BioSystems 147 (2016) 78–93

Contents lists available at ScienceDirect

BioSystems

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

Natural chymotrypsin-like-cleaved human mitochondrial peptides confirm tetra-, pentacodon, non-canonical RNA translations

Hervé Seligmann

Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes, Faculté de Médecine, Université d'Aix-Marseille, URMITE CNRS-IRD 198 UMER 6236, Marseille, France

ARTICLE INFO

Article history:
Received 10 April 2016
Received in revised form 15 July 2016
Accepted 26 July 2016
Available online 28 July 2016

Keywords:
Frameshifting transcription
Translational frameshift
Bijective transformation
Digestion enzymes
Proteome
Systematic deletions
delRNA

ABSTRACT

Mass spectra of human mitochondrial peptides match non-canonical transcripts systematically (a) deleting mono/dinucleotides after trinucleotides (delRNA), (b) exchanging nucleotides (swinger RNA), translated according to tri, (c) tetra- and pentacodons (codons expanded by a 4th (and 5th) silent nucleotide(s)). Swinger transcriptions are 23 bijective transformations, nine symmetric (X <-> Y, e.g. A <-> C) and fourteen asymmetric exchanges (X->Y->Z->X, e.g. A->C->G->A). Here, proteomic analyses assuming cleavage after W,Y, F (chymotrypsin-like, for trypsinized samples) detect fewer chymotrypsinized than trypsinized peptides. Detected non-canonical peptides map preferentially on detected non-canonical RNAs for chymotrypsinized peptides, as previously found for trypsinized peptides. This suggests residual natural chymotrypsin-like digestion detectable within experimentally trypsinized peptide data. Some trypsinized peptides are detected twice, by analyses assuming trypsin, and those assuming chymotrypsin cleavages. They have higher spectra counts than peptides detected only once, meaning that abundant peptides are more frequently detected, but detection certainties resemble those for peptides detected only once. Analyses assuming 'incorrect' digestions are inadequate negative controls for digestion enzymes naturally active in biological samples. Chymotrypsin-analyses confirm non-canonical transcriptions/translations independently of results obtained assuming trypsinization, increase non-canonical peptidome coverage, indicating mitogenome-encoding of yet undetected proteins.

1. Introduction

Recent analyses of human mitochondrial proteomic data detected numerous peptides matching alternative translations of the mitogenome. These peptides correspond to two major non-canonical phenomena. One phenomenon is transcription that systematically exchanges nucleotides, producing 'swinger' transcripts whose homology with the template mitogenomic DNA can't be detected unless assuming the adequate systematic exchange between nucleotides. The other corresponds to translation of codons expanded by a 4th or a 4th and 5th (silent) nucleotides. Fig. 1 describes some of these alternative transcriptions and translations for the 5′ extremity of the human mitochondrial gene's ND2, for expanded codons, systematic deletions, and systematic nucleotide exchanges.

1.1. Swinger polymerizations

Swinger transcripts follow bijective transformations of regular sequences (Michel and Seligmann, 2014). There are 23 transformations, 9 symmetric (X <> Y, e.g. A <> C, Seligmann, 2012a,b,c,d,e, Seligmann, 2013a,b,c), and 14 asymmetric (X > Y > Z > X, e.g. A > C > > G > A, Seligmann, 2013b; Fimmel et al., 2015a,b). Properties of mitochondrial swinger RNAs detected among GenBank's EST database sequenced by classical sequencing methods converge with those detected among mitochondrial transcriptomic data based on next generation massive sequencing RNA seq methods (Seligmann, 2016a). GenBank's EST database includes also chimeric RNAs, matching in part regular, and in part swinger transcriptions of otherwise contiguous mitogenomic regions (Seligmann, 2015a). Swinger DNA, only along exchange rule A <> T + C <> G, has also been detected for mitochondrial and nuclear sequences, mainly for ribosomal RNA genes (Seligmann, 2014a,b). One swinger-transformed insect mitochondrial 16 S rRNA gene inserted within an otherwise regular mitogenome also exists in GenBank (Seligmann, 2015b).

