The mitochondrial ND8 gene from *Crithidia oncopelti* is not pan-edited

Alexey A. Aravin, Vyacheslav Yu. Yurchenko, Ekaterina M. Merzlyak, Alexander A. Kolesnikov*

Department of Molecular Biology, Biology Faculty, Moscow State University, Moscow 119899, Russia

Received 17 June 1998

Abstract RNA editing in trypanosomatid mitochondria is a process involving the insertion and deletion of uridine residues within the coding region of maxicircle messenger RNA transcripts. Twelve of the 17 known genes need editing to produce functional molecules. We have analyzed the predicted editing sites for the *Crithidia oncopelti* mitochondrial NADH-ubiquinone oxidoreductase subunit 8 (ND8) gene based on known mRNAs from other trypanosomatid species. All studied ND8 mRNAs undergo editing throughout the coding (and 3′ non-coding) sequences (pan-editing). The 5′ part of the *C. oncopelti* ND8 gene undergoes editing (like in *Leishmania tarentolae* and *Trypanosoma brucei*) while the 3′ part of the pre-edited gene corresponds to the 3′ part of edited ND8 mRNAs from other organisms. The organization of the ND8 gene in *C. oncopelti* mitochondrial DNA differs from all organisms investigated so far – this gene is not pan-edited. We have also localized the guide RNA for cytochrome *b* between 9S rRNA and the ND8 gene. This RNA shows high homology to the gCYb-II gene of *L. tarentolae* and the gCyb gene of *Crithidia fasciculata*. A hypothetical editing pattern for the cytochrome *b* gene in *C. oncopelti* maxicircles is proposed.

Key words: *Crithidia oncopelti*; RNA editing; Maxicircle; ND8

1. Introduction

RNA editing in mitochondria of trypanosomatids is a process involving the insertion and deletion of uridine residues within the coding region of maxicircle messenger RNA transcripts [1]. Twelve of the 17 known mitochondrial genes need editing to produce functional molecules. Some homologous genes from different systematic groups have different editing patterns. In some trypanosomatid species, entire genes are edited throughout their whole length (pan-editing) whereas in others editing is limited to the 5′ termini of the edited domain (5′-editing) [2,3]. Maslov et al. [4] proposed a pan-editing organization model for ancestral cryptogenes and generation of 5′-edited homologues from partially edited RNAs. The editing genetic information was found to reside in small 3′ oligo-uridylated RNAs (termed guide RNA (gRNA)) [5]. These molecules could form anchor hybrids with pre-edited RNAs and mediate the insertion and deletion of Us at precise sites [6]. A hypothetical mechanism for RNA editing ('enzymatic cascade') has been proposed. It includes the endonuclease cleavage of mRNA molecules and the insertion and/or deletion of Us up to full-length complementation with the gRNA molecule [5]. An editing site specific endonuclease, a terminal uridylyl transferase (TUTase) and RNA ligase are present in mitochondrial extracts from trypanosomatids. However, the detailed mechanism of RNA editing still remains unclear.

We report here the primary structure of the ND8 gene from *Crithidia oncopelti*. The results demonstrate the partially edited pattern of this gene.

2. Materials and methods

C. oncopelti strains ATCC 12982 and ATCC 30264 were used for isolation of kinetoplast DNA as described before [7]. The ND8 pre-edited gene and the predicted template for the Cyb gRNA were amplified using PCR with primers 9S3F: TT*AAGCTT*GTTGCCCACCATTCTTTGTAATAAAGACAACGTGCAGT and S960com: TT*GGATCC*AYTGYTCWACWSTTTTATATTCRCATAACTTTTCTGTACC (Y = C/T, W = A/T, S = C/G, R = A/G) at the following temperature conditions: template denaturation 30 s at 94°C; primer annealing 1 min at 45°C (first five cycles) and 50°C (last 25 cycles) and primer extension 1 min at 72°C. The PCR product was purified (QIAgen), cloned to pBluescript KS^+ (Stratagene) vector and sequenced using the Sequenase v. 2.0 sequencing kit (Amersham-USB). *Hin*dIII and *Bam*HI sites (shown by italic) were introduced into the primers.

