Conserved genes encode guide RNAs in mitochondria of Crithidia fasciculata

Hans van der Spek, Gert-Jan Arts,
Richard R.Zwaal, Janny van den Burg,
Paul Sloof and Rob Benne

E.C.Slater Institute for Biochemical Research, University of
Amsterdam Academic Medical Centre, Meibergdreef 15, 1105 AZ
Amsterdam, The Netherlands

RNA editing is the post-transcriptional alteration of the nucleotide sequence of RNA, which in trypanosome mitochondria is characterized by the insertion and deletion of uridine residues. It has recently been proposed that the information for the sequence alteration in *Leishmania tarentolae* is provided by small guide (g) RNAs encoded in the mitochondrial DNA [Blum *et al.* (1990) *Cell*, **60**, 189–198]. We are studying the mechanism of RNA editing in the insect trypanosome *Crithidia fasciculata* and report that: (i) a full length, conventional DNA gene or an independently replicating RNA gene that could encode the edited MURF3 transcript is absent when probed for in sensitive, calibrated assay systems; (ii) in all cases (seven) investigated in *C.fasciculata* so far, putative gRNA genes are found in a position in the mitochondrial DNA virtually identical to that in *L.tarentolae* and (iii) also in *C.fasciculata*, the putative gRNA genes are transcribed into small RNAs with discrete 5' ends. These results provide strong evolutionary evidence in support of the participation of gRNAs in RNA editing. Remarkably, in *C.fasciculata* the basepaired region of some putative gRNA:mRNA hybrids contains a C:A non-Watson–Crick basepair.
Key words: gene expression/guide RNA/mitochondrion/ RNA editing/trypanosomes

Introduction

Mitochondrial (mt) DNA of trypanosomes is composed of maxicircles, containing the mt genes, and minicircles (reviewed in Benne, 1985; Simpson, 1986). Analysis of gene expression in trypanosome mitochondria revealed the existence of an RNA editing process that alters transcript sequences by inserting and/or deleting uridines (Benne *et al.*, 1986; Feagin *et al.*, 1987; reviewed in Simpson and Shaw, 1989; Stuart, 1989; Benne, 1989, 1990). Other, probably non-related types of RNA editing have recently been shown to occur in diverse genetic systems such as mammalian cells (C to U conversion in apolipoprotein B100 RNA, Chen *et al.*, 1987; Powell *et al.*, 1987), plant mitochondria (C to U conversions, Gualberto *et al.*, 1989; Covello and Gray, 1989, and U to C conversions, Gualberto *et al.*, 1990; Schuster *et al.*, 1990), paramyxo viruses (G insertion, Thomas *et al.*, 1988; Cattaneo *et al.*, 1989; Vidal *et al.*,

1990) and *Physarum polycephalum* mitochondria (C insertion, Mahendran *et al.*, 1991).

In trypanosomes such as *Leishmania tarentolae* and *Crithidia fasciculata*, RNA editing is frequently a local process, which only affects small areas of an RNA. In those cases, editing creates transcripts which are translationally competent: DNA encoded shifts in the reading frame are removed and/or translational initiation codons created (reviewed in Simpson and Shaw, 1989; Stuart, 1989; Benne, 1989, 1990). In *Trypanosoma brucei*, however, transcripts from the cytochrome *c* oxidase III (COIII), MURF3 (most likely encoding an NADH dehydrogenase subunit, termed ND7 by Bhat *et al.*, 1990) and ATPase 6 (formerly called MURF4) genes are altered over their entire length by massive U insertion/deletion (Feagin *et al.*, 1988; Koslowsky *et al.*, 1990; Bhat *et al.*, 1990).

