K4(Lys->PyrLys); M14(Oxidation); K20(Lys->PyrLys) |
| 2 | 282-293 | D | KFNdAMLTPGL|V|LWARSRNN | 3.56 | 4 | 0 | 1 | M6(Oxidation) |
| 4 | 4257-4264 | E | KGGEVKGA|FeWISELVFMILLAQRMGSDWLPSGE | 2.72 | 61 | 0 | 0.73 | M25(Oxidation) |
| 4 | 460-473 | G | KFVITVAPQNDIW|P|RGYgSVgLgEgPVSSVDDVMPPCGDg | 2.44 | 2 | 0 | 0.412 | C37(Carbamidomethyl) |
| 2 | 260-280 | G | RKESQTAA|S|KRLAgPHPHGKQQWLTFSNK | 3.53 | 1 | 0 | 0.838 | |
| 4 | 3176-3187 | G | RMYgKDWgLLVAggKSMALMKQPW|G|HSGSGLQRSTC | 2.58 | 48 | 0 | 0.865 | K5(Lys->PyrLys); K15(Lys->PyrLys); C36(Carbamidomethyl) |
| 5 | 1645-1652 | I | RNSGCECViGM|A|DWiVCNE | 2.57 | 65 | 0 | 1 | C5(Carbamidomethyl); C7(Carbamidomethyl); M11(Oxidation); C17(Carbamidomethyl) |
| 2 | 948-959 | M | RAVHAKTSPVKA|MLQYHIAmKSREPLL | 2.02 | 1 | 0 | 0.296 | |
| 6 | 3343-3360 | M | KmmLMMVLPGRK|G|VEVAVCmmYSDASSmDWEmmE | 2.18 | 1 | 0 | 0.774 | C19(Carbamidomethyl) |
| 2 | 282-292 | M | KFNmAMLTPGL|ESSDRSLTI | 4 | 289 | 0 | 0.986 | |
| 2 | 282-292 | N | KFNnAMLTPGL|ESSDRSLTI | 3.46 | 80 | 0 | 0.531 | M6(Oxidation) |
| 4 | 2272-2287 | N | KWWSGPGQNCRIVKVG|TRSTLNLVGGNNNDPV | 2.84 | 1 | 0 | 0.716 | C10(Carbamidomethyl) |
| 5 | 1645-1652 | C | RNSGCECVqGM|A|DWqVCNE | 2.11 | 9 | 0 | 1 | C5(Carbamidomethyl); C7(Carbamidomethyl); C17(Carbamidomethyl) |
| 3 | 3879-3889 | Y | KVNKAMHEyQTHYTYP|TYPSyyQPFSS | 2.33 | 4 | 0 | 0.388 | |
| 4 | 4257-4264 | I | KGGEVKG|A|FiWISELVFMILLAQRMGSDWLPSGE | 2.64 | 65 | 0 | 0.618 | M18(Oxidation); M25(Oxidation) |
| 5 | 1645-1652 | K | RNSGCECVkGM|A|DWkVCNE | 2.11 | 10 | 0 | 1 | C5(Carbamidomethyl); C7(Carbamidomethyl); C17(Carbamidomethyl) |
| 3 | 5166-5176 | K | KQTIQDP|A|TQTIMPkPTP | 2.98 | 128 | 0 | 0.538 | K1(Lys->PyrLys); K15(Lys->PyrLys) |
| 2 | 3153-3162 | K | RKHQPTPCKGSTI|P|IYYLKSFFL | 2.44 | 1 | 0 | 0.307 | C8(Carbamidomethyl); K9(Lys->PyrLys) |

Table 2. Chimeric peptides transcribed in part according to regular, and in part according to expanded codons (tetra- and pentacodons) from the human mitogenome, detected in MS data from Alberio et al (2014). PLGS is the score estimating goodness of fit between observed and expected MS in the PLGS peptide detection software. Δ ppm is the difference between expected and observed MS total mass. Cl indicates cleavage expected by the MS/MS search that detected the specified peptide, C indicates cleavage at the carboxyl-, and N the amino-end of the amino acid. Chym and elas indicate cleavage by chymotrypsin elastase.

