A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*

Dennis V. Lavrov, Wesley M. Brown, and Jeffrey L. Boore

Department of Biology, University of Michigan, 830 North University Avenue, Ann Arbor, MI 48109-1048

Edited by Paul R. Schimmel, The Scripps Research Institute, La Jolla, CA, and approved October 11, 2000 (received for review August 21, 2000)

We determined the complete mtDNA sequence of the centipede *Lithobius forficatus* and found that only one of the 22 inferred tRNA genes encodes a fully paired aminoacyl acceptor stem. The other 21 genes encode tRNAs with up to five mismatches in these stems, and some of these overlap extensively with the downstream genes. Because a well-paired acceptor stem is required for proper tRNA functioning, RNA editing in the products of these genes was suspected. We investigated this hypothesis by studying cDNA sequences from eight tRNAs and found the editing of up to 5 nt at their 3′ ends. This editing appears to occur by a novel mechanism with the 5′ end of the acceptor stem being used as a template for the de novo synthesis of the 3′ end, presumably by an RNA-dependent RNA polymerase. In addition, unusual secondary structures for several tRNAs were found, including those lacking a TΨC (Ψ) or a dihydrouridine (D) arm, and having an unusual number of base pairs in the anticodon or anticodon stems.

RNA editing has been defined as “any programmed alteration of RNA primary structure to generate a sequence that could have been directly encoded at the DNA (gene) level” (1). First discovered less than 15 years ago, it is now regarded as a widespread phenomenon occurring in all types of eukaryotic RNA, mainly in mitochondria, less frequently in chloroplasts, and, in a few cases, in the nucleus (2). Four different types of tRNA editing have been found to date (all of them in mitochondria): cytidine to uridine conversion, cytidine or uridine insertion, template-dependent editing of the first three nucleotides at the 5′ ends of tRNAs, and template-independent editing at the 3′ ends of tRNAs (for review, see ref. 3). With a single exception (4), only the last type has been reported so far in animal tRNAs. None of the previously described types of tRNA editing, however, could serve as a potential editing mechanism for the aberrant tRNA structures encoded by the mtDNA of the centipede *Lithobius forficatus*.

We have sequenced the complete mtDNA of *L. forficatus* (GenBank accession no. AF309492) and found that all but one of its 22 tRNA genes code for tRNAs with at least one and as many as five mismatches in their aminoacyl acceptor stems (henceforth termed acceptor stems). Some, but not all, of these tRNA genes, if they encode full-length acceptor stems, would also overlap with their downstream genes. Because a well-matched acceptor stem is important for defining tRNA structure (5), directing tRNA processing (6), and assisting tRNA recognition by aminoacyl-tRNA synthetase (7), we suspected that the poorly matched acceptor stems of *L. forficatus* tRNAs are posttranscriptionally edited. We studied this hypothesis by analyzing cDNA sequences from eight tRNAs and found that a previously unknown type of tRNA editing is present in the mitochondria of *L. forficatus*.

**Materials and Methods**

**tRNA and DNA Preparation.** Total RNA was prepared by using TRizol reagent (GIBCO/BRL) from a frozen and homogenized *L. forficatus* centipede collected in Ann Arbor, MI. To eliminate most of the rRNA, we modified previously published protocols (8, 9) as follows. Total RNA was dissolved in 200 μl of 2 M LiCl/0.1 M potassium acetate, pH 5.0. The mixture was chilled on ice for 5 min and centrifuged for 20 min at 5,000 × g. The supernatant, containing the salt-soluble RNA fraction (mostly tRNA) was transferred to a new tube and RNA was precipitated by the addition of 3 vol of 100% ethanol and centrifugation at 12,000 × g for 15 min. The RNA pellet was washed with 80% ethanol and dissolved in 30 μl of water. The DNA was prepared from the same centipede by using the 2× cetyltrimethylammonium bromide extraction buffer, phenol/chloroform extraction and ethanol precipitation.