3. Results and discussion

The *C. oncopelti* maxicircle mitochondrial DNA fragment covering the region with the 3′ part of the 9S rRNA gene, gRNA for cytochrome *b* (gCyt *b*), the full-length ND8 gene and the 3′ region of the ND9 gene has been cloned, sequenced (GenBank accession number AF060882) and analyzed. The order of genes in *C. oncopelti* corresponds to the usual order in all previously investigated kinetoplastids [8]. We did not find any differences between the analyzed sequences of the two different strains, ATCC 12982 and ATCC 30264.

3.1. The ND8 gene from C. oncopelti is not pan-edited

We found hardly any homology within the 5′ region of the *C. oncopelti* ND8 gene with known sequences of the ND8 genes from *Leishmania tarentolae* and *Trypanosoma brucei* analyzed previously [8,9]. However, such absence of homology is a common characteristic of almost any heavily edited trypanosomal gene. The sequence analysis of the *C. oncopelti* ND8 gene showed the start of a reading frame at approximately 100 bp distance from the 3′ end of the 9S rRNA gene. The G-rich area in *L. tarentolae* and *T. brucei* maxicircles starts from the same position and corresponds to the pan-edited ND8 gene. The 5′ part of the ND8 gene (ca. 66%) undergoes RNA editing (as in all kinetoplastids investigated so far). But followed by the edited region the rest of the ND8 *C. oncopelti* sequence corresponds to the 3′ region of ND8 mRNAs from other organisms. This leads us to conclusion that the 3′ part of the ND8 gene might not be edited (Fig. 1A).

*Corresponding author. Fax: (7) (095) 939-46-58.
E-mail: sasha@protein.bio.msu.su

PII: S0014-5793(98)00813-8

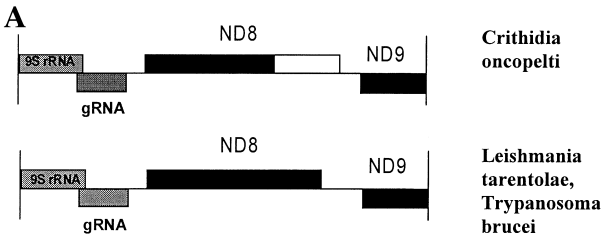

Fig. 1. A: The editing schemes of the pan-edited ND8 gene from *L. tarentolae* and *T. brucei* and the partially edited ND8 gene from *C. oncopelti* (the edited domains are black). B: Alignment of the ND8 genes from different organisms. DNA and mRNA (in italic) sequences are shown for *L. tarentolae* and *T. brucei* (lines 1, 2, 5, 6) [8,9]. The aligned gene sequences from *C. fasciculata* [9] and *C. oncopelti* are shown in lines 3 and 4. The sites of insertion or deletion of Us are shown by dashes. The initiation codon (AUG) and termination codons (UAG or UAA) are in bold.