The RNA editing concept is largely based on the fact that despite considerable efforts of several groups (reviewed in Simpson and Shaw, 1989; Stuart, 1989; Benne, 1989, 1990) no evidence for the existence of 'edited' gene versions could be obtained. The mechanism of the editing process, however, remains largely unclear. Most of the data indicate that RNA editing occurs post-transcriptionally: first, unedited RNAs which can serve as a substrate for the editing process are always present (Feagin *et al.*, 1988; Shaw *et al.*, 1988; van der Spek *et al.*, 1988) and second, editing proceeds under conditions that block transcription (Harris *et al.*, 1990). Moreover, numerous partially edited RNAs exist that would each require a separate template. Instead, it can be argued from the characteristics of these RNAs, that they are created by an insertion/deletion process proceeding in the 3' to 5' direction (Abraham *et al.*, 1988; Sturm and Simpson, 1990a). Recently, Blum *et al.* identified small transcripts in mitochondria of the lizard trypanosome *L.tarentolae* that can basepair with edited mRNA sequences. The authors proposed an attractive model, which accommodates most of the data and accounts for the precision and specificity of the RNA editing process. The small RNAs, which are thought to provide the sequence information and to lead the RNA editing machinery (Blum *et al.*, 1990), were termed guide (g) RNAs. They cannot be considered as templates in the conventional sense, since apart from A:U and G:C, also non-Watson–Crick G:U basepairs are included in the mRNA:gRNA hybrids. Some gRNA genes are located in intergenic regions of maxicircle DNA, others particularly in *T.brucei* have been localized to minicircle DNA (Bhat *et al.*, 1990; Koslowsky *et al.*, 1990; Sturm and Simpson, 1990b; Pollard *et al.*, 1990). The precise mode of action of the gRNAs is still obscure.

The picture that is emerging from these data appears to be in conflict with a study performed by Volloch *et al.* These authors suggest that in *T.brucei* no precursor–product relationship between unedited and edited COIII transcripts exists and that editing is a transcriptional process (Volloch *et al.*, 1990). This could be in line with the proposal by Maizels

and Weiner that independently replicating RNAs provide the editing information (Maizels and Weiner, 1988). Moreover, all of the available published evidence demonstrating the absence of alternative, edited, gene versions stems from, relatively, insensitive Southern blot analyses, which might have missed low copy number genes (reviewed in Simpson and Shaw, 1989; Stuart, 1989; Benne, 1989, 1990). In an attempt to settle this issue, we have searched for sequences in (mt) DNA and RNA that could serve as a template or guide for edited mt RNAs in *C.fasciculata*. Our experimental approach involves computer aided sequence analysis of maxicircle DNA, PCR analysis of DNA and Northern blot and sequence analysis of RNA. Our data virtually rule out the existence of DNA and RNA templates and, instead, fully support the gRNA model proposed for *L.tarentolae*. Remarkably, next to the more familiar G:U basepairs, some of the putative mRNA:gRNA hybrids appear to contain another non-Watson−Crick basepair: A:C.

Results

Absence of conventional templates

The MURF3 transcript in *C.fasciculata* is edited in two domains: five uridines are inserted at an internal position, removing a (+1) shift in the reading frame and 22 uridines are inserted near the 5' end, thereby shifting an AUG codon

out of phase with respect to the rest of the reading frame (van der Spek *et al.*, 1988). To exclude the possibility of having missed low copy 'edited' genes (or fragments), PCR analysis was used. Figure 1 shows an experiment in which an edited version of the MURF3 gene was amplified using two oligonucleotides complementary to the two editing domains of this transcript (Figure 1A). Using 250 ng of total *C.fasciculata* DNA, an amplified fragment fails to show up when edited primers are used, under conditions that allow amplification with unedited primers and 5 pg of DNA (Figure 1B, lanes 4 and 5 respectively). To calculate the detection limit in this experiment, control reactions were performed with samples that contained low amounts of edited cDNA in addition to the total DNA. In a background of total DNA representing 5×10^6 cells, 500 copies of cDNA are readily amplified (Figure 1B, lane 2).

