Complete set of mitochondrial pan-edited mRNAs in *Leishmania mexicana amazonensis* LV78

Dmitri A. Maslov*

Department of Biology, University of California, Riverside, 900 University Avenue, Riverside, CA 92512, USA

ARTICLE INFO

Article history:
Received 14 December 2009
Received in revised form 13 May 2010
Accepted 14 May 2010
Available online xxx

Keywords:
*Leishmania amazonensis mexicana* RNA editing kinetoplast Trypanosomatidae

ABSTRACT

Editing of mRNA transcribed from the mitochondrial cryptogenes ND8 (G1), ND9 (G2), G3, G4, ND3 (G5), RPS12 (G6) was investigated in *Leishmania mexicana amazonensis*, strain LV78, by amplification of the cDNA, cloning and sequencing. For each of these genes, extensively and partially edited transcripts were found to be relatively abundant compared to the respective pre-edited molecules. Moreover, the editing patterns observed in a majority of transcripts of each gene were consistent among themselves which allowed for inferring consensus editing sequences. The open reading frames contained in the consensus sequences were predicted to encode polypeptides that were highly similar to their counterparts in other species of Trypanosomatidae. Several kinetoplast DNA minicircles from this species available in the public domain were found to contain genes for guide RNAs which mediate editing of some of the mRNAs. The results indicate that the investigated strain of *L. m. amazonensis* has preserved its full editing capacity in spite of the long-term maintenance in culture. This property differs drastically from the other *Leishmania* species which lost some or all of the G1–G5 mRNA editing ability in culture.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Mitochondrial genes in trypanosomatid protists are encoded by maxicircles which together with more abundant minicircles are catenated in the form of a kinetoplast DNA (kDNA) network [1–3]. The genes localized in maxicircles include two ribosomal RNA genes (12S and 95 rRNA) and eighteen protein-coding genes that represent ribosomal protein S12 (RPS12) and subunits of mitochondrial respiratory complexes I, III–V [4]. There are no tRNA genes, and all tRNAs participating in mitochondrial translation are imported from the cytoplasm [5]. The most striking feature of the kinetoplast–mitochondrial genetic system is the process of RNA editing whereby messenger RNAs of several protein-coding genes are modified by specific insertions and (less frequently) deletions of uridylic residues [6]. Positions of the insertion/deletion sites and numbers of the U’s involved are defined by sequences of guide RNAs (gRNA) that are encoded mainly in the kDNA minicircles [7–9]. Although each minicircle class encodes only 1–3 different gRNAs, a single kDNA network contains dozens or hundreds of minicircle classes, thereby providing the required multitude of gRNAs [10]. Editing information is passed on from the guide RNAs to the messengers by interaction of the two RNA molecules mediated by complex enzymatic machinery (reviewed in [11–15]). Besides the 20S core “editosome” or “L-complex” [16–18], several additional macromolecular complexes have been identified including the complexes involved in gRNA biogenesis [19,20] and mRNA polyadenylation [21].

Editing results in conversion of a pre-edited transcript with a defective reading frame into a mature mRNA [22–25]. Translation of the edited apocytochrome b (Cyb) mRNA was demonstrated [26]. The magnitude of RNA editing varies from a few inserted U-residues, such as those needed to repair a −1 frameshift in the COI mRNA [6,27], to hundreds of residues, as exemplified by ND7 or A6 transcripts that contain no recognizable ORF prior to editing [22,23,25]. Such extensively edited (pan-edited) mRNAs are processed with involvement of multiple gRNAs each of which being responsible only for a short ‘editing block’ (30–50 nt) of mRNA sequence [28–30]. Editing begins at the 3′-end of a pre-edited transcript by formation of a 10–15 nt ‘anchor’ duplex between the first gRNA and the mRNA and then expands towards the 5′-end until the entire gRNA is base-paired with the mRNA [7]. This is followed by annealing of the second gRNA which displaces the first gRNA from the overlap region and then extends the edited sequence further upstream, and so on until the entire mRNA is edited [31].
D.A. Maslov / Molecular & Biochemical Parasitology xxx (2010) xxx–xxx