B

```
T. brucei DNA   (1)    A-G---------GA------GTTTTT-------G---G-----
T. brucei mRNA  (2)    AUGUUUUUUUUGAUUUUUUGUUUUUUUUUUUUGUUUGUUUUU
Cr. fas.  DNA   (3)    A-G---A---G-GA-----GCTTT-CCC-----G---G-----
Cr. onc.  DNA   (4)    ATG--TG---G-GA-----GT----C-------G---G-----
L. tar.   mRNA  (5)    AUGUUUGUUUAUGAUUUUUGUUUUUCUUUUUUUGUUUGUUUUU
L. tar.   DNA   (6)    A-G---G---A-GA-----GTTTT-C-------G---G-----

(1) A----A---G-GT---G---G--G-G--ATTTTC-ATTTTG---ATTTCCCA--GAG--AAC
(2) A---UA--UGUGUUUUGUUUGUUGUGUUA----CUAUUU-GUUUA---CCCAUUGAGUUAAC
(3) ATTTTATTTG-G------A-G--G-G--A----C---AA-----A---CCA--GGAA--AA-
(4) A----A---G-G------G-G--G-G--A----C---AG-T---A---CCT--GGAG--AAC
(5) A---UA--UGUGUUUUUUAUGUUGUGUUA----CUUUGG-UUUUA---CCAUUGGAGUUGAC
(6) A----A---G-G------A-G--G-G--AT---C---GG-----A---CCA--GGAG--GAC

(1) CA--G--AG---A--GG---CG-GG-AACCA------GC-----GTTTT-ATTTTTGG----
(2) CAUUGUUAGUUUAUUGGUU-CGUGGUAACCAUUUUUUGC-----GUUUUUAUU---GGU---
(3) CA--G--AG-A---G-G---CGTGG-AGCCA------ACTTTTTG-----A-----GGTTTT
(4) CA-----AG-AGG-G-G---CG-GG-AA-CATTA----C-----G-----A-----GA----
(5) CAUUGUUAGUAUUUGUGUU-CGUGGUAACCAUUUUUUGC-----GUUUUUAUU---GAU---
(6) CA--G--AG-A---G-GTTTCG-GG-AACCA------GC-----G-----ATTT--GATTT-

(1) G-GG---AGAGCG--G-A----GC--G-CG---A-G-GA---AA---------GCC-------
(2) GUGGUUUAGAGCGUUGUAUU--GCUUGUCGUUUAUGUGAUUUAAUUU-----GCC-------
(3) G-GG---AGAACG--G-ATTTTGCATGTCG---A-G-GA----A-ATTTTTTGTC-------
(4) G-GGT--AGAGCG--G-A----GCTTGTCG---G-GTGA----A-A------GCCTTTTTTT
(5) GUGGUUUAGAACGUUGUAUU--GCCUGCCGUUUAUGUGAUUUUAUAU-----GCC-------
(6) G-GG---AGAACG--G-A----GCC-GCCG---A-GTGA----A-ATTTT--GCC-------

(1) C-ATTTTG---A-----GCA-------GGA-GTTTTTC---G----------G--GGG-GGAG
(2) CUA----GUUUA-----GCAUU-----GGAUGUU---C---G---------UGUUGGGUGGAG
(3) CAA----GT--ATTTTTGC-C------AGA-G-----C---GTTTTTTTTTTTTTGCTCGCAG
(4) CTA----G---A-----GCA-------AGA-G-----CTTTG----------G--GGG-GAAG
(5) CAA----GUUUA-----GCUCU-----AGAUGUU---C---G---------UUGUGUUAGAAG
(6) CAA----G---A-----GC-CTTTTTTAGA-G-----C---G----------G-G--AGAAG

(1) ----GG--GG---CATTTCTG----GCGGA--GA---ACA--GAG--ATTCTTG-TTTC----