As a next step, we set out to look for RNA genes coding for the edited sequences of MURF3. Total *C.fasciculata* RNA was isolated and Northern blots were hybridized with sense MURF3 RNA (Figure 2). Hybridizing transcripts cannot be detected using this approach. The experiment was calibrated by adding fixed amounts of *in vitro* synthesized antisense 'edited' MURF3 RNA to the total RNA samples before running the gel (Figure 2, lanes 1−4). We calculated that the presence of one copy or less of an edited RNA gene
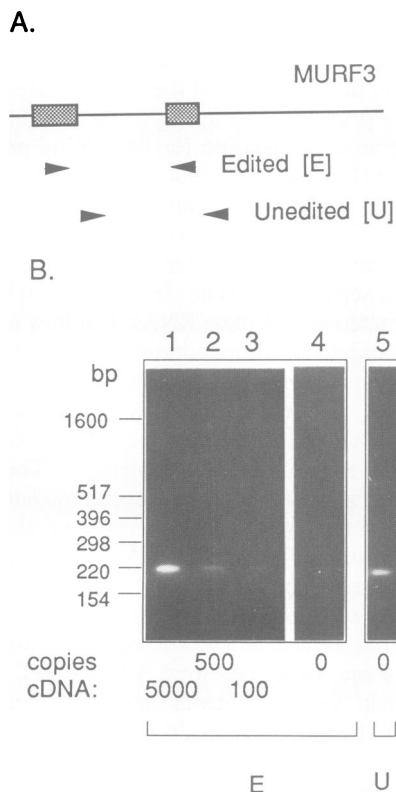

Fig. 1. Search for an edited MURF3 DNA gene by PCR analysis. **A.** Schematic representation of the MURF3 transcript containing two editing domains. The two primer sets used in the PCR shown in panel B are indicated (E = 'edited', U = 'unedited'). **B.** PCR amplification of total DNA (250 ng/lane) of *C.fasciculata* using 'edited' oligonucleotides (lanes 1, 2, 3 and 4). As an internal control, edited cDNA is added prior to amplification in lanes 1 (25 fg), 2 (2.5 fg) and 3 (0.5 fg). Lane 5 contains a control reaction using 5 pg of total DNA and unedited oligonucleotides as a positive PCR control of total DNA.
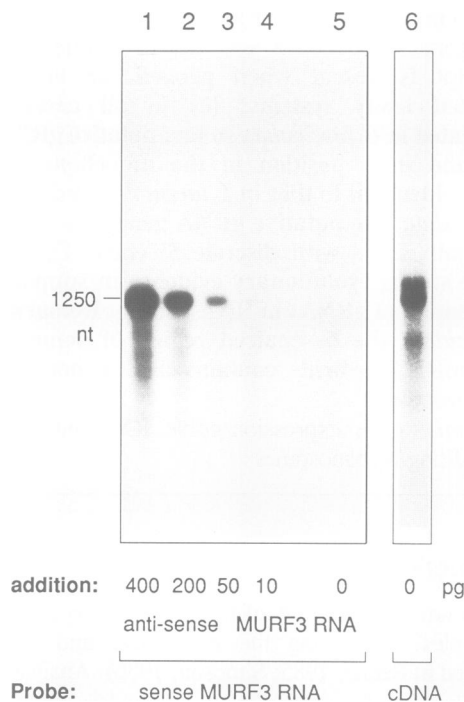
Fig. 2. Search for an (anti-sense) MURF3 RNA gene. Northern blot of *C.fasciculata* total RNA (30 μg/lane). As an internal control, *in vitro* synthesized 'edited' anti-sense MURF3 RNA was added in lanes 1 (400 pg), 2 (200 pg), 3 (50 pg) and 4 (10 pg). The blot was hybridized with *in vitro* synthesized [α-³²P]UTP labelled MURF3 sense RNA. Calculation: the yield of total RNA is routinely 20 mg from one litre of culture (= 5×10^{10} cells). Lanes 1−6 contain 30 μg total RNA, which corresponds to 7.5×10^7 cells. The added MURF3 RNA has a mol. wt of (1250 nucleotides × 340 daltons=) 425 kd. Thus, the signal of 50 pg (lane 3) corresponds to 7.5×10^7 copies, in a background of RNA from 7.5×10^7 cells: 1 copy/cell. In lane 6 a control hybridization was performed with (double-stranded) MURF3 cDNA as a probe.

