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

Laurie K.Read<sup>1</sup>, Wallace R.Fish<sup>2</sup>, Anthony M.Muthiani<sup>2</sup> and Kenneth Stuart<sup>1,3\*</sup> <sup>1</sup>Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, WA 98109, USA, <sup>2</sup>International Laboratory for Research on Animal Diseases, PO Box 30709, Nairobi, Kenya and <sup>3</sup>Pathobiology 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 nuclearly 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  $\mu$ g 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  $BdG_{10}$  (CCGGAT-CCGGGGGGGGGGGG) 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<sup>-</sup> (Stratagene) that was digested with Eco RV (BRL) and Bam HI, and transformed into Escherichia coli DH5a 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-CGAAGGTGGTGGACCAGACTCCG), 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 (CTACACTTTTATATTCACATAA-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 vectorspecific primers and the dye terminator technique with the following insert-specific primers:

TCFt1-1: CTCTTGTCAGTCTTCTTATCCTCC, TCFt1-2a: TTAAATCCATTATTAATTTGACTAAACATACTAC, TCFt1-3: GTAGGATGTTTAGTCAAATTAATAATCC,

TcFr1-4: GGAAAGAGGGCGTTATGGGGGTTGGAGG.

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: CAAAAACCTTGTCTCCGCCAATCCCTCTG, TcFr2-2: CAGGGAGGAGAGAAAGGGGAACG, TcFr2-3b: CCTCAAAACTCCTCTCTTTGATCACGC, TcFr2-4: GAGGAGGCGGGTGACAGAGGG, TcFr2-5: CACCGCCTCCTCTTGTTCTGCCCG, TcFr2-6a: CTTTTGGAAGGGGATTTTGGGGGGGC, TcFr2-7b: CACGATTCCTCTCCCGGAG, TcFr2-8: GTTTGTGTCGCAGAAGATTGTGC.

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

#### 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 BdG<sub>10</sub> 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-dT<sub>17</sub> as described (4) and PCR amplified using TcCR1 and XSC-dT<sub>17</sub> for 30 cycles of 94°C for 1 min, 50°C for 30 sec,

and  $72^{\circ}$ C for 1 min. The resulting product was gel purified as described (4), digested with Eco RI and Sal I (BRL), and cloned as described above using vector digested with Eco RI and Sal I. Eight transformants identified by colony PCR were sequenced as described.

## RESULTS

#### DNA and edited mRNA sequences

We cloned and sequenced two DNA fragments that correspond to a 2.07 kb portion of the T. congolense maxicircle spanning the 3' 277 bp of the NADH dehydrogenase subunit 8 (ND8) gene. all of the ND9 (previously called CR2; Souza, et al., submitted), US (unidentified sequence between ND9 and ND7 (3)), ND7, and cytochrome oxidase subunit III genes, and the 5' 83 bp of the apocytochrome b (CYb) gene. The sequence showed that the same genes are contiguous and in the same order as in T.brucei and L. tarentolae (3, 22, 24, 25). The T. congolense sequence has only 34% identity with that of T.brucei when a gap penalty of 100 is applied during alignment, but 83% identity if no gap penalty is applied. Gapping increases identity because the two sequences differ primarily in the positions of Ts as described below. The low identity of the non-gapped sequences is consistent with restriction polymorphisms and reduced hybridization with T.brucei probes in Southern blot analysis (23, 26).

The 5' portion of the *T.congolense* CYb gene, which specifies the editing domain of the mRNA, has an interesting pattern of sequence divergence from that of *T.brucei* (Fig. 1). The 5' untranslated region (UTR) sequence is very diverged and cannot be aligned even if Ts (Us) are deleted. The sequence downstream of the initiation codon which is created by editing differs from that of *T.brucei* by the presence of three additional Ts in *T.congolense* and a G to A transition. *T.congolense* edited CYb mRNA sequence was determined from cDNA clones generated by anchor PCR of C-tailed first strand cDNA with B-dG<sub>10</sub> and a 3' primer corresponding to the region just 3' of the editing domain. Comparison of the *T.congolense* and *T.brucei* CYb mRNA sequences reveals that U addition to the *T.brucei* mRNA during editing abrogates the gene sequence differences resulting from the additional Ts in the *T.congolense* DNA (Fig. 1). In

| TC CYB DNA<br>TC CYB RNA<br>TB CYB RNA            | GCAGGAAGCAGTTAGTGTGTAGTTAGAGTAAATA A A A G CG G AGA A A A A AGAAA G TT G TT   |
|---|---|
| protein   | NFRCRFLLFFLLFRNL C C  |
| Tc CYb DWA<br>Tc CYb RNA<br>Tb CYb RNA<br>protein | TCTTTTATGTCAGGTGTTGTTTGTATAGAATA<br>UCULINANGCUCGGUGLUGUNUMGANA<br>UCULINANGCUCGGUGLUGUNUMGANA<br>L L N S C L Y R I |