**tRNA Circulation and cDNA Synthesis.** RNA was ligated by using T4 RNA ligase (GIBCO/BRL) and cDNA was synthesized by using avian myeloblastosis virus reverse transcriptase (Promega) under conditions as described (1). Primers used for reverse transcription were Litho-trnC-R (5′-ataaggatccAAACTAG-GATTCAATATACTCATGAC-3′) for tRNA(C), Litho-trnM-E (5′-ataaggatccGTGTGGTGATATTTT-3′) for tRNA(E), Litho-trnH-R (5′-ataaggatccCAAGCCATATTCTCT-3′) for tRNA(H), Litho-trnM-R (5′-cacaagcttATGAACCCAG-CCTTTTTTTGTG-3′) for tRNA(M), Litho-trnL-R (5′-ataaggatccCAGTGTAATGTCCTTCTGAC-3′) or Litho-trnR-R (5′-ataaggatccCAGTGTAATGTCCTTCTGAC-3′) for tRNA(L), Litho-trnQ-R (5′-ataaggatccAAACTAG-CTTCTATTATGCTA-3′) for tRNA(Q), Litho-trnS1-R (5′-ataaggatccGAAACTGATTGCAATATATCGC-3′) for tRNA(S1), Litho-trnS2-R (5′-ataaggatccTAGCTTAATTTAGC-3′) for tRNA(S2), and Litho-trnX-R (5′-ataaggatccAAATTATTATGCTA-3′) for tRNA(X). The 5′-ataaggatccAAATTATTAATGCTAACATC-3′ for tRNA(O), Litho-trnR-R (5′-ataaggatccGAAACTGATTGCAATATATCGC-3′) for tRNA(R), and Litho-trnS1-R (5′-ataaggatccGAAACTGATTGCAATATATCGC-3′) for tRNA(S). Throughout, tRNAs are designated by the one-letter amino acid code as tRNA(X), with the two leucine (L) and two serine (S) tRNAs differentiated by anticodon sequences as L1 (anticodon = UAG), L2 (UAA), S1 (UCU) and S2 (UGA); the tRNA genes are named *trnX*. The 5′ end 9 nt in these primers (in lowercase type) were added to create a restriction site for *HindIII* or *BamHI* plus three terminal nucleotides.

**PCR Amplification, Cloning, and Sequencing.** PCR amplification of cDNA was performed by using a Fisher Biotech Taq DNA polymerase kit and the following primer pairs: Litho-trnC-R and Litho-trnM-F (5′-cacaagcttTTGATGCAAGTCTTACTCAGC-3′) for tRNA(C), Litho-trnE-R and Litho-trnF-E (5′-cacaagcttCA-...
Results and Discussion

Genomic Sequences of L. forficatus tRNAs. The mitochondrial genome of the centipede L. forficatus is typical for arthropods in terms of its size, nucleotide composition, gene content, and gene arrangement (Fig. 1 and unpublished data). However, all but one of the 22 inferred tRNA genes encode tRNAs with at least one mismatch in their acceptor stems (Fig. 1). The 5’ end of the acceptor stems inferred from mtDNA sequence. The 5’ end 9 nt in all of these primers were added to create a restriction site for HindIII or BamHI plus three terminal nucleotides. An aliquot of each PCR product was electrophoresed on a 1% agarose gel and visualized; the remainder was extracted with phenol/chloroform and digested with restriction enzymes HindIII and BamHI (Promega). One microliter (∼0.1 μg) of Bluescript (Stratagene) plasmid digested with the same restriction enzymes and 2 μl (∼0.2 μg) of PCR product were ligated in a 10-μl reaction containing 1.5 units of T4 ligase (Promega) for 2 h at room temperature. Stratagene Gold competent cells were transformed, single colonies were harvested, and plasmid DNA was prepared by using the alkaline lysis method. The sequence of each cloned insert was determined on an Applied Biosystems model 377 automated DNA sequencer using a BigDye DNA sequencing kit (Applied Biosystems) and modified T3 primer (5’-GAA-CAAAAGCTGGAGCT-3’).