ARTICLE IN PRESS

G Model
MOLBIO-10442; No. of Pages 8

Please cite this article in press as: Maslov DA. Complete set of mitochondrial pan-edited mRNAs in Leishmania amazonensis LV78. Mol Biochem Parasitol (2010), doi:10.1016/j.molbiopara.2010.05.013

Such a complex process can be vulnerable to deleterious mutations and random genetic drift [32]. Indeed, only the indispensable mRNAs types, such as mRNA encoding subunits of cytochrome c oxidase, cytochrome bc1, and F1F0 ATPase or RPS12, are invariably edited in the investigated trypanosomatid species with functional oxidative phosphorylation suggesting the presence of a strong selection pressure to maintain intactness of the editing system [29,33]. This is not the case with respect to the five short cryptogenes, earlier referred to as G-rich regions G1–G5 (also called CR1–CR5) [29,33]. These are thought to encode subunits of Respiratory Complex I (NADH dehydrogenase), although sequence-based identification was only possible for G1 (ND8) [34], G2 (ND9) [35] and, with lower confidence, G5 (ND3) [36]. In line with the uncertain function, and even the very existence, of Complex I in cultured Leishmania cells [37], it was found that productive editing of the G1–G5 mRNA is disrupted in the old laboratory strain of Leishmania tarentolae [29,38]. It was hypothesized that in the absence of selection to maintain the full repertoire of guide RNAs, including those necessary for editing of G1–G5 mRNAs, the minicircles that encode these gRNAs had been lost in culture [32]. This view was supported by finding that in a more recently isolated strain of the same species the editing was almost intact [28].

Consequently, it can be thought that periodic passages of cultures through an animal or a sandfly should prevent the loss of editing at the population level. It was surprising, therefore, to find that a virulent strain of Leishmania donovani, with a presumed ability to propagate in animal and insect hosts [39], also showed disruption of the G1–G5 mRNA editing, as well as insertional inactivating of one additional subunit of Complex I [33]. However, the other investigated mRNAs were efficiently edited in the same strain, indicating that the editing machinery in general was intact. These findings suggest that the selective loss of editing in this strain might occur relatively rapidly (since the most recent passage through the host or vector). Alternatively, the G1–G5 subunits or possibly even the entire Complex I, might be dispensable not only in culture but also in hosts, at least in the laboratory settings that might not fully recapitulate crucial steps of the natural life cycle. In this case the observed anomalies might have occurred over a longer period of time (since the initial isolation in culture in 1962).

Another question to be addressed is to what extent the observed dispensability of the G1–G5 products represents a general feature which, therefore, can also be observed in other Leishmania species, especially those that also were maintained in the laboratory for long time. In this work the editing has been investigated in Leishmania mexicana amazonensis initially isolated in 1972. The strain (LV78) is virulent and has been used intensively for investigation of various aspects of Leishmania cell biology and pathogenesis (see e.g. Refs. [40–43]). The results presented below indicate that in this organism, unlike in L. tarentolae and L. donovani, the editing of G1–G5 transcripts is intact. Such maintenance of editing indicates that there are significant, although not yet fully appreciated, differences among Leishmania species which allow for the loss of editing in some cases but prevent this in others.

2. Materials and methods

2.1. Culture conditions, isolation of kDNA and recombinant DNA procedures

A culture of L. m. amazonensis (MPRP/BR/72/M1845), strain LV78, virulent clonal line 12-1, was provided by K.-P. Chang. Promastigotes were maintained at 26 °C in the M199 medium (Invitrogen) supplemented with 10% (v/v) heat inactivated fetal bovine serum and other components as described previously [39]. Other aspects of culture maintenance and growth of large (1–2 L) batch cultures were performed essentially as described for L. tarentolae [44]. Kinetoplast DNA was isolated by the CsCl cushion sedimentation procedure [45]. PCR amplification, cloning and sequencing were performed using standard procedures as described elsewhere [33]. DNA sequences were analyzed using Vector NTI.