E-mail address: timonuslepidus@gmail.com

http://dx.doi.org/10.1016/j.biosystems.2016.07.010
0303-2647/

Table 1
Peptides detected in human mitochondrial proteome data (from Gueugneau et al., 2014), after translating all six frames of the 23 swinger-transformed versions of the human mitogenome according to the vertebrate mitochondrial genetic code. Analyses assume digestion by chymotrypsin, samples were actually trypsinized. Analyses assume that the same amino acid is inserted at all stops within a single peptide, testing 19 possibilities for peptides translated from sequences with stop(s) (mass spectrometry does not differentiate leucine and isoleucine because of their identical weights). Columns are: 1. Xcorr, goodness of fit between observed mass spectrum and expected mass spectrum for hypothetical peptide; 2. Peptide sequence; 3. Number of matching mass spectra; 4. Adjusted false discovery rate q; 5. Posterior error probability of peptide; 6. Swinger transformation assumed to predict the peptide, and the strand (F or R) and corresponding translation frame; 7. Amino acid inserted at stop(s); 8–9. Positions of 5′ and 3′ extremities on positive strand of sequence coding for the trypsinized peptide including the detected peptide. When two PSM values are shown, the second indicates PSMs detected assuming trypsin-digestion, underlined peptides map on previously detected swinger RNAs (Seligmann, 2016a).

