

Maxicircle DNA and edited mRNA sequences of closely related trypanosome species: implications of kRNA editing for evolution of maxicircle genomes

Laurie K. Read¹, Wallace R. Fish², Anthony M. Muthiani² and Kenneth Stuart^{1,3*}

¹Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, WA 98109, USA, ²International Laboratory for Research on Animal Diseases, PO Box 30709, Nairobi, Kenya and ³Pathobiology Department, SC38, University of Washington, Seattle, WA 98195, USA

Received May 4, 1993; Revised and Accepted July 12, 1993

ABSTRACT

kRNA editing produces functional mRNAs by uridine insertion and deletion. We analyzed portions of the apocytochrome b and NADH dehydrogenase subunits 7 and 8 (ND7 and 8) genes and their edited mRNAs in *Trypanosoma congolense* and compared these to the corresponding sequences in *T. brucei*. We find that these genes are highly diverged between the two species, especially in the positions of thymidines and in nucleotide transitions. Editing eliminates differences in encoded uridines producing edited mRNAs that are identical except for the nucleotide substitutions. The resulting predicted proteins are identical since all nucleotide substitutions are silent. A *T. congolense* minicircle-encoded gRNA which can specify editing of ND8 mRNA was identified. This gRNA can basepair with both *T. congolense* and *T. brucei* ND8 mRNA despite nucleotide transitions due to the flexibility of G:U basepairing. These results illustrate how editing affects the characteristics of maxicircle sequence divergence and allows protein sequence conservation despite a level of DNA sequence divergence which would be predicted to be intolerable in the absence of editing.

INTRODUCTION

Many primary transcripts from the maxicircle component of mitochondrial (kinetoplast) DNA (kDNA) are not translatable in kinetoplastid parasites. They require the posttranscriptional addition and deletion of uridines (Us) by kRNA editing to produce translatable mRNA (1–3). Several of these mRNAs are so extensively edited that over 50% of their nucleotides are the result of editing (4–8). The edited sequence information is specified by small RNAs called guide RNAs (gRNAs) many, if not all, of which are encoded in the minicircle component of kDNA in *T. brucei* (9,10, Corell, et al., submitted). A minimum of 180 gRNAs are estimated to be necessary to direct all of the editing observed in *T. brucei*. However, detection of multiple redundant and overlapping gRNAs (10,11, Corell, et al., submitted)

suggests that the actual number of gRNAs is higher in *T. brucei*, perhaps approaching the total coding capacity of minicircles of about 1200 different gRNAs. Editing is probably performed by a macromolecular machinery and candidate complexes composed of several proteins which are presumably nuclear encoded have been identified (Göringer, et al., submitted).

The selective advantage of a process which requires such a substantial allocation of genetic information has been questioned (12). Comparison of DNA, mRNA, and predicted protein sequences of the two most studied kinetoplastids, *T. brucei* and *Leishmania tarentolae*, suggests one selective advantage of kRNA editing: retention of relatively conserved protein sequences despite substantial gene sequence divergence. However, comparison of nucleotide sequences from species as distantly related as *T. brucei* and *L. tarentolae* is complicated by a large number of frameshifts and mutations of non-thymidine residues (for example, compare (4) and (13)). In these cases, it is often difficult to discern the stepwise changes which have occurred during the divergence of maxicircle sequences and the influence of RNA editing on sequence divergence. In contrast, comparison of more closely related species may provide insight into the effects of RNA editing on maxicircle sequence divergence in the evolutionary short term. We compared the DNA and edited mRNA sequences of three *T. brucei* maxicircle genes with those of *T. congolense*. These two species apparently diverged from a common ancestor following its divergence from both the *L. tarentolae* and *T. cruzi* branches (14). Since different mitochondrial protein coding genes can have different rates of change (15), we examined portions of the sequences of three genes whose transcripts have different patterns of editing: apocytochrome b (CYb) and NADH dehydrogenase subunits 7 and 8 (ND7 and ND8). In *T. brucei*, CYb is edited in a small region at its 5' end (16), ND7 is extensively edited in two domains (7), and ND8 is extensively edited throughout its length (5). We find that the edited mRNA sequences of all three transcripts are much more highly conserved between *T. brucei* and *T. congolense* than are the corresponding DNA sequences. The DNA sequences exhibit considerable divergence, primarily in the positions of thymidine (T) residues and in

* To whom correspondence should be addressed at: Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, WA 98109, USA

nucleotide transitions. In each case, editing compensates for gene divergence by U addition and deletion in mRNA resulting in completely conserved protein sequence. In addition, we have identified a minicircle-encoded gRNA in *T. congolense* which could direct editing of ND8 mRNA. Alignment of this gRNA with edited mRNA of both *T. congolense* and *T. brucei* illustrates how gRNA-mRNA interactions may affect the evolution of maxicircle DNA.