2.2. Reverse-transcription PCR and related procedures

Trizol (Invitrogen) extraction procedure was used to isolate total cell RNA of promastigotes. RT-PCR was performed using SuperScript™ III One-Step RT-PCR system (Invitrogen). The following oligonucleotides were used for RT-PCR of mitochondrial mRNA:

- ND8 (G1), M268 (TAAATATAATAGAATATTTAT) and M269 (AACATTACTAATTTTTACTGCA);
- ND9 (G2), M270 (TATAATAATCATTTATA) and M271 (AAATAATTCTAGTTTAAAGT);
- G3, M272 (AAAAATAATATATGTATTAAC) and M273 (CTTTTTACTCCAAATAATAGCTA);
- G4, M274 (TAGATTATAAAATTATTTAAG) and M275 (TAAATAAAATTTATCACTC);
- ND3 (G5), M276 (TACTCTAAAAATACCTGAGG) and M277 (TATATTATTGCTCTAGAAAGCTTT);
- RPS12 (G6): M278 (CATAAAACTCTATAGCTT) and M279 (TATTTTTATAAAAATATATTTAAAAC).

cDNA was purified, cloned and sequenced as described previously [33]. Guide RNA search was performed using the UWGCG program BESTFIT [46].

3. Results and discussion

3.1. Partial sequence analysis of L. m. amazonensis LV78 maxicircle kDNA

The maxicircle DNA region encompassing two adjacent G-rich cryptogenes ND8 and ND9 and also the region containing the G3 cryptogene were amplified by PCR using flanking oligonucleotide primers chosen with the aid of the available L. tarentolae and L. donovani maxicircle sequences. Subsequently, the regions were expanded and merged into a single 9373 bp contig by amplification of additional segments and by primer walking. The sequenced region represented more than a half of the maxicircle conserved region and included nearly the entire 12S rRNA gene (except for 60–65 nt missing at the 5′-end), followed by the genes for 9S rRNA, ND8, ND9, MURF5, ND7, COII, Cyb, A6, MURF1 (putative ND2) [47], G3 (a cryptogene), ND1, and flanked on the other side with the 5′ half of the COII gene. The genes were found in the same order and with the same polarity as in other investigated trypanosomatid species. The region including the G4 cryptogene (574 bp) and another region including the adjacent cryptogenes ND3 and RPS12 (876 bp) were amplified using conserved sequences in the flanks.

The determined gene sequences from L. m. amazonensis show a high degree of similarity (85–90%) to homologous genes from other Leishmania species. The maxicircle genome conservation includes positions and lengths of the pre-edited regions which therefore can be easily identified by sequence comparisons. Thus, in L. m. amazonensis, as in other investigated Leishmania species, the ND7 gene contains two short pre-edited regions: the first one at the 5′-end of the coding sequence and the second one ~170 nt downstream; the COII and Cyb genes are 5′-edited; the A6 gene contains an extensively edited 5′-region; and transcripts of the ND1, MURF1 and MURF5 genes do not require editing (data not shown).

The conservation even includes some idiosyncratic features of the MURF5 and MURF1 ORFs as illustrated in Fig. 1. The in sil-
ico translated MURF5 amino acid sequences from several species (Fig. 1A) show a high degree of similarity except for the very N- and C-termini. The homology in that region is also shared by the Phytomonas serpens and Trypanosoma brucei sequences. The conserved part, therefore, represents the functional polypeptide product. However, positions of the start and the end codons are less certain. Unlike in the other two species, none of the predicted Leishmania sequences begins with methionine. A conserved leucine residue, encoded by a TTG codon and found at alignment position 15, probably takes its place. This suggests that instead of ATG, the MURF5 initiation involves a noncanonical codon TTG in Leishmania. Earlier, TTG was proposed to serve as initiation codon in pan-edited G4 and ND3 mRNA of T. brucei [36,48]. It also frequently functions as initiation codon in mitochondria of nematodes [49].