Figure 1. The editing domain (16) of *T. congolense* (Tc) CYb DNA and edited RNA aligned with *T. brucei* (Tb) edited CYb RNA and the predicted amino acid sequence. The *T. congolense* edited RNA sequence was deduced from the consensus of eight dC-tailed cDNA clones. Six of the cDNA clones (which had a total of three different 5' ends) contained identical edited sequence matching the consensus shown. The other two clones were partially edited in the 3' end of the editing domain and their edited sequence did not match the consensus. The 5' end shown is the 5'-most site identified and was present in one clone. The 5' end shown is the 5'-most site identified and was present in one clone. The 5' end shown of clone lone) nucleotides. Lower case u's indicate uridines inserted by editing. Boxed regions are editing sites within the coding sequence which have a different number of encoded Ts between *T. congolense* and *T. brucei* and single underlined nucleotides indicate a nucleotide substitution. The initiation codon created by editing is double underlined. Sequences 5' of the initiation codon cannot be aligned as described in the text.

addition, the nucleotide substitution is at the third position and is silent. The in-frame AUG previously noted (16) is also conserved. Thus, while editing extends a few sites further 5' in *T. congolense*, the CYb mRNA encodes an identical amino acid sequence in the region examined.

The pattern found for CYb also occurs in the ND7 gene which encodes a much more extensively edited RNA (Fig. 2). The 5' and 3' UTRs are highly diverged and cannot be easily aligned even if Ts (Us) are disregarded although the divergence is less for the 3' UTR. In contrast, the amino acid coding sequence of the *T.congolense* ND7 DNA can be readily aligned with that of *T.brucei* if gaps are introduced opposite Ts or Us. 59 of 549 editing sites (sites between non-T nucleotides; ESs) differ in the number of Ts between *T.congolense* and *T.brucei* DNAs (boxed regions in Fig. 2). These ESs are throughout the ND7 sequence

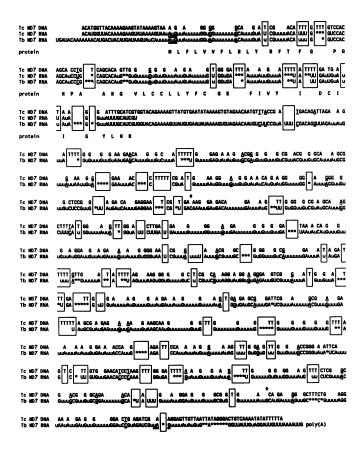


Figure 2. T. congolense ND7 DNA and 5' domain edited RNA sequences aligned with T.brucei edited ND7 RNA and deduced amino acid sequences. T.congolense edited RNA sequence was determined from the consensus of five dC-tailed cDNA clones. Two clones had identical edited sequence matching the consensus shown; a third clone differed from the consensus by one additional T 5' of the initiation codon. The other two cDNAs were partially edited at their 5' ends and matched the consensus sequence from their 3' ends to 6 and 26 ESs from the initiation codon, respectively. Four of the five clones had identical 5' ends at the site shown; the 5' end of the fifth clone was 25 nt 5' of this site. The ND7 poly(A) addition site was not determined and the same number of encoded nucleotides 3' of the termination codon as are present in T.brucei are shown. Asterisks indicate encoded uridines deleted by editing. Diamonds above the sequence denote non-T nucleotides in T.brucei which align across from a gap in the T.congolense sequence. Initiation and termination codons created by editing are double underlined. The 5' and 3' UTRs of the T. congolense and T. brucei sequences cannot be aligned as described in the text. All other symbols as in Fig. 1.