MATERIALS AND METHODS

Cells and nucleic acids

All *T. congolense* IL3000 life cycle stages were grown as described (17,18). *T. congolense* epimastigote and bloodstream form (BF) RNA was isolated as previously described (19) and kDNA was isolated from *T. congolense* procyclic forms as published (8).

DNA cloning and sequencing

T. congolense maxicircle fragment 1, which spans the 3' 277 bp of ND8, all of ND9 and US, and the 5' domain of ND7, was sequenced as follows. The ND8 DNA sequence was initially obtained from cDNA clones that were prepared by synthesizing cDNA from 20 µg BF RNA, dC-tailing the cDNA (4), and PCR amplifying twenty per cent of the dC-tailed cDNA using oligonucleotides TbHR1 (CCCTCCTGATAGCGTTGGC-AAAAA), which corresponds to the unedited *T. brucei* ND8 sequence near the 3' end, and BdG₁₀ (CCGGAT-CCGGGGGGGGGG) for thirty cycles of 94°C for 1 min, 45°C for 30 sec, 72°C for 1 min. The resulting product was gel purified (4), filled and phosphorylated as described (20), digested with Bam HI (BRL), ligated into pBluescript II SK⁻ (Stratagene) that was digested with Eco RV (BRL) and Bam HI, and transformed into *Escherichia coli* DH5α F'IQ cells (BRL). Four clones identified by colony PCR with ZL (GGTACCGGGCCCCCCCC) and ZR (4) primers were sequenced with vector-specific primers and an ABI automated sequencer; all had identical sequence. This sequence was assumed to be primarily unedited since editing proceeds generally 3' to 5' (21) and a 3' primer corresponding to *T. brucei* unedited ND8 sequence was used. Based on the cDNA sequence, the TcCR1-1 oligonucleotide (GCGAATT-CGAAGGTGGTGGACCACTCCG), corresponding to sequence 55 nucleotides from the 5' end of the cDNA clones, was generated. Three ng of kDNA was then PCR amplified using TcCR1-1 and MURF 3-1 (CTACTTTTATATTCACATAA-CTTTTCTGTACC), which corresponds to the *T. brucei* sequence in the HR3 region of ND7 (7,22), for 30 cycles of 94°C for 1 min, 55°C for 30 sec, 72°C for 2 min. The resulting product was gel purified, filled, phosphorylated, ligated, and cloned as described above except that the ligation was blunt-ended and the vector was digested with Eco RV and treated with calf intestinal alkaline phosphatase (CIP; Boehringer Mannheim). Two recombinants which were identified by colony PCR with ZL and ZR primers were sequenced in both directions using an ABI automated sequencer and the dye primer technique with vector-specific primers and the dye terminator technique with the following insert-specific primers:

TcFr1-1: CTCTTGTCAGTCTTCTATCCTCC,
TcFr1-2a: TTAATCCATTATTAATTTGACTAAACATACTAC,
TcFr1-3: GTAGGATGTTAGTCAAATTAATAATCC,
TcFr1-4: GGAAAGAGGGCGTTATGGGGTTGGAGG.

The sequences of the two clones were identical with the exception of one base which clearly differed in the region of the US gene (23).