The position of translation termination codon is also problematic in the MURF5 sequences. Conservation of the amino acid sequences ends at positions 93–97 which coincide with the end of the L. m. amazonensis ORF, but not in the other species (Fig. 1A). A possibility exists that in these cases the termination codon is created by polyadenylation of an in-frame TA or T that are localized in this region but this has not been verified experimentally.

Most of the predicted MURF1 N-terminal sequences, including L. m. amazonensis (Fig. 1B), also lack methionine. The exception is the P. serpens sequence where it is found at the start of the conserved region (alignment position 39). The corresponding position in the other sequences is occupied by leucine which is encoded by TTC. This again points at the potential role of this triplet as a noncanonical initiation codon in trypanosomatids.

The encoded initiation and termination codons in other L. m. amazonensis genes are either ATG (Met), as in ND1 and COII, or ATA (Ile) in ND7, although in the last case there is also a conserved in-frame ATG six codons downstream (data not shown). The initiation codons of COIII, Cyb and A6 are likely to be localized within the respective 5′ pre-edited regions and created by editing out of the encoded AG dinucleotides, as in the other Leishmania species. The encoded termination codons were canonical TAA (ND7, Cyb, A6, MURF1, ND1) or TAG (COII). The TGA codon, as in other trypanosomatids, is predicted to code for tryptophan.

3.2. Editing of pan-edited cryptogene mRNAs

Reverse-transcription PCR was employed to investigate editing of transcripts derived from the six pan-edited cryptogenes of L. m. amazonensis. The amplification strategy was based on the presence of unedited sequences at the 5′- and 3′-ends of the transcripts that...
Fig. 3. Pan-edited mitochondrial ND8 (NADH dehydrogenase subunit 8) mRNA sequence of *L. m. amazonensis*. The sequence (*Lam*, upper lane) and its *in silico* translation (the lane underneath the mRNA sequence) are compared with the respective protein and mRNA sequences from *L. tarentolae* LEM125 (*Lt*). Edited regions on both mRNA are underlined. Unedited 5'- and 3'-ends expected to be present in *Lam* mRNA and used as PCR primers are shown with bold italics. Encoded nucleotides preserved in the edited transcripts are shown with upper case letters; inserted uridylates are shown with lower case u's and deleted uridylates are shown with asterisks. Dashes represent alignment gaps. The exact 5'- and 3'-ends of the *Lam* mRNAs are unknown, while the actual ends of the *Lt* mRNA are as shown.

Fig. 4. Pan-edited mitochondrial ND9 (NADH dehydrogenase subunit 9) mRNA sequence of *L. m. amazonensis*. Most designations are as in Fig. 3. The 3'-end alternative editing pattern caused by the presence of an extra U (highlighted with black) in some *Lam* transcripts (see the text) is also shown. A minicircle-encoded guide RNA, gND9-V, is shown matched with the *Lam* edited sequence with canonical base pairs indicated with vertical lines and noncanonical G–U base pairs indicated with colons.

Please cite this article in press as: Maslov DA. Complete set of mitochondrial pan-edited mRNAs in *Leishmania mexicana amazonensis* LV78. Mol Biochem Parasitol (2010), doi:10.1016/j.molbiopara.2010.05.013
flank the pan-edited mRNA regions [46]. The flanking sequences are used to design PCR primers which, therefore, allow for amplification of all types of mRNA (pre-edited, partially edited and fully edited) transcribed from a targeted cryptogene. The flanking sequences in the L. m. amazonensis were identified by comparison with the sequences from L. tarentolae [28]. Due to the noticeable size increase (nearly twofold) after extensive editing, cDNA products representing pan-edited transcripts can be distinguished from pre-edited products. Indeed, after gel electrophoresis of the amplified material for each cryptogene, a discrete cDNA band with the size larger than the pre-edited product was observed (Fig. 2, arrow).

Moreover, judging by relative abundance of the edited cDNA products (arrows) compared to the pre-edited cDNA products (dots), it can be tentatively concluded that the edited products represent a significant fraction in the steady-state mRNA pool. These results were the first indication that productive editing takes place in each case.