| A   |   |  |  |  |  |  |
|---|---|--|--|--|--|--|
| TH NDS DNA<br>Th NDS RNA                          | ANCA E TAS E AS CE SE ANTT CCA. ACC TTT A TTT ES E SE AGAGES & AC CC A CC AND A CC A ACC A CC A CC A C  |  |  |  |  |  |
| TC NDS DNA<br>TC NDS RNA<br>Tb NDS RNA            | СА. С.  |  |  |  |  |  |
| protein   | A LDV RVGUSFGG H R  |  |  |  |  |  |
| sin08(1012  |   |  |  |  |  |  |
| Tc NDS DNA<br>Tc NDS RNA<br>Tb NDS RNA<br>protein | C C C C C C C C C C C C C C C C C C C   |  |  |  |  |  |
|   |   |  |  |  |  |  |
| TC NDS DNA<br>TC NDS RNA<br>TL NDS RNA            | CCBACMATTITTECCA ACACATTTT CA E E A E E E BOCC A T E A A E TT ECCC TTT<br>CCBACMATTITTECCA ACACATTT CA E E A E E E BOCC A T E A A E TT ECCC TTT<br>CCBACMATTITTECCA ACACATTL CA UN CAUGULAUGULAUGULAUGUCAUABOCC |  |  |  |  |  |
| protein   | PTD AITH SLFVHCFCCLA NYLL AP  |  |  |  |  |  |
| TC NDS DNA<br>TC NDS RNA<br>Tb NDS RNA<br>protein |   |  |  |  |  |  |
| procenn   |   |  |  |  |  |  |
| TC ND8 DNA<br>TC ND8 RNA<br>Tb ND8 RNA            | TTTTGGGTTTAGAAATCAAGTATTATAAAGTATAAAATTAAAAATTAAAGT<br>"Uluggglungaaatcaagnungaagnungaagnungaantaaattaang doly(a)<br>Gugulaagaulaa  |  |  |  |  |  |
| В   |   |  |  |  |  |  |
| D   |   |  |  |  |  |  |

| T. congolense mRNA | CAUCOUUUOCĂ*GAU | uGAUUUACAUU§A****G | uuAUC (  | ;**** <u>C</u> CGuuGuAuuuAuuGu |
|--------------------|-----------------|--------------------|----------|--------------------------------|
|                    |                 | =   =    =         | 11       | ×  ]]                          |
| Tc-gND8(101)       | UAAAGUGU AUA    | AUUAAGUGUAGUU C.   | AGU AG ( | : AGCAACAUAAA                  |
|                    | 1   11 1 ×      | =   =    =         | I=       |                                |
| T. brucei mRHA     | CAUCGUUUGCG GAU | uGAUUUACAUUGA G    | uuA*UC** | ** <u>U</u> CGuuGuAuuuAuuGu    |

Figure 3. Partial T. congolense ND8 DNA and edited RNA sequences aligned with partial T.brucei edited ND8 RNA sequence, the deduced amino acid sequence encoded by the region of T. congolense RNA shown, and Tc-gND8[101] gRNA. A. T. congolense edited RNA sequence was deduced from the consensus of eight cDNA clones, seven of which were edited up to the point indicated by the arrow and one of which was edited almost to its 5' end. The 3' end of the T. congolense sequence was tentatively identified by poly(A) in one clone. Six other clones had poly(A) 12 nucleotides 5' of this site (five of the As could be encoded) and one clone had poly(A) 95 nucleotides further 5'. The overlined sequence indicates the region of Tc-gND8[101] homology and the diamond denotes a cytidine in T. congolense opposite a gap in T. brucei. Other symbols as in Figs. 1 and 2. The ND8 gene and mRNA sequences of T.brucei extend 5' for 68 and 134 nucleotides, respectively, to sequence presented here (5). B. The Tc-gND8[101] gRNA aligned with the region of T.congolense and T.brucei ND8 mRNA to which it is homologous. Tc-gND8[101] was identified in the previously published T.congolense Tck.1 minicircle sequence (27). Vertical lines signify Watson-Crick basepairs, colons signify G:U basepairs, and Xs signify mismatches. All other symbols as in Figs. 1 and 2.

except for the internal region that encodes a portion of the mRNA sequence, termed the HR3 region, that is not edited and is also conserved between T. brucei and L. tarentolae (7,22). In addition, 48 nucleotide substitutions occur out of 557 potential positions (underlined in Fig. 2). Thirty-nine of the substitutions are A to G transitions, seven are C to T transitions, and one is an A to T transversion. A to G transitions have occurred equally in both species, but all C to T transitions result in Ts in T. congolense and Cs in T. brucei. T. congolense edited mRNA sequence of the ND7 5' editing domain (the sequence 5' of the HR3 region (7)) was obtained from cDNAs amplified using anchor PCR with BdG<sub>10</sub> and a 3' primer corresponding to the T.brucei (and, fortuitously, the T. congolense) DNA sequence in the HR3 region. Despite the sequence differences between T. congolense and T.brucei at the DNA level, the edited mRNA sequences in the region examined are identical within the coding sequence except for the nucleotide substitutions (Fig. 2). Thus, editing has compensated for the differences in the number of Ts between