per cell would have been detected in this experiment (for calculation, see legend of Figure 2). The products of low molecular weight, visible as a smear in lane 5, do not hybridize to an edited sense oligonucleotide (result not shown). These may be derived from symmetrical transcription of both maxicircle DNA strands, and subsequent processing and degradation.

gRNAs

Recently, Blum *et al.* (1990) discovered short sequences in the *L. tarentolae* maxicircle that, when transcribed, can basepair with edited RNA sequences if G:U basepairs are allowed: guide (g)RNAs. The possible involvement of

gRNAs in *C. fasciculata* editing was investigated by performing a computer analysis of published and newly sequenced maxicircle regions (see Materials and methods). The result of this analysis is shown in Figure 3 where the RNA hybrids are given for seven edited transcript segments (upper lines) and the respective sequences found in this analysis (lower lines). The positions of all the putative gRNA genes in *C. fasciculata* maxicircle DNA are virtually identical to those of the homologous gRNA genes in *L. tarentolae* (see also Figure 7). Surprisingly, two of the putative *C. fasciculata* gRNA sequences do not fully basepair to their respective edited transcript segment. In CYb-II and MURF3(5') gRNAs (nomenclature according to Blum *et al.*, 1990) a C is present

Fig. 3. Identification of *Crithidia fasciculata* gRNA sequences. In the figure, edited mRNA sequences (top) are aligned with the putative gRNA sequences (bottom) as found in a computer search of *C. fasciculata* maxicircle sequence (see Materials and methods for details). Nomenclature as in Blum *et al.* (1990). The 5' end of three gRNAs was determined (see also Figure 6) and indicated in the figure (arrows). Coordinates and references of the gRNA sequences: MURF3(FS): 105–56[a], MURF3(5'): 90–156[b]; MURF2-I: 598–629[c]; MURF2-II: 126–63[a]; COII: 863–823[c]; CYb-I: 90–128[a]; CYb-II: 2081–2142[b]. For comparison, the *L. tarentolae* CYb-II gRNA sequence is given at the bottom of the figure (Blum *et al.*, 1990). Compensatory base changes between the two organisms are indicated by boldface letters. Two C:A mismatches in the CYb-II and MURF3(5') gRNA:mRNA alignments are indicated (▲). Short oligo(U) sequences are underlined. For general location of the gRNA genes see also Figure 7. References: [a] This paper, Figure 9A and B; [b] Sloof *et al.*, 1987; [c] van der Spek *et al.*, 1989.
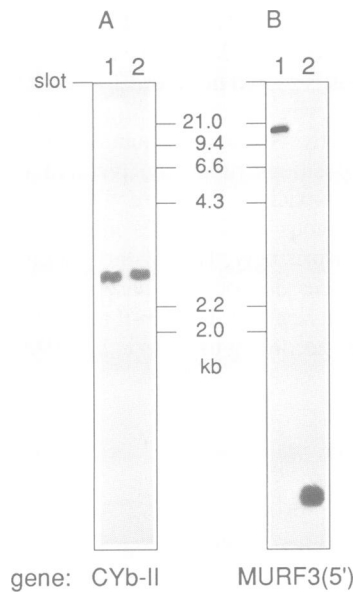
Fig. 4. Southern blot hybridization of kinetoplast DNA with gRNA probes. Southern blots of kinetoplast DNA digested with restriction enzymes *Eco*RI (lanes 1) or *Eco*RI and *Sal*I (lanes 2) were hybridized with oligonucleotides complementary to the CYb-II (panel A, C81) and MURF3-5' (panel B, C80) gRNAs.
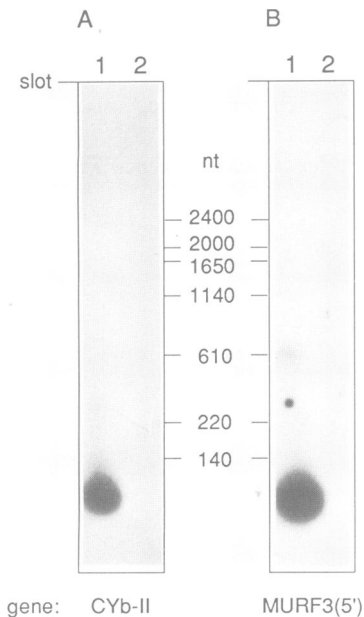
Fig. 5. Northern blot hybridization of total and mtRNA with gRNA probes. Northern blots of mitochondrial (lanes 1) and total RNA (lanes 2) were hybridized with labelled oligonucleotides complementary to the CYb-II (panel A, C81) and MURF3(5') (panel B, C80) gRNAs. The size of the hybridizing transcripts is indicated. Control experiments with probes for MURF3 RNA and the mt ribosomal RNAs of 9S and 12S showed that these RNAs were intact in this mtRNA preparation (results not shown).