T. congolense maxicircle fragment 2 (spanning the 3' domain of ND7, all of COIII, and the 5' 83 bp of CYb) was sequenced as follows. Three ng of kDNA was PCR amplified using TcND7-1 (GCGGATCCGATGATATAGGATTTGCATCG), which corresponds to the ND7 5' domain sequence obtained from fragment 1, and CYb-CS5 (GCGGATCCTAAACTAAA(A/T)-CC(A/T)AC(A/C)CCATA), a degenerate primer designed based on *T. brucei* and *L. tarentolae* CYb sequence just 3' of the edited domain. The conditions were 94°C for 1 min, 50°C for 30 sec, 72°C for 2 min. The PCR product was gel purified, digested with Bam HI, and cloned as described above using vector digested with Bam HI and treated with CIP. Three transformants identified by colony PCR with the ZL and ZR primers were sequenced in both directions using the ABI automated sequencer and the dye primer technique with vector-specific primers and the dye terminator technique with the following insert-specific primers:

TcFr2-1: CAAAAACCTTGCTCCGCCAATCCCTCTG,
TcFr2-2: CAGGGAGGAGAGAAAGGGGAACG,
TcFr2-3b: CCTCAAACTCCTCTCTTTGATCAGCG,
TcFr2-4: GAGGAGGGCGGTGACAGAGGG,
TcFr2-5: CACCGCCCTCTCTTGTCTGCCCC,
TcFr2-6a: CTTTGGAAAGGGATTTGGGGGGG,
TcFr2-7b: CACGATTCCTCTCCGGAG,
TcFr2-8: GTTTGTGTCGAGAAGATTGTGC.

The sequences of all three clones were identical. Sequences were analyzed using ESEE (E.Cabot, Univ. of Rochester) and DNASTAR (Madison, WI).

Determination of edited RNA sequence

T. congolense edited mRNA sequences were determined using previously described strategies (4-7), which are detailed for each mRNA below.

CYb

cDNA was synthesized from epimastigote stage RNA using XSC-dT₁₇ (GACTCGAGTCGACATCGATTTTTTTTTTTTTTTTTT) and 20 µg RNA and dC-tailed as described (4). Twenty percent of the dC-tailed cDNA was PCR amplified with BdG₁₀ and CYb-CS5 for 30 cycles of 94°C for 1 min, 50°C for 30 sec, and 72°C for 2 min. The resulting product was digested with Bam HI, gel purified, and cloned as described above. Eight clones identified by colony PCR were sequenced using an ABI automated sequencer and the dye primer technique with vector-specific primers.

ND7

T. congolense epimastigote stage RNA was isolated and cDNA synthesized and dC-tailed as described above. Twenty per cent of the dC-tailed cDNA was PCR amplified with MURF 3-1 and BdG₁₀ for thirty cycles of 94°C for 1 min, 50°C for 30 sec, 72°C for 1 min. The resulting product was gel purified, filled, phosphorylated, and cloned as described using vector digested with Eco RV and treated with CIP. Five transformants identified by colony PCR were sequenced as described above.

ND8

T. congolense epimastigote stage cDNA was synthesized using XSC-dT₁₇ as described (4) and PCR amplified using TcCR1 and XSC-dT₁₇ for 30 cycles of 94°C for 1 min, 50°C for 30 sec,

This gRNA can form a duplex with both *T.congolense* and *T.brucei* edited ND8 mRNA, including a 10 or 11 basepair Watson-Crick anchor duplex (Fig. 3B). Thus, if this gRNA were used during editing in both species, it could eliminate sequence differences between pre-edited transcripts arising from different numbers of encoded Ts at ESs in this region. The two gRNA-mRNA mismatches may reflect the origin of the gRNA from a different isolate of *T.congolense* than that used in the present study (17,27). Since gRNA-mRNA interaction involves G:U as well as Watson-Crick basepairing, the A to G transitions in the ND8 mRNA do not affect duplex formation because there are Us in the gRNA at the corresponding positions. This predicts that C to T transitions would also not affect duplex formation providing there was a G in the gRNA. Thus, the same gRNA can direct editing to the same final mRNA sequence despite gene sequence divergence entailing both changes in the positions of Ts and many transitions.

DISCUSSION

Maxicircle sequence divergence between *T.brucei* and *T.congolense* is limited primarily to variations in sequence that can be modified in mRNA by editing and to silent nucleotide substitutions, the majority of which are transitions. In the amino acid coding regions of the genes reported here, 10–12% of ES differ in the number of Ts and 3–10% of the non-T nts are substituted between the two species. These results suggest two implications of kRNA editing for the evolution of maxicircle genomes. Firstly, the positions of Ts in the maxicircle genome appears to be able drift almost at random since gRNA sequences ultimately determine the number and position of Us in the final edited mRNA as illustrated by the alignment of Tc-gND8[101] with edited ND8 mRNA of both *T.congolense* and *T.brucei*. Secondly, editing imposes a constraint on DNA sequence divergence not found in other organisms. Not only must the sequence conserve the ability to encode a functional protein but the RNA must also maintain the ability to duplex with gRNA. This is consistent with the nucleotide substitutions between *T.brucei* and *T.congolense* maxicircle-encoded genes being largely confined to transitions, which can often maintain gRNA-mRNA interaction due to G:U basepairing. Thus, editing allows genes to diverge dramatically, but with certain restrictions.