Results and Discussion

Genomic Sequences of L. forficatus tRNAs. The mitochondrial genome of the centipede L. forficatus is typical for arthropods in terms of its size, nucleotide composition, gene content, and gene arrangement (Fig. 1 and unpublished data). However, all but one of the 22 inferred tRNA genes encode tRNAs with at least one and as many as five mismatches in their acceptor stems (Fig. 1). We analyzed cDNA sequences from eight L. forficatus mitochondrial tRNAs, those for C, E, H, M, N, Q, R, and S1. The genes trnR, trnN, trnS1, and trnE form a contiguous cluster in the mtDNA with trnR and trnN separated by one intervening nucleotide, trnN and trnS1 overlapping by 4 nt, and trnS1 and trnE overlapping by 2 nt (Fig. 2A). trnC, trnH, trnM, and trnQ are located in different parts of the mtDNA and are not adjacent to other tRNA genes on the same strand. trnQ is the only tRNA gene in L. forficatus mtDNA that does not encode any mismatches in the acceptor stem; other tRNA genes studied encode between one and four mismatches, as enumerated in Fig. 1.

Analysis of cDNA. For all tRNAs analyzed, at least some cDNAs have sequence at the 3’ end of the acceptor stems that is different from that encoded by the mtDNA and perfectly matches the 5’ end of the stems (Fig. 2B). In all of these cases, the discriminator nucleotide (always A) and, in all but one, the trinucleotide CCA are also present at the 3’ end of the tRNA, indicating that these cDNA sequences represent mature tRNAs (Table 1). Because the rest of the cDNA sequence matches that of the mtDNA perfectly and because no import of tRNAs has been reported for arthropod mitochondria, we conclude that these differences between genomic and cDNA sequences are the result of RNA editing.

Characteristics of tRNA Editing. In total, 28 nt are inferred to be edited in eight tRNAs studied (22 forming base pairs at the 3’ end of the acceptor stem and six at the discriminator nucleotide position). The observed editing is not limited to a subset of nucleotides and always creates a perfect match between the edited nucleotide and its complement in the 5’ end of the acceptor stem. Even the U-G pair in the completely matched acceptor stem of tRNA(Q) is replaced with a Watson-Crick U-A pair. Therefore, we infer that the editing is template-dependent and uses the 5’ end sequence of the acceptor stem as a template.

By contrast, the editing found in several other species of animals (11-14) is hypothesized to occur by a template-independent mechanism, most likely posttranscriptional polyadenylation. Indeed, in all of those cases but one, only adenosines are added, even when this recreates a mismatch in the acceptor stem. The single exception is platypus tRNA(S) where the 3’ end sequence in the acceptor stem is edited to CCCA (15). This editing also recreates one of the mismatches in the acceptor stem and, therefore, also appears to be template-independent. The mechanism for the latter editing might involve the participation of tRNA nucleotidyltransferase, which is otherwise used for 3’ end CCA addition and/or repair in various organisms (16). Consistent with this idea, the repair of adenosines and cytidines in the tRNA acceptor stem has also been recently reported in human and mouse mitochondrial in vitro systems (17).

Alternative Editing Found in Some tRNAs. Interestingly, the type of editing found in other animal species seems also to occur in L. forficatus mitochondria. Indeed, alternatively edited cDNA sequences were found for several tRNAs. The proportion of these sequences increases with the number of mismatches in the acceptor stem (Table 1). The alternative editing includes polyadenylation and, in two cases, C insertions, exactly the type of editing observed in other animal mitochondrial tRNAs. However, in L. forficatus, this alternative editing does not correct mismatches in the acceptor stem and is usually associated with cDNAs of aberrant length or/and sequence. Thus, it seems likely that these are the products of a different editing system, possibly