opposite an A in the edited transcript (see also Figure 6).

To test the data obtained in the computer analysis experimentally, the following experiments were performed. First, Southern blots of digested kDNA were hybridized with
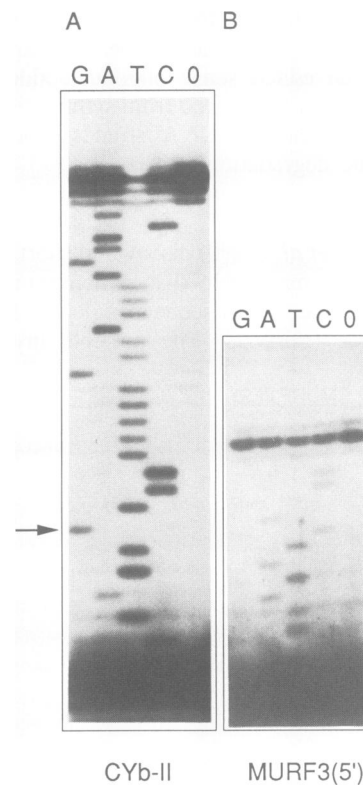
Fig. 6. Sequence analysis by primer extension of gRNAs. Mitochondrial RNA was sequenced using a labelled oligonucleotide complementary to CYb-II (panel **A**, C89) or MURF3(5') (panel **B**, C80) gRNA. The CYb-II primer is located 6 nucleotides 3' of the cytidine that would give rise to an A:C pair in the mRNA:gRNA hybrid. This nucleotide is indicated in the figure by an arrow. the precise 5'-end sequence of the gRNAs could not be read, most likely due to heterogeneity in length at the 5'-end (also observed in some *L.tarentolae* gRNAs, Blum *et al.*, 1990).

oligonucleotides directed against putative gRNA gene sequences. A representative example is shown in Figure 4. The CYb-II- as well as the MURF3(5') primer hybridized only with fragments of a size expected on the basis of the known maxicircle sequence (a 2.7 kb *Eco*RI fragment that is not cut by *Sal*I and an 11 kb *Eco*RI fragment cut by *Sal*I to give an 800 bp *Eco*RI − *Sal*I fragment respectively), indicating that no heterogeneity exists among maxicircles.

Second, Northern blot analysis showed that these two putative gRNA genes are indeed transcribed. In an experiment utilizing the same oligonucleotides (Figure 5), stable transcripts of ~60−80 nucleotides are detected only in mt RNA, analogously to the situation in *L.tarentolae* (Blum *et al.*, 1990).