DNA sequences of the two species. This includes the AUG initiation codon which is created by editing in the same position in *T. congolense* as in *T. brucei* (double underline in Fig. 2). In addition, all nucleotide substitutions are at silent third positions. The resulting predicted ND7 N-terminal amino acid sequences are 100% conserved between the two species within the 49 amino acid region encoded by the *T. congolense* edited mRNA consensus sequence. If the number of Us is kept equal to those in *T. brucei* at each ES throughout the remainder of the RNA, and the two gaps (diamonds in Fig. 1) are compensated for by U addition (see ND8 RNA sequence), the predicted 386 amino acid proteins would have only three amino acid differences, all of which are conservative.

We determined most of the T. congolense ND8 DNA and edited mRNA sequences and found that, again, despite substantial gene sequence divergence, the RNA predicts a protein identical to that encoded in T.brucei edited mRNA (Fig. 3A). ND8 DNA sequences of T. congolense align with those of T. brucei if gapping is allowed for T residues with only minor similarities between sequences immediately 3' of the termination codon reminiscent of ND7. Within the coding sequences, 21 out of 177 ES contain different numbers of Ts. Of the 18 nucleotide substitutions (underlined in Fig. 3A) spaced throughout the ND8 sequence, 16 are A to G transitions and two are C to T transitions. Edited mRNA sequences were determined from cDNA clones obtained by PCR amplification with  $XSC-dT_{17}$  and a 5' primer corresponding to ND8 DNA sequence near the 5' end of the gene. We find that editing of ND8 mRNA compensates for the differences in the numbers of Ts such that the final edited mRNAs differ only in the nucleotide substitutions which are silent. A termination codon (double underlined in Fig. 3A) is created by editing at the same site in both species (UAA in T. congolense vs. UAG in T.brucei). One C in the T.congolense DNA sequence opposite a gap in the *T.brucei* sequence (diamond in Fig. 3A) is compensated for by deletion of the U immediately 3' of the C in T.congolense and insertion of a U in this position in T.brucei. The resulting predicted amino acid sequences are identical in the two species despite considerable divergence at the DNA level.

In all three genes examined, sequences of a given gene contain approximately the same total number of Ts in both species; therefore, transcripts from these genes require approximately the same amount of editing to achieve the final mRNA sequence. However, DNA sequence divergence affecting the number of Ts at each ES does not appear to be random. Inter-species variations in the number of encoded Ts in both the ND7 and ND8 genes are strongly biased toward ESs from which Us will be deleted from the mRNA by editing. In fact, 86% of *T.brucei* ND7 and ND8 ESs from which Us are deleted from the mRNA are altered in the DNA compared to *T.congolense*, while only 6% and 4% of ESs requiring U addition in the mRNA are altered in the ND7 and ND8 genes, respectively.

## gRNAs

Using computer assisted analysis (5) of two published *T.congolense* minicircle sequences (27) we have identified a gene for a putative gRNA (Tc-gND8[101]; for description of gRNA nomenclature, see (10)) which could direct editing of ND8 mRNA. Tc-gND8[101] is encoded on the same DNA strand of the minicircle as all gRNAs identified in *T.brucei* and lies between 19 basepair inverted repeats (28) in a position relative to the 13 bp conserved sequence roughly equivalent to cassette four in *T.brucei* (29, R.A.Corell, P.J.Myler, and K.Stuart, unpublished).

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.

- Bhat,G.J., Koslowsky,D.J., Feagin,J.E., Smiley,B.L. and Stuart,K. (1990) Cell, 61, 885-894.
- 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.
- Koslowsky, D.J., Riley, G.R., Feagin, J.E. and Stuart, K. (1992) Mol. Cell. Biol., 12(5), 2043-2049.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Read, L.K., Jacob, A.N.K., Fish, W.R., Muthiani, A.M., and Stuart, K. (1993) Mol. Biochem. Parasitol. In press.
- Stuart,K. and Feagin,J.E. (1992) In Int.Rev.Cytol, Vol. 141. Academic Press, pp. 65-88.
- 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 Verlagsgesellschft/VCH Publishers, pp. 10-23. In press.
- Borst, P., Fase-Fowler, F., Weijers, P.J., Barry, J.D., Tetley, L. and Vickerman, K. (1985) Mol. Biochem. Parasitol., 15, 129-142.
- 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.