Fig. 1. Gene map of L. forficatus mtDNA. Protein and rRNA genes are abbreviated as follows: atp6 and atp8 (genes for subunits 6 and 8 of the F0 ATPase), cox1–cox3 (genes for cytochrome c oxidase subunits 1–3), cob (gene for apocytochrome b), nad1–nad6 and nad4L (genes for NADH dehydrogenase subunits 1–6 and 4L), and rrnS and rrnL (genes for small and large subunit rRNAs). The 22 tRNA genes are identified by the one-letter code for the corresponding amino acid. Two leucine and two serine tRNA genes are differentiated by their anticondon sequence with trnL(uaa) marked as L1, trnL(uaa) as L2, trnS(ucu) as S1, and trnS(ugg) as S2. The direction of transcription for each gene is shown by an arrow. trn (shaded area) is the only gene translocated relative to the arthropod primitive gene arrangement exemplified by the mtDNA of the horseshoe crab Limulus polyphemus (10). The numerals near each tRNA gene indicate the number of mismatches in the acceptor stems inferred from mtDNA sequence.
A Candidate Enzyme for tRNA Editing in *L. forficatus*. The inferred mode of *L. forficatus* mitochondrial tRNA editing would require a 5'-to-3' RNA-dependent RNA polymerase (RdRp). RdRp is encoded by most RNA viruses (except retroviruses), including mitochondria-associated double-stranded RNAs that are widespread in eukaryotes (18). A functional viral RdRp enzyme has been found in mitochondria of the fungus *Ophiostoma novo-ulmi* (19) and a gene homologous to viral RdRp has been incorporated into the *Arabidopsis thaliana* mitochondrial genome (20). Proteins homologous to RdRps have also been found in the nuclear genomes of a variety of eukaryotes; in some cases, they are part of the gene-silencing viral-defense mechanism (21). Thus, the enzyme involved in tRNA editing could be encoded by the nuclear genome and targeted to mitochondria or have a viral origin. If a viral enzyme is participating in the tRNA editing observed in *L. forficatus*, the virus, or at least the RdRp-encoding portion of its genome, would be indispensable for this animal.

Implication for tRNA Processing. The inability of the unedited tRNA gene sequences to form paired acceptor stems raises questions about RNA processing. According to the generally accepted tRNA punctuation model for animal mitochondria (22), the secondary structures of tRNAs serve as the signals for processing of the polycistronic transcript. Their enzymatic removal not only creates tRNAs, but also liberates the intervening mRNAs. Most of the tRNA processing enzymes are rather specific in their requirements (23, 24) and most use the paired acceptor stem of tRNA as at least a part of the template. Consistently, it was demonstrated in plants that the editing of the tRNA acceptor stem has to precede the tRNA 3' end excision (25). In animal mitochondrial tRNAs, however, the processing of the 3' end seems to precede its editing and, therefore, was hypothesized to occur by a passive mechanism, such as the cleavage at the 5' end of the properly folded downstream pre-tRNA present in the same polycistronic transcript (11). Indeed, in all cases but one, the edited part of the tRNA acceptor stem is encoded by the portion of the gene that overlaps with the downstream tRNA gene (for review, see ref. 3) and thus is removed when the transcript of the downstream gene is cleaved at the 5' end. An alternative view would be that the edited portion of the acceptor stem is not encoded at all in these tRNA genes, which are, consequently, immediately adjacent to their downstream tRNA genes.

The presence of similarly overlapping (abbreviated) tRNA genes in *L. forficatus* suggests that the tRNA processing also should precede tRNA editing in this organism. However, because the editing occurs in the transcripts of both overlapping and nonoverlapping genes, the latter often followed by a protein gene, the excision of unedited pre-tRNAs from polycistronic pre-tRNA present in the same polycistronic transcript (11). Indeed, in all cases but one, the edited part of the tRNA acceptor stem is encoded by the portion of the gene that overlaps with the downstream tRNA gene (for review, see ref. 3) and thus is removed when the transcript of the downstream gene is cleaved at the 5' end. An alternative view would be that the edited portion of the acceptor stem is not encoded at all in these tRNA genes, which are, consequently, immediately adjacent to their downstream tRNA genes.