| T | Pos | S | Peptide | PLGS | PSM | Δ ppm | Gene | Cl |
|---|-----|---|---------|------|-----|-------|------|-----|
| | | | Tetracoding | | | | | |
| 3 | 4428-4435 | T | NRHQPTTP|TSSLLPPStTNF | 6.79 | 48 | -2.7085 | | Chym |
| 2 | 179-187 | E | HNQPAIYQTTTLe|PYP|EPTKPQe | 6.51 | 41 | -4.7043 | | Hn |
| 2 | 1612-1620 | E | HSSPeYQA|P|SDIRPASS | 6.51 | 37 | -6.2392 | | Hn |
| 1 | 1340-1352 | D | SHANHNLYMTP|TTTIFLGTTYDAL | 6.54 | 41 | 1.9161 | ND1, 234-250 | Sn |
| 5 | 3210-3220 | H | KMNPhAQSTA|A|IFMCSWVGSS | 6.49 | 54 | -3.1235 | | Kn |
| 1 | 2296-2304 | E | TEAMWNDL|L|eLDPGSLLSRGADGFMA | 6.64 | 41 | -0.7789 | | Tn |
| 5 | 4157-4165 | I | CRFiNGGI|VG|SWWQNML | 6.79 | 38 | 2.4082 | | Cn |
| 2 | 3525-3531 | R | YTLSPMSW|r|NNTIAVH | 6.74 | 22 | -2.4967 | | Yn |
| 3 | 1639-1658 | Q | NVSLLLTLSILSIMAGSWGG|QPTSTKQYP | 6.56 | 39 | -1.0766 | ND2, 150-169 | Nn |
| 1 | 1769-1779 | N | AFPNGISnFQKnAHI|P|PSnPPSPSLT | 6.59 | 58 | 1.8674 | | Tc |
| 1 | 5433-5443 | Q | TTGTTTTT|L|TVHSTqSHLP | 6.72 | 31 | 0.5116 | | Pc |
| 2 | 249-264 | P | pSHLNHTS|kEQASSTQQCSSkRLA | 6.83 | 39 | -1.917 | | Pn |
| 2 | 5240-5248 | R | YNPSLT|qTFPqPqTAH | 6.75 | 74 | 1.0238 | CytB, 325-333 | Yn |
| 1 | 2182-2197 | K | YWLLAADL|L|FNWSrHHN | 6.66 | 35 | -0.3615 | | Yn |
| | | | Pentacoding | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 853-861 | S | FPCTKSSQPMsPCs|HV|sRPRYPN | 6.6 | 33 | -1.8371 | | Nc |
| 3 | 5074-5082 | F | TfFNESEEA|TVHPLTSTSSFLFAPQ | 6.94 | 40 | -2.8644 | | Qc |
| 3 | 4612-4622 | D | LSNSALSSNL|S|PdPQLPNQQT | 6.49 | 58 | 0.6481 | | Tc |
| 2 | 399-415 | H | hGACSVIDKPRSTS|P|SLPMALAPMGQ | 6.51 | 58 | 2.9743 | | Hn |
| 1 | 3776-3783 | Y | HHySKFLHSAyYN|L|SQQLNMT | 6.47 | 35 | 0.8352 | | Tc |
| 1 | 1287-1294 | H | PDLAHPGh|W|FISTLAE | 6.9 | 33 | 1.6579 | ND1, 185-192 | Ec |
| 1 | 651-667 | S | LCSKMVGsFMGsGDKP|T|AVSVPMISNS | 7.02 | 63 | -0.6754 | | Ln |
| 5 | 1151-1158 | A | WVaaFLL|Q|MASSGaGGLM | 6.72 | 40 | -1.7249 | | Wn |
| 6 | 457-467 | C | RMVSLcLLWPLcM|ISSGMVcGLF | 6.84 | 23 | -7.7922 | | Chym |
| 2 | 3143-3154 | F | QHTMNWRfRKHQPTP|CPKfPSMRDNPI | 6.57 | 36 | -0.9708 | | Elas |
| 6 | 5476-5483 | S | SLRVMSG|s|QESKTDTA | 6.8 | 71 | -6.8916 | | Ac |
| 4 | 4033-4041 | E | KICAAVECADeeDVAG|e|LVREGYNQ | 6.53 | 29 | 0.04081 | | Qc |
| 1 | 2017-2024 | N | VTTTSTT|L|FSLLDTFSN | 6.73 | 34 | -2.3462 | | Nc |
| 6 | 5475-5483 | M | MLRVMMG|m|QESKTDTA | 6.83 | 26 | 1.7577 | | Ac |
| 2 | 925-933 | Y | KNHGyYLHNHT|Q|VLNYQTCI | 6.69 | 28 | -3.8878 | | Kn |
| 1 | 1122-1136 | H | NAYRTKNShLYTTT|Q|ANWAHAHP | 6.62 | 32 | 0.2141 | | Pc |
| 4 | 2632-2643 | G | DgSLLGGDgSVV|EDLGGKgDSEVAGGSWGMWRSF | 6.63 | 31 | -1.6834 | | Fc |
| 2 | 4073-4081 | Y | SQELTLYyA|Q|ELLTHAPM | 6.83 | 38 | -2.4583 | | Sn |

Chimeric translation according to tri- and expanded codons (tetra- and pentacodons)

Tetracodons

1 ATGACCCCAATACGCAAAACTAACCCCCTAATA|AAATTAATTAACCACTCATTCATCGACCTCCCCACC|CCATCCAACATCTCC

  M  T  P  M  R  K  T  N  P  L  M  |K   *   *   H   H   H   D   S   H  |P  S  N  I  S

2 TGACCCCAATACGCAAAACTAACCCCCTAATAA|AATTAATTAACCACTCATTCATCGACCTCCCCACCC|CATCCAACATCTCCG

  W  P  Q  Y  A  K  L  T  P  *  *  |N   N   N   T   I   I   T   P   T  |H  P  T  S  P

Pentacodons

1 ATGACCCCAATACGCAAAACTAACCCCCTAATA|AAATTAATTAACCACTCATTCATCGACCTCCCCACCCCATCCAAC|ATCTCC

  M  T  P  M  R  K  T  N  P  L  M  |K   N   T   S   H   T   P   P   P  |I  S

2 TGACCCCAATACGCAAAACTAACCCCCTAATAA|AATTAATTAACCACTCATTCATCGACCTCCCCACCCCATCCAACA|TCTCCG

  W  P  Q  Y  A  K  L  T  P  *  *  |N   I   P   H   I   P   P   P   Q  |S  P