To verify this, the amplified material was cloned in the plasmid vector and sequenced. The presence of inserted or deleted uridylates in the cDNA sequences was deduced from comparison with the corresponding gene sequences. Edited sequences of individual clones were arranged in the order reflecting the 3′–5′ progression of editing and used to derive the editing consensus sequences [46].

The ND8 (=G1) edited sequence was derived from the sequences of 41 cDNA clones that represented various stages of editing. Translation of the consensus yielded a 146 aa polypeptide that was 91.1% identical to the ND8 from L. tarentolae LEM125 (Fig. 3), the only other Leishmania species from which this sequence was available [28], indicating that the functional, fully edited NDB mRNA is also
present in L. m. amazonensis. The canonical initiation and termination codons are created by editing in both species.

An ND9 (=G2) edited sequence (Fig. 4) was derived from the analysis of 16 cDNA clones including six clones that were consistently edited throughout the entire length. The 203 amino acid sequence derived from the editing consensus sequence is highly (84.7%) similar to the 190 aa L. tarentolae sequence except for the very C-terminal part. It needs to be mentioned that there was a minor group of partially edited cDNA clones that deviated from the editing consensus in this region by the presence of a single additional U-residue (highlighted with black in Fig. 4). Interestingly, the amino acid sequence derived from this alternative editing pattern is nearly identical to the amino acid sequence derived from this alternative editing pattern is only 18.1% (Fig. 6B). Remarkably, the entire L. m. amazonensis sequence is either (yet more conserved) pattern were present in the fully edited L. m. amazonensis mRNA, then the proteins from the two species would be 92.7% identical. It is uncertain if this pattern is present in at least some of the mature mRNA molecules, so that two types of proteins (yet more conserved) pattern were present in the fully edited L. m. amazonensis mRNA, then the proteins from the two species would be 92.7% identical. It is uncertain if this pattern is present in at least some of the mature mRNA molecules, so that two types of proteins. Both encoded polypeptide is 166 amino acids long and includes six transmembrane segments but shows no identifiable homology besides the counterparts in other trypanosomatids. The previously determined sequence from L. tarentolae LEM125 is highly similar (67.3% identity) with the L. m. amazonensis sequence in the C-terminal half, however the level of identity in the N-terminal half is only 18.1% (Fig. 6B). Remarkably, the entire L. m. amazonensis sequence is rather similar to the sequence from Trypanosoma brucei (55.8% identity in the overlapping region), which lends additional credence to the L. m. amazonensis consensus. There are several alternatives for the position of the translation initiation codon in the L. m. amazonensis sequence. One possibility represents a canonical AuG that is conserved in T. brucei (Fig. 6A). However, this would leave an unusually long untranslated edited sequence upstream and that would represent a situation not typically observed in trypanosomatids. The alternatives represent several in-frame AuU (Ile) codons in the upstream edited sequence. The predicted termination codon is canonical uAA found in the same position as in T. brucei.

The inferred G4 edited sequence is shown in Fig. 6A. The edited consensus was based on the analysis of 30 cDNA clones including two clones that were edited over the entire length. The encoded polypeptide is 166 amino acids long and includes six transmembrane segments but shows no identifiable homology besides the counterparts in other trypanosomatids. The previously determined sequence from L. tarentolae LEM125 is highly similar (67.3% identity) with the L. m. amazonensis sequence in the C-terminal half, however the level of identity in the N-terminal half is only 18.1% (Fig. 6B). Remarkably, the entire L. m. amazonensis sequence is rather similar to the sequence from Trypanosoma brucei (55.8% identity in the overlapping region), which lends additional credence to the L. m. amazonensis consensus. There are several alternatives for the position of the translation initiation codon in the L. m. amazonensis sequence. One possibility represents a canonical AuG that is conserved in T. brucei (Fig. 6A). However, this would leave an unusually long untranslated edited sequence upstream and that would represent a situation not typically observed in trypanosomatids. The alternatives represent several in-frame AuU (Ile) codons in the upstream edited sequence. The predicted termination codon is canonical uAA found in the same position as in T. brucei.