Unusual tRNA Secondary Structures. In addition to the tRNA editing, three types of unusual secondary structures for *L. forficatus* mitochondrial tRNAs have been revealed by the analysis of cDNA data:

(i) Three of the eight tRNAs analyzed deviate from the
Table 1. cDNA sequence variation in circularized Lithobius
mitochondrial tRNAs

<table>
<thead>
<tr>
<th>tRNA</th>
<th>No. clones</th>
<th>cDNA sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA (C)</td>
<td>7</td>
<td>TCA&lt;sup&gt;G&lt;/sup&gt;TCTACGACATCTGAAACCACATCAAGATT</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TCA&lt;sup&gt;G&lt;/sup&gt;T---tatcacacagcagt---AGATT</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TCA&lt;sup&gt;G&lt;/sup&gt;---------actag-----------AGATT</td>
</tr>
<tr>
<td>tRNA (E)</td>
<td>10</td>
<td>TGTAAAAAACCTAGAAAAGACCTTTCTCTATGGAAG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TGTAA---------AGAAAAGACCTTTCTCTATGGAAG</td>
</tr>
<tr>
<td>tRNA (H)</td>
<td>4</td>
<td>TAA&lt;sup&gt;C&lt;/sup&gt;TTTTACCTACCAAGTAAAATAGTTTGAAG</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>TAA&lt;sup&gt;C&lt;/sup&gt;TTTTACC&lt;sup&gt;C&lt;/sup&gt;---ACCAAAAAAAAAAGAG</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>TAA&lt;sup&gt;C&lt;/sup&gt;TTTTGA---------GTAAAATAGTTTGAAG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TAA&lt;sup&gt;C&lt;/sup&gt;TTTTGA---------GAAAATAGTTTGAAG</td>
</tr>
<tr>
<td>tRNA (M)</td>
<td>2</td>
<td>ATCTTTTACAC&lt;sup&gt;C&lt;/sup&gt;ACATGAATAGTCATGAAC</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>ATCTTTTCTTTTT---------TGTAAGATAGTCATGAAC</td>
</tr>
<tr>
<td>tRNA (N)</td>
<td>4</td>
<td>TAA&lt;sup&gt;G&lt;/sup&gt;TTTTACCTACCAAGTAAAATAGTTTGAAG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TAA&lt;sup&gt;G&lt;/sup&gt;TTTT---------AAATGAGATAGTTTGAAG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TAA&lt;sup&gt;G&lt;/sup&gt;TTTTACCTACCAAGTAAAATAGTTTGAAG</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>TAA&lt;sup&gt;G&lt;/sup&gt;TTTTGAA-----TGCTGCAAAGTCATGAAC</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>TAA&lt;sup&gt;G&lt;/sup&gt;TTTTGAA--TGCTGCAAAGTCATGAAC</td>
</tr>
<tr>
<td>tRNA (Q)</td>
<td>6</td>
<td>AGTCTGTTAAATGAAAACCACATCAATAGAG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>AGTCTGTTAATGAAAAC---TATAGATAAG</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>AGTCTGTT---------tttt----TATAGAG</td>
</tr>
<tr>
<td>tRNA (R)</td>
<td>4</td>
<td>GGC&lt;sup&gt;G&lt;/sup&gt;GCCCCCCTAATAGACCACTATAGAGACG</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>GGC&lt;sup&gt;G&lt;/sup&gt;GCCCCCTAAA---------AACACAA</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>GGC&lt;sup&gt;G&lt;/sup&gt;GCCCCCAAAA---------TACACCA</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>GGC&lt;sup&gt;G&lt;/sup&gt;GCCCCCAAAA---------AACACCA</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>GGC&lt;sup&gt;G&lt;/sup&gt;GCCCCCAAAA---------AACACCA</td>
</tr>
<tr>
<td>tRNA (S1)</td>
<td>12</td>
<td>CAC&lt;sup&gt;C&lt;/sup&gt;CCTGTTAATACCTACCAAGTAATCACCG</td>
</tr>
</tbody>
</table>

Underlined nucleotides are part of the primer sequence. Nucleotides in boldface are those inferred to be edited at the 3' end of the acceptor stem. Nucleotides in italics are inferred to be added by some alternative (non-templated) editing process(es). Nucleotides in lowercase are inferred to be artifacts of the experimental procedure. D indicates nucleotide variation (A, T, or G) among the four edited cDNA clones for tRNA (R), possibly caused by posttranscriptional modification at this tRNA position (26).