(2) UUUUGGU-GGU--CAU--C-GUUUUGCGGAUUGAUUUACAUUGAGUUAU-C--G---UC----
(3) ---A-A--GGTTT-A---C-GCTTCTCAGA-A-A---ACAA--AG--ATTC--G----CTTTT
(4) ---A-G--GGTTTCAT--C-GG----CCGG-G-G---GG----AGTTAT-C--G----C----
(5) UUUAUGU-GGU-U-AU--C-GGUUUUCCGAUGUGUUUAAUAUUAGUUAU-C--G---UC----
(6) ---A-GTTGG----AT--C-GG----CCGA-G-G---AA-A--AG--AT-C--GTTTTC----

(1) --------G--G-A---A--G-GG-----G-A-GC------A-G---GCCCGACA----GATT
(2) --------GUUGUAUUUAUUGUGGUUUUUGUAUGC------AUGUUUGCCCGACA----GAU-
(3) TTTTTTTTG--G-A---A--G-GG-----G-A-GCTTTTTTA-G---G-CCTACA----GATT
(4) --------GTTGTATTTATTGTGGTTTTTGTATGC------ATGTATGTCCAACT----GAT-
(5) --------GUUGUAUUUAUUGUGGUUUUUGUAUGC------AUGUUUGUCCAACA----GAC-
(6) --------G--G-A---A--G-GGTTT--G-A-GC------A-G---G-CCAACATTTTGACT
```

```
(1) TTT--GCCA---AC---GCATTCA--G---G--A-G-G-----G--G---AGCCTTATTTG-A
(2) -----GCCAU-UAC---GCAUUCAUUGUUUGUUAUGUGUUUUUGUUGUUUAGCC--AU--GUA
(3) TC---GCTA---ACTTTGCA--CA-G------A--A----G--A--G-A--GCC--A---G-A
(4) -----GCCGATAAC---CCATTCAATTTTTGTTTTATATTGTGTTTGCTTAGCT--AT--GTA
(5) -----GCCAU-UAC---GCAUUCAUGUUUUUUGUUAUUUUGUUGUUGUAUUGCC--AU--GUA
(6) TTTTTGCCA---ACTT-GCA--CA-G------G--A----G--G--G-A--GCC--A---G-A

(1) ---A---GTGCGCTTT-CTTTCAAG-----A--G---G----G--G--G----A-G--A----
(2) UUUAUU-G-GCGC----C---CAAGUUUUUAUUGUUUG----GUUGUUGUUUUAUGUUAUUU-
(3) ---A-G---GCAC----C---TAAA---G----A---G----G--G--GG---A-G--GTTA-
(4) TTTATT-A-GCCC----C---TAAATTTTTATTATTCG----GTTGTTGTTTTATGTTGTTT-
(5) UUUAUG-C-GCAC----C---UAAAUUUGUUUUAUUUG----GUUGUUGUUUUAUGUUAUUU-
(6) ---A-GTCTGCAC----C---TAAA---GTTT-A---GTTTTG--G--G----A-G--ATTTT

(1) --GA-----A---G-G----G-G-AG--A---A----G--GG-GA---A--GTG---A-GA--
(2) --GAUUUUUAUUUGUGUUUUGUGUAGUUAUUUAUUUUG--GGUGAUUUAUUGUGUUUAUGAUU
(3) --GA-----A---G-G----G---AA---AA-A-ATTA--TAATTTGTTATCTTCTTTTAAAT
(4) --GATTTTTATTTATGTTTTGTTTAATTAATTTACCAA--CACAAAAAACCCCTAACCCCTAA
(5) --GAUUUUUAUUUGUGUUUUGUUUAGUUUAGUGUAUCU--CGAGAAGUAUAUUUGAUUAAUAA
(6) TTGA-----A---G-G----G---AG---AG-G-A-CTTTCGAGAAGTATATTTGATTAATAA

(1) -AATTTAGAATTA--CACGGT
(2) UAA---AGAA--AUUCACGGU
(3) CAAGTTTAACCTAAAATTTAT
(4) ATT---AAGG--TCGCCCTAA
(5) UUU---AAUA--AAUUAAUUU
(6) TTT---AATA--AATTAATTT
```