1	2	3	4	5	6	7	8	9
Xcorr	Peptide	PSM	q	PEP	Swing	x	5′	3′
	Chymotrypsin							
4,05	HVLLLHILNDYY	59	0	0,0196	agcR1	V	15147	15084
2,69	NLNVNNLLHYHGCW	157	0	0,032	atgcR1	N	6627	6504
2,8	NYEGGLAW	110	0	0,0542	atR0	H	11922	11775
2,3	GGSSSVDVQGW	19	0	0,0543	gtF2	G	10059	10119
3,33	NDLYAMCSTISGW	7	0	0,0585	agcR1	F	4548	4440
2,56	NLNVTTLLHYHGCW	77	0	0,0845	atgcR1	T	6627	6504
2,42	MNIMNMITSCNITNHTNIF	389	0	0,1	acR1	N	6435	6222
2,66	EVCDGAVGLMAFAGPGGF	102	0	0,122	actR1		10467	10416
2,35	VSVGAGTGGFVSSGF	19	0	0,145	ag-ctF0	T	13605	13716
3,69	TFTPPMMLTSGASSNPQSPLY	289	0	0,19	agcR2	S	3372	3234
3,61	TFTPPMMLTTGASSNPQSPLY	416	0	0,209	agcR2	T	3372	3234
2,48	LWXCGSPLDGDXY	103	0	0,21	cgF2	I	11694	11742
2,4	MLSNHEQSSAATLCTLTHHWVGY	13	0	0,213	ctgF1	H	6837	6981
2,29	MELGGAEW	25	0	0,0573	ctgF0	A	5004	5259
2,5	MAMMMMMSMYGW	263	0	0,223	atgR1	M	12846	12792
3,51	SEESSISAWDTDPMEDGGAGIF	1865	0	0,23	agR0	A	16455	16380
3,82	ETLGAMGESMNNIALPFGIPY	930	0	0,241	atcF2	G	1362	1575
2,78	PTSTQTLLAAAEVQALTMDEW	39	0	0,244	agtcF1	D	11541	11622
2,37	EMLFSLSMISILF	9	0	0,255	agctF0	S	6813	6867
3,62	LLKPVVSSNLHFVLVEKKY	430	0	0,26	ctgF1	K	14250	14304
2,76	MVGALPSVLGSSW	2	0	0,304	agctF2	S	8889	9123
3,39	MELLMEMANSLW	21	0	0,307	atgcF0	I	11274	11409
2,79	NLAIYLKMVEVMF	31	0	0,328	ctgF2		11181	11217
2,11	SMKLIDVKEGKMTMEFVVMAVGSDKNINVMGGMSW	123	0	0,333	actR0	K	13323	13185
3	LLLTITNAYLCNF	29	0	0,333	atcF0	A	11550	11619
2,79	NLAIYLQMVEVMF	30	0	0,343	ctgF2	Q	11130	11484
4,31	MLASWAMTLSSLLMSAALLW	2217	0	0,345	cgtR0	W	6147	6042
2,35	MLLMLTLVLLLGDLLVQLGVLSSVMCLGFF	64	0	0,426	regR0	L	2259	1896
3,56	XTMETXGFDXLXDTXSLVSF	447	0	0,428	ctgF0	X	16308	16422
2,07	NCESHMKSDKEKLMTCWNMKW	7	0	0,432	cgF1	K	12582	12753
2,25	CGSSVGSSCF	12	0	0,451	gtF2	G	10839	10869
2,31	GAWPAPSLCCPCITWLW	77	0	0,519	agF2	I	11523	11574
3,35	TFPTNTAAEXNNAAESW	35	0	0,574	agcF0	L	2793	2862
2,28	ASQPVMPLLPSSTF	645	0	0,648	cgtR1	X	13401	13260
2,23	SYSPPSLSLLTSLF	63	0	0,822	atgcR2	P	5343	5304
2,51	QLVPNCPTCHEEPMVNY	2	0	0,83	agctF2	T	9801	9879
2,44	YHTTDGTNKTHSADY	9	0	0,834	atf2	K	7221	7341
3,27	NLAIYLEMVEVMF	38	0	0,391	ctgF2		11181	11217
3,4	CLEDLVLFLTSCVSPLLSF	654	0	0,837	cgtR0	E	5571	5259
2,8	GCQSSHVSLSLVSLGSSLAHVYLISTPASGPHF	5	0	0,958	agctR0	V	7611	7497
2,19	NGSLAETQFHY	518	0	1	actF2	F	6615	6705
2,19	NGSLAETQMHY	373	0	1	actM2	M	6615	6705
2,23	LLGCLTSPGMQPNLGSPSQVGLMQTITMHNLILHHGGLF	160	0	1	ctgR2	G	13644	13278
2,34	QAQGEDGQAHGGLEAGAAF	2	0	1	agR2	E	6357	6300
2,48	MMELEEDELNMNEDENW	175	0	1	ctR0	E	5379	5328
3,22	SVQLPIXLLDXGEYXIXIMF	284	0	1	ctF1	X	3594	3693
	Trypsin							
2,72	SSLLSNPLK	11	0	0,191	acF0	N	10374	10443
2,21	IMLTNLEMEVLLIQDPLHDGTIMLTILGR	27	0	0,338	acF0	L	9123	9207
2,01	LAWSILR	12	0	0,333	acgF0	W	9129	9171
2,08	DLSQEAAPQHMQK	3	0	0,358	acgR0	S	12504	12468
3,36	SEESSSQVSSPSGK	48, 55	0	0,357	ac-gtF0	E	13986	14025
2,6	EAQGDSSGHDTSHPGSAGAAENELDTR	23	0	0,334	acgtF2	H	1155	1260
3,32	LPMVMCPNLSMGTQELPK	3	0	0,132	ac-gtF2	C	9132	9183
2,21	ISAVSLVR	1236, 2373	0	0,502	acgtR1	X	5589	5568
2,42	NSAVSLVR	429, 1220	0	1	acgtR1	N	5589	5568
2,1	AGAPTLMLSAVQGSAVPFVTGR	36	0	0,183	acgtR2	G	7716	7611
2,77	EQGPMAGDGGEPGR	111	0	0,173	acR0	Q	2256	2208
2,77	EKGPMAGDGGEPGR	105	0	0,187	acR0		2247	2208
2,9	MVSSSSDFLILGCNDSGGR	9	0	0,776	acR1	S	2967	2913
3,24	MAGGNIAASAMIASASWSVCIK	268	0	0,83	acR1	A	4554	4446
2,07	QCPGMEVLGVIR	7	0	0,022	acR2	V	7443	7410
2,53	SINNMMMMPTPAHPSNMVK	88	0	1	acR2	M	12735	12681
3,06	MELLQMR	80, 160	0	0,0815	actgR0	E	11655	11637
2,05	LWTSVAR	3	0	0,0841	actR1	A	1269	1251
3,84	DIIGGMDVFHGAK	347, 813	0	0,184	actR1	F	13623	13587