Third, the sequence of the 5' region of three gRNAs [MURF3(FS), MURF3(5') and CYb-II] was determined with reverse transcriptase and oligonucleotide primers; two of the resulting sequences [of MURF3(5') and CYb-II gRNAs] are shown in Figure 6. The sequence clearly matches that of the corresponding DNA (see Figure 3), confirming the results of Figure 5. As in *L.tarentolae*, the transcripts have discrete 5' ends, arguing against the possibility that they have arisen as non-specific breakdown products. Moreover, the sequence of the CYb-II gRNA clearly shows that the C predicted by the DNA sequence

L. tarentolae
guide RNA genes

Fig. 7. Map localization and conservation of gRNA genes. Linear map of the maxicircle. The position of the genes is indicated. Shaded areas correspond to the transcript sequences that are subject to editing in different trypanosome species. The position of the *L.tarentolae* gRNA genes is indicated by the boxes. Below the figure is indicated which gRNA genes are conserved (+) or not (−) in *C.fasciculata* and *T.brucei*. The nucleotide sequence of the putative *T.brucei* gRNAs is given in Figure 8.

COII

```
mRNA: 5'  CAAUAUCAAGUUUAGGUAUAAAAGUAGA   G A ACCUGGUAGGUGUAAUGAAAUAAU  3'
                                     uu u u
          IIII I II* I*    I I I   *IIIIII*IIIIIII  * *    II    II I*
gRNA: 3'  GUUAGAAUUUUAGGGGAAAGUAAUAUCU  a a a a UGGAAAGAUAUAUUUUAAAUAAUG  5'
                                          U U
```

MURF2-I

```
mRNA: 5'  AGGGGAUUUUAAGAUUGGCUUUGAUUG AGUCGUGUUUUUGAUUUGUUAUGUAUUAGAAC  3'
                                     u
          I* *I    I  *I  *  II  I IIII**I*III II *  IIIIII II I    I
gRNA: 3'  UUAUCCUCUAAAUUCUUAUAAUGUCAC a UUGGUACAUAACUGUAACAAUUCAAACCAAUA  5'
```

MURF2-II

```
mRNA: 5'  CUAUAAUGA    AA G    GG  G    AA   AG    A   UUG GCUUUGAUUGAGUCGUGUUUUU  3'
                   uuuu  u uuu  uu uuuu  uuu  uuuu uuu    u
          I   I*IIIIIII**II***II**I*IIIIIII*IIIIIIIIIIIIIIII   I*I*I  *  II*II
gRNA: 3'  AUUUAAAUU a a a a  U 9U a a 9 UU a a U 9a9a UU a a a UU a a a a U a a a AAC a CGAAAGAUAUUUAAAGGAAGAA  5'
```

Fig. 8. Identification of gRNA genes in *T.brucei* maxicircle DNA. Computer analysis of the *T.brucei* maxicircle (including the variable region) revealed the presence of three putative gRNA genes. The location of these genes is similar to the corresponding gRNA genes in *L.tarentolae* (Blum *et al.*, 1990) and *C.fasciculata* (this paper, see Figure 7). Short oligo(U) sequences are underlined. Coordinates and reference: COII: 2642−2585; MURF2-I: 2348−2407; MURF2-II: 5663−5594; Hensgens *et al.*, 1984.

is indeed present in the gRNA (arrow), confirming the presence of an A:C pair in the mRNA:gRNA hybrid.

Discussion

In this paper we have collected experimental evidence supporting the hypothesis that gRNAs participate in RNA editing in trypanosome mitochondria. As part of this study we have searched for conventional templates for the edited sequences, using more sensitive methods in previous work. The results of Figure 1 conclusively show the absence of edited DNA gene versions in carefully calibrated experiments. The level of detection of the analysis is at least four orders of magnitude below that required for the detection of a single copy gene. Unfortunately, PCR analysis could not be used in the search of (−) strand MURF3 RNA templates, due to the fact that first strand cDNA synthesis occurred in the absence of added (+) strand MURF3 primer. Most likely, the RNA preparations contained a sizeable fraction of small nucleic acid fragments, that can act as (non-specific) primers on endogenous (+) strand MURF3 RNA. Alternatively, one of the MURF3 gRNAs (see below) might act as such. The approach followed in the experiment of Figure 2, nevertheless, still would have allowed the detec-

tion of the presence of one copy per cell of a (−) strand MURF3 RNA. The conclusion that such RNAs do not exist seems therefore justified.