(1) Classical cloverleaf secondary structure: tRNA(M) has a TV-replacement loop instead of a T-arm and variable (V) loop, whereas tRNA(S1) and tRNA(N) have D-arm replacement loops (Fig. 2). Although mitochondrial tRNA(S1) lacks the D-arm in all metazoan species examined (26), the deviation in the structure of the two other tRNAs is unusual but not unprecedented. The processing of such tRNAs from a presumed polycistronic transcript would present additional problems, because the D-arm and T-arm are generally needed for tRNA recognition by 5' and 3' tRNA end processing enzymes (23, 24). Interestingly, the genes for tRNA(M) and tRNA(N), along with some adjacent nucleotides, have the potential to encode secondary structures that are more like the typical cloverleaf form (Fig. 3) and RNA processing has been observed at their 5' and 3' ends (Table 1). However, no cDNAs were found that indicated production of such tRNAs with unusual secondary structures.

(2) Three tRNAs (those for H, M, and N) have other than the typical 7 bp in their acceptor stem in at least some cDNA clones. All previously studied histidine tRNAs have an additional nucleotide, almost universally a G, at their 5' ends (27), either encoded in the sequence or added posttranscriptionally (28). This nucleotide is located opposite to the discriminator nucleotide in the acceptor stem and may or may not be complementary to it. The inferred L. forficatus tRNA(H) also has an additional nucleotide at the 5' end, an unusual A, potentially encoded by the mtDNA. However, this tRNA differs from all other histidine tRNAs by also having an additional (discriminator) nucleotide at the 3' end of an 8-bp acceptor stem (Fig. 2B).

(3) tRNA(C) is inferred to have an unusual secondary structure with 7 bp in the anticodon stem. Similar structures have been described for mammalian mitochondrial tRNA(S) (29), fungal mitochondrial tRNA(C), echinoderm mitochondrial tRNA(T), and some others (30). The L. forficatus trnC codes for two T-T mismatches in the acceptor stem. These two mismatches were confirmed to be present in the cDNA from the mature transcript product of this tRNA gene (data not shown). Either two T-T mismatches or a single mismatch and bulged T are also present in the anticodon stem of mitochondrial tRNA(C) of Neurospora crassa (30).

Fig. 3. Alternative foldings for the unedited L. forficatus tRNAs with unusual secondary structures. Arrows indicate the deduced processing at the boundaries of these structures as observed in cDNA.

Evolutionary Considerations. The editing observed in L. forficatus tRNAs is more extensive than any previously reported. It appears to occur in all mitochondrial tRNA genes and represents a further step in the degeneration of tRNA genes observed in mitochondria (31). It is likely that, once evolved, this editing would persist in the lineage, because it is difficult to imagine a mechanism that would correct the sequences of all tRNA genes simultaneously and, thus, allow for its loss. Thus, the presence of
this type of tRNA editing may be a nearly irreversible phylogenetic character. However, its usefulness for phylogenetic studies would depend on the probability of it arising independently in different lineages, which itself depends on the underlying mechanism. So far, in addition to L. forficatus, we have observed tRNA genes that encode sequences with the potential for similar editing only among species of onychophorans (unpublished data). Further studies are needed to confirm whether this represents an independent origin or a phylogenetically useful character.

We thank K. G. Helfenbein and J. V. Moran for comments on the manuscript. This work was supported by National Science Foundation Dissertation Improvement Grant DEB 9972712 to W.M.B. and D.V.L., National Science Foundation Grant DEB 9807100 to W.M.B. and J.L.B., and a University of Michigan predoctoral fellowship to D.V.L.