The conservation of edited mRNA sequence in the face of extensive differences in the positions of encoded Ts requires that gRNAs of appropriate sequence are retained during evolution. Indeed, studies in our laboratory on gRNA diversity in *T.brucei* suggest that gRNAs which specify mature mRNA sequence are conserved despite a rapid rate of minicircle evolution (Riley et al., submitted). Nevertheless, gRNA genes also undergo sequence divergence. Because of the flexibility of G:U basepairing gRNAs can diverge quite extensively and still specify the same edited sequence (10,11; Corell et al., submitted; Riley et al., submitted). Furthermore, minicircle classes are present in multiple copies, so there can be multiple divergent gRNAs, all of which can specify correct editing. For example, we have demonstrated that three *T.brucei* CYb gRNAs which specify the same edited sequence are encoded in minicircles that diverged from a common ancestral molecule (Riley et al., submitted). gRNA genes can also diverge to the extent that they no longer specify the fully edited sequence (Riley et al., submitted). Minicircles encoding these gRNAs may not be eliminated by selection if they encode other functional gRNAs. In addition, genes for gRNAs that do

not specify fully edited mRNA may be agents of evolutionary change if they specify mRNAs encoding proteins with a selective advantage. Thus, kRNA editing allows substantial divergence of both maxicircle and minicircle sequences without consequence to protein sequences. However, the capacity for evolutionary change is also maintained, with mutations in gRNA gene sequence potentially having more effect than mutations in maxicircle gene sequence since the latter are often eliminated in RNA by editing.

DNA sequence mutations affecting the positions of Ts in the maxicircle genome are not completely random. Although edited RNA has an order of magnitude more U insertions than U deletions, DNA sequence differences between *T.congolense* and *T.brucei* are highly skewed toward sites from which Us are deleted by editing (Figs. 2 and 3A). One explanation is that aspects of U deletion and addition differ; there may be fewer constraints on the deletion process. Regions thought to be in the process of editing frequently contain fully edited ESs arising by U deletion surrounded by incompletely edited ESs which require U addition (30; Read and Stuart, unpublished). Thus, deletions may occur preferentially. Furthermore, gRNA-mRNA chimeras with gRNAs attached at ESs requiring deletions are found less frequently than those with gRNAs attached at sites requiring additions (30). This suggests that deletions occur more rapidly than additions, lessening accumulation of the corresponding intermediates in steady state RNA. Thus, U additions may be more complex than U deletions, leading to a bias in selection for mutations at deletion sites. An alternative explanation is that because deletion sites contain multiple Ts in the DNA whereas addition sites typically do not, the former are more prone to mutation by thymine dimerization, which could result in the loss of Ts.

The results presented here are the first report of kRNA editing in *T.congolense*. The high degree of edited mRNA and protein sequence conservation between *T.congolense* and *T.brucei* indicates that they are closely related, a fact that would be obscured if only the DNA sequence were examined. The nearly absolute conservation of edited mRNA and hence protein sequence despite this DNA sequence divergence strikingly illustrates one beneficial aspect of kRNA editing, and suggests one reason why this process has been retained during evolution.

ACKNOWLEDGMENTS

This work was supported by NIH AI14102 to KS who is also a Burroughs Wellcome Scholar in Molecular Parasitology. LKR is supported by NIH postdoctoral fellowship AI08491. The authors thank K.Stankey, J.L.Bennett, T.Hill, J.Bertles, and T.Morales for excellent technical assistance and the members of the Stuart lab for stimulating discussions. We also thank Drs R.Corell and M.Parsons for critical reading of the manuscript and Dr L.Landweber for suggesting the sequence of the Cyb-CS5 primer.