Guide RNA genes in C.fasciculata and T.brucei

Our analysis identified seven putative gRNA genes in maxicircle DNA of *C.fasciculata* (Figure 3). We show that two of these, [MURF3(5′) and CYb-II, see Figures 5 and 6] are transcribed into stable transcripts of a discrete size, comparable to that of *L.tarentolae* gRNAs and we infer that RNAs of a similar size are produced from the other gRNAs in *C.fasciculata*. Southern blot analysis demonstrated that only one version of these putative gRNA genes can be found in mt DNA (Figure 4).

All seven putative gRNA genes are located at positions virtually identical to those of their *L.tarentolae* counterparts (summarized in Figure 7). Beside this, the length of the gRNAs is very similar in the two organisms, as illustrated by the 5′ end sequence of MURF3(FS), MURF3(5′) and CYb-II, which extends only 2−4 nucleotides beyond that of the corresponding *L.tarentolae* gRNAs. Moreover, in cases in which the nucleotide sequence of edited regions *C.fasciculata* transcripts differs from that of the corresponding *L.tarentolae* RNA segments, compensatory base

Table I.

Oligonucleotide primer		Coordinates and reference
C41C	ACCGT<u>A</u>TTT<u>T</u>GAT<u>A</u>GATTAG	3336−3350[a]
C67	ACAGATAACATAAACCAATC	63−45[b]
C80	TATTTATTTTGTCTTTATTGCCTTTTTG	110−130[c]
C81	TTACGTTCTCTCATATTAAACCCCTATT	2088−2115[c]
C87	CGTATAAATATTATTTTTAC	17−36 (This paper, Figure 9B)
C89	CATTACGTTCTCTCATATTAAACC	2086−2109[c]

[a]Edited sequence (nucleotides not present in the DNA are underlined) of the MURF3 frameshift region, Sloof *et al.*, 1987; [b]number of nucleotides downstream of the A in ATG of COI, van der Spek *et al.*, 1990; [c]Sloof *et al.*, 1987.

changes in the putative *C.fasciculata* gRNA genes maintain a full potential for basepairing in the gRNA:mRNA hybrids (illustrated by the two CYb-II gRNA:mRNA pairs in Figure 3). Such a strict conservation, both in terms of sequence and gene location, provides strong evolutionary evidence in support of a role in RNA editing for the gRNAs. As in *L.tarentolae*, a maxicircle COIII gRNA gene has not been identified in *C.fasciculata*. The likely possibility that this gRNA is encoded by minicircle DNA (as in *L.tarentolae*), is under investigation.

Little is known about the sequence and location of gRNA genes in *T.brucei*. Our analysis of *T.brucei* maxicircle sequences suggests the presence of three gRNA genes: COII, MURF2-I and MURF2-II (Figure 8). Although we have not further substantiated the computer analysis with Northern and Southern blot experiments, they appear to be bona fide gRNA genes, since their location on the *T.brucei* maxicircle corresponds precisely to that of their *C.fasciculata* and *L.tarentolae* counterparts. The COII and MURF2 transcripts are only edited in a small section and the pattern and extent of editing of these RNAs does not differ much between the three organisms. Remarkably, CYb-I and CYb-II gRNA genes are not found in the *T.brucei* maxicircle, in spite of the fact that the CYb transcript is also edited in a very similar fashion in the three species. Recently, several putative gRNA genes encoding gRNAs for the extensively edited COIII, MURF3 and ATPase 6 transcripts, have been localized to minicircle DNA in *T.brucei* (Bhat *et al.*, 1990; Koslowsky *et al.*, 1990; Pollard *et al.*, 1990). It is to be expected that the CYb and many of the remaining gRNA genes will also be located in minicircles, providing an explanation for the greater minicircle sequence complexity in *T.brucei*, in comparison with *L.tarentolae* and *C.fasciculata*, (also suggested by Bhat *et al.*, 1990; Koslowsky *et al.*, 1990; Sturm and Simpson, 1990b).