Fig. 1. (*Continued*).

The results of ND8 edited (*L. tarentolae* and *T. brucei)* and pre-edited (*L. tarentolae, T. brucei, C. fasciculata, C. oncopelti*) sequence alignment (Fig. 1B) allow us to prognosticate precise editing sites for *C. oncopelti* and predict the protein sequence. In all species studied the termination codon is a result of RNA editing (UAG for *T. brucei* and *L. tarentolae;* UAA for *C. fasciculata* and *T. congolense*), but it is encoded without editing in the *C. oncopelti* ND8 gene (UAA). The initiation codon is also encoded in the *C. oncopelti* DNA sequence (in *T. brucei*, *L. tarentolae* and *C. fasciculata* these sites are created by RNA editing). The 3′ tailer third of the *C. oncopelti* ND8 gene showed extremely high homology to the 3′ part of *L. tarentolae* ND8 mRNA (Fig. 1B).

There are two important points in these results. (i) The ND8 gene from the *C. oncopelti* maxicircle is the first of the known ND8 genes with partial editing (ca. 66% of a sequence from the 5′ end is edited while the 3′ part of the gene is not edited). (ii) The encoded 3′ part of the *C. oncopelti* ND8 gene is practically identical with the *L. tarentolae* mRNA sequence.

The alignment of the hypothetical *C. oncopelti* ND8 protein and published ND8 protein sequences from *T. brucei* and *L. tarentolae* is shown in Fig. 2. One hundred and twelve amino acids are identical for all four species; 33 amino acids are non-conservative (13 of them do not change the hydrophilic-hydrophobic protein pattern).

These results are consistent with the hypothesis proposed by Simpson et al. [10,11] on substitution of pan-edited genes by their partially edited analogs. Such substitution could take place several times during protozoan evolution. One possible mechanism is reverse transcription of mRNAs from pan-

```
          1                                                  50
T. bru. MFFFDFLFFFFVCFYMCFVCCVTICLPIELTIVSLLVRGNHFLRFYWCGL
C. onc. MFVCDFCFSFFVCFYMCFLCCVTLVLPLELTIVSVCVRGNHYFRFYWCGL
L. tar. MFVYDFCFSFFVCFYMCFLCCVTLVLPLELTIVSICVRGNHFLRFYWCGL
cons.   MF--DF-FSFFVCFYMCF-CCVT--LP-ELTIVS--VRGNH--RFYWCGL

          51                                                 100
T. bru. ERCIACRLCDLICPSLALDVRVGWSFGGHRFADWFTLSYRRCIYCGFCMH
C. onc. ERCIACRLCDFICPSLDIDVRVGWSLCGHRFSGVFGFSYRRCIYCGFCMH
L. tar. ERCIACRLCDFICPSLALDVRCVRSLCGYRFSDVFNISYRRCIYCGFCMH
cons.   ERCIACRLCDFICPSL--DVR---SL-G-RF---F--SYRRCIYCGFCMH

          101                                          145
T. bru. VCPTDAITHSLFVMCFCCLAMYLLAPKFLLFGCCFMLFDFYLCFV
C. onc. VCPTDAETHSIFVLYCVCLAMYLLAPKFLLFGCCFMLFDFYLCFV
L. tar. VCPTDAITHSCFLLFCCCIAMYLCAPKFVLFGCCFMLFDFYLCFV
cons.   VCPTDA-THS-F-----C-AMYL-APKF-LFGCCFMLFDFYLCFV
```

Fig. 2. Comparison of predicted *C. oncopelti* ND8 protein structure with the published sequences of the same proteins. T. bru., *T. brucei*; C. onc., *C. oncopelti*; L. tar., *L. tarentolae*. The non-conservative amino acids are shown by dashes.

```
L.tar. DNA TTATAAA-A-G------CG-G--AGA-----G--A--------A--A---AGAAAAG

L.tar mRNA UUAUAAAuAuGuuuuuuCGuGuuAGAuuuuuGuuAuuuuuuuuuAuuAuuuAGAAAAG-3′
           || ||||||:||||:||:||||||:|::||||||||:||||||||||||||||||||||

L.tar.gRNA UAUUUUUaUaUaaaagaGUaCaaUUUggaaaCaaUgaaaaaaaUaaUaaaUCUUUUC-5′
          ***** *********|****:*****||***:***:******* *****:*:*

C.onc.gRNA UAUUUCUaUaUaaaaaaGUaUaaUUUaaaaaUaaUaaaaaaaaCaaUaagUUUGCGU-5′
```

Fig. 3. *L. tarentolae* and *C. oncopelti* gRNAs for Cyt *b* editing. The edited region of the Cyt *b* cryptogene, the corresponding mRNA and gRNA are shown for *L. tarentolae* [12]. The initiation AUG codon created by RNA editing is underlined. The nucleotides encoded in the DNA template are in capitals. Identity between *L. tarentolae* and *C. oncopelti* gRNAs is indicated by asterisks. Dashes indicate the editing sites; vertical lines indicate Watson-Crick base-pairing; colons indicate allowable G-U pairing.

edited genes into cDNAs. Then these cDNAs could be rebuilt into the maxicircle genome by homology recombination.