The G5 (putative ND3) edited sequence of L. m. amazonensis was derived from 15 cDNA clones including seven clones consistently edited almost over the entire length. The canonical initiation and termination codons flank the ORF found in the sequence (Fig. 7A). The 125 aa polypeptide has three transmembrane segments but shows no identifiable homology. The alternatives represent several in-frame AuU (Ile) codons in the upstream edited sequence. The predicted termination codon is canonical uAA found in the same position as in T. brucei.
Fig. 7A) in the L. tarentolae sequence shifts the translation frame with respect to the L. m. amazonensis sequence, and that leads to reduction of the similarity between the C-terminal parts of the two proteins. The possibility exists that in either species this difference is eliminated by a subsequent re-editing within this region or that there is a minor fraction of alternatively edited transcripts (as discussed above for ND9 mRNA) but this has not been investigated. The C-terminal, as well as the N-terminal part of the L. m. amazonensis protein shows a noticeable conservation with respect to the T. brucei protein (Fig. 7B). The N-terminal segment of the L. tarentolae protein is not available, apparently due to incomplete editing of the G5 mRNA in that strain [28].

The RPS12 (=G6) edited sequence (Fig. 8) was based on the analysis of 25 cDNA clones including 19 clones that were consistently edited almost over the entire length. The edited consensus encoded a polypeptide that was 76.1% identical to the L. tarentolae sequence and 87.5% identical to the L. donovani sequence. In the Leishmania species the reading frame starts with a conserved noncanonical Aua (Ile) codon that is created by editing and ends with an encoded conserved UAA. The conservation also includes the presence of two Aua (Ile) codon that is created by editing and ends with an encoded

The gene potentially encoding a gRNA for the last editing block of the G3 mRNA is found in the minicircles M21235 and M21236 (Fig. 6A). Two putative guide RNAs were identified for the G4 mRNA (Fig. 6A). The first editing block (gG4-I) is encoded by the minicircles M94088, M94089, Z11549, Z11550, Z11554, Z11556. Another guide RNA is encoded by the minicircles M94091, Z11553, Z11555. According to its position with respect to the edited sequence it represents block VI gRNA (gG4-VI). A potential gene encoding a guide RNA for editing block IV of ND3 mRNA (gND3-IV) was found in the minicircle entry M21327 (Fig. 7A).

A gene for potential block II Cyb gRNA is contained within the sequenced region of the maxicircle in the same locus as in the other species (positions 1791–1740 of the GenBankTM HM439238 entry, the complementary strand).

4. Conclusion

I have found that mRNA derived from each of the six mitochondrial pan-edited cryptogenes is productively edited in promastigotes of L. m. amazonensis, strain LV78. So far, this organism was the only investigated Leishmania that possessed the full editing capacity; in the other investigated species of this genus the editing of G1–G5 mRNA was fully or partially disrupted [28,29,33]. It is also noteworthy that the productive editing is observed in L. m. amazonensis in spite of the fact that this is another example of an old laboratory strain which had passed through numerous replication cycles in culture under a supposedly relaxed selection to maintain the complete minicircle and the gRNA repertoire. This lack of selection pressure was believed to be the main factor for the loss of editing in the other cases. This new finding suggests that some important, yet not fully appreciated, biochemical differences among the strains or species play an important role in maintaining intactness of the editing system or allowing for its loss in some cases. The availability of a representative set of genomic sequences from this genus will shed light in this and other questions addressing evolution and comparative biology of this group.

Note: Nucleotide sequence data reported in this work are available in the GenBankTM database under the accession numbers HM439238–HM443080.

Acknowledgments

I am indebted to Prof. K.-P. Chang for providing the strain of L. m. amazonensis and for valuable advice and discussions. Technical assistance by C. Kim and A. Banerjee is greatly appreciated. This work was supported by the NIH grant AI070927 to D.A.M.


Maslov DA, Simpson L. The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell 1992;70:459–67.