REFERENCES

1. Feagin, J.E. (1990) *J. Biol. Chem.*, **265**, 19373–19376.
2. Stuart, K. (1991) *Annu. Rev. Microbiol.*, **45**, 327–344.
3. Stuart, K. (1991) *Trends Biochem. Sci.*, **16**, 68–72.
4. Read, L.K., Myler, P.J. and Stuart, K. (1992) *J. Biol. Chem.*, **267**, 1123–1128.
5. Souza, A.E., Myler, P.J. and Stuart, K. (1992) *Mol. Cell. Biol.*, **12**(5), 2100–2107.

6. Bhat,G.J., Koslowsky,D.J., Feagin,J.E., Smiley,B.L. and Stuart,K. (1990) *Cell*, **61**, 885–894.
7. Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L., Feagin,J.E. and Stuart,K. (1990) *Cell*, **62**, 901–911.
8. Feagin,J.E., Abraham,J.M. and Stuart,K. (1988) *Cell*, **53**, 413–422.
9. Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
10. Koslowsky,D.J., Riley,G.R., Feagin,J.E. and Stuart,K. (1992) *Mol. Cell. Biol.*, **12**(5), 2043–2049.
11. Pollard,V.W., Rohrer,S.P., Michelotti,E.F., Hancock,K. and Hajduk,S.L. (1990) *Cell*, **63**, 783–790.
12. Weiner,A.M. and Maizels,N. (1990) *Cell*, **61**, 917–912.
13. Maslov,D.A., Sturm,N.R., Niner,B.M., Gruszynski,E.S., Peris,M. and Simpson,L. (1992) *Mol. Cell. Biol.*, **12**(1), 56–67.
14. C.A.Hoare, (1972) *The Trypanosomes of Mammals. A Zoological Monograph*. Blackwell, Oxford.
15. M.W.Gray, (1989) *Annu. Rev. Cell Biol.*, **5**, 25–50.
16. Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *Cell*, **49**, 337–345.
17. W.R.Fish, C.W.Muriuki, A.M.Muthiani, D.J.Grab, J.D. Lonsdale-Eccles (1989) *Biochemistry*, **28**, 5415–5421.
18. E.J.Bienen, P.Webster, and W.R.Fish (1991) *Exp. Parasitol.*, **73**, 403–412.
19. Feagin,J.E., Jasmer,D.P. and Stuart,K. (1985) *Nucl. Acids Res.*, **13**, 4577–4596.
20. Bhat,G.J., Lodes,M.J., Myler,P.J. and Stuart,K.D. (1991) *Nucl. Acids Res.*, **19**, 398.
21. Abraham,J.M., Feagin,J.E. and Stuart,K. (1988) *Cell*, **55**, 267–272.
22. Simpson,L., Neckelmann,N., de la Cruz,V.F., Simpson,A.M., Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *J. Biol. Chem.*, **262**, 6182–6196.
23. Read,L.K., Jacob,A.N.K., Fish,W.R., Muthiani,A.M., and Stuart,K. (1993) *Mol. Biochem. Parasitol.* In press.
24. Stuart,K. and Feagin,J.E. (1992) In *Int.Rev.Cytol*, Vol. 141. Academic Press, pp. 65–88.
25. Stuart,K., Corell,R.A., Göringer,H.U., Koslowsky,D.J., Myler,P.J., Read,L.K., Riley,G.R., Shu,H-H. and Souza,A.E. (1993) In Brennicke,A. and Kück,U. (eds.), *Plant Mitochondria*. VCH Verlagsgesellschaft/VCH Publishers, pp. 10–23. In press.
26. Borst,P., Fase-Fowler,F., Weijers,P.J., Barry,J.D., Tetley,L. and Vickerman,K. (1985) *Mol. Biochem. Parasitol.*, **15**, 129–142.
27. Nasir,A., Cook,G.A. and Donelson,J.E. (1987) *Mol. Biochem. Parasitol.*, **24**, 295–300.
28. Read,L.K. and Stuart,K. (1993) *Mol. Biochem. Parasitol.* In press.
29. Jasmer,D.P. and Stuart,K. (1986) *Mol. Biochem. Parasitol.*, **18**, 321–331.
30. Read,L.K., Corell,R.A. and Stuart,K. (1992) *Nucl. Acid. Res.*, **20**(9), 2341–2347.