These results allow us to suppose that ND8 genes might be partially edited in other lower trypanosomatids (e.g. *Blastocrithidia* sp., *Leptomonas* sp.).

3.2. The sequences neighboring the ND8 gene in C. oncopelti

The 3′ neighbor of ND8 is the ND9 gene encoded on the opposite DNA strand. Our preliminary results suggest a pan-editing pattern of ND9 RNA processing in *C. oncopelti*. The size of the spacer between the ND8 and ND9 genes can vary among different species. For example, the spacer sizes are 71, 89 and 36 bp for *L. tarentolae*, *C. fasciculata* and *T. brucei* respectively but for *C. oncopelti* the spacer size is 24 bp only. This may indicate an overlap of the transcribed (and edited) 3′ part of the ND9 gene with the 3′ part of the ND8 gene (this situation has been demonstrated for *L. tarentolae* and *T. brucei*).

The Cyt *b* gRNA gene (gCYb-II) in the *L. tarentolae* maxicircle is localized between the ND8 and 9S rRNA genes on the opposite strand. This gRNA edits the 5′ region of the CYb gene (including creation of the initiating AUG codon [5,12]). In this area and at the same position the Cyt *b* gRNA gene in *C. fasciculata* has been localized [13,14]. The comparison of *C. oncopelti* gCyt *b* with the *L. tarentolae* gCYb-II gene is shown in Fig. 3.

The gRNAs for the Cyt *b* gene from different organisms (*L. tarentolae* and *C. oncopelti*) have almost identical sequences (Fig. 3). This means that the structure of these molecules remained very conservative during evolution of protozoan kinetoplastids.

It would be interesting to analyze the same region of other members of kinetoplastid protozoa and trace the direction of RNA editing evolution among organisms from different trypanosomatid genera.

Acknowledgements: We would like to thank Dr. D.A. Maslov for his interest in our work and helpful discussion; Drs. N.S. Entelis, M.N. Merzlyak and I.A. Krascheninnikov for their critical reading of the manuscript and T.A. Yurchenko for her help with pictures. This work was supported by a grant from the Russian Foundation for Basic Research.

References

[1] Benne, R., Van Der Burg, J., Brakenhoff, J., Sloof, P., Van Boom, J. and Tromp, M. (1986) Cell 46, 819–826.
[2] Bhat, G.L., Koslowsky, D.J., Feagin, J.E., Smiley, B.L. and Stuart, K. (1990) Cell 61, 885–894.
[3] Shaw, J., Feagin, J.E., Stuart, K. and Simpson, L. (1988) Cell 53, 401–411.
[4] Maslov, D.A., Avila, H.A., Lake, J.A. and Simpson, L. (1994) Nature 368, 345–348.
[5] Blum, B., Bakalara, N. and Simpson, L. (1990) Cell 60, 189–198.
[6] Blum, B. and Simpson, L. (1990) Cell 62, 391–397.
[7] Simpson, L., Frech, G.C. and Maslov, D.A. (1996) Methods Enzymol. 264B, 99–121.
[8] Thiemann, O.H., Maslov, D.A. and Simpson, L. (1994) EMBO J. 13, 5689–5700.
[9] Sloof, P., Van den Burg, J., Voogd, A. and Benne, R. (1987) Nucleic Acids Res. 15, 51–65.
[10] Simpson, L. and Maslov, D.A. (1994) Curr. Opin. Genet. Dev. 4, 887–894.
[11] Simpson, L. and Maslov, D.A. (1994) Science 264, 1870–1871.
[12] Feagin, J.E., Shaw, J.M., Simpson, L. and Stuart, K. (1988) Proc. Natl. Acad. Sci. USA 85, 539–543.
[13] Sugisaki, H. and Takanami, M. (1993) J. Biol. Chem. 268, 887–891.
[14] Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R. and Noller, H. (1985) Nucleic Acids Res. 13, 4171–4190.