The mode of action of gRNAs

According to the model of Blum *et al.* (1990), editing is initiated by duplex formation between the 5'-region of a gRNA (the so-called anchor sequence) and a stretch of nucleotides downstream of an editing site of the pre-edited RNA. Besides A:U and G:C basepairs the duplex may also contain less conventional G:U pairs. In *C.fasciculata* gRNAs, possible anchor sequences are also present and their lengths are very similar to those of their *L.tarentolae* counterparts, except that of *C.fasciculata* MURF2-I gRNA, which is considerably smaller (5 instead of 11 nucleotides). Short anchor sequences are not unusual, however, e.g. those of COII in *L.tarentolae* and *C.fasciculata*, which are also 5 nucleotides. Apparently, the editing machinery is able to cope with short anchors, suggesting the involvement of proteins
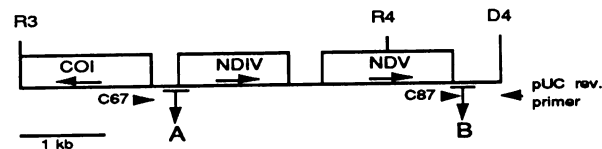
A.

```
        10          20          30          40          50
TCTCTCCCTA  TTGCATTTTT  ATTACTATGT  TTAAAAGTTA  AAATGTTTTC

        60          70          80          90          100
TCTCACATTT  AATTCAACTC  AGTCAAAGCA  TAGAAATGGG  GTTAAATTAA

        110         120         130         140
AGTGATTAAA  CATGTTTAAA  TTTTTTTGAT  ATCGGAAATT
```

B.

```
        10          20          30              40          50
TTTTAATATT  AATATACGTA  TAAATATTAT  TTTTACTAAA  ATATTTTAGC
                       C87 →
        60          70          80          90          100
TTGTTCATAA  CAGCATTAGT  CTAATCTATC  AGAATGCTAC  TCAAAAATTT

        110         120         130         140         150
ATATTATCTT  CTAACGTCAA  GGTTTTTGCC  TTTTACTATT  AAATCGATGA

        160         170
TATCCATAAT  AACGCTAATG
```

Fig. 9. Cloning and sequencing strategy. The figure shows a schematic map of a part of the *C.fasciculata* maxicircle. The relevant restriction sites are indicated (R = *Eco*RI, D = *Hind*III). The sequence, obtained with primers C67, C87 and the standard pUC reverse primer, is given in **A** and **B** respectively. The start codon of NDIV (see A) and the stop codon of NDV (see B) are in boldface letters. Position and coordinates of gRNA sequences: A. MURF2-II: 126−63; B. MURF3(FS): 105−56; CYb-I: 90−128.

during this stage. In a next phase, the editing machinery is thought to move from 3' to 5' on the pre-edited RNA and, when a mismatch is encountered, Us are inserted or deleted depending on the mismatch in question. This process can be repeated at upstream editing sites until basepairing between gRNA and mRNA is maximal, again including G:U pairs. In this model the insertion/deletion process is 'guided' by the gRNA, for details see Blum *et al.* (1990).

A different view is held by Decker and Sollner-Webb (1990), who assume that uridine insertion/deletion processes occur randomly within a certain section of an RNA, the correctly edited sequence being protected from further editing by duplex formation with the gRNA. In both models, however, a mismatch between the mRNA and the gRNA is a signal that further editing is required. Surprisingly, therefore, our data imply that two putative gRNA:mRNA hybrids [MURF3(5') and CYb-II] contain a C:A non-Watson−Crick basepair, the C being present in the gRNA. The sequence of the 5' moiety of CYb-II gRNA, including the C (Figure 6), completely matches that of the correspond-

ing area of the gene, which eliminates the possibility of alternative (U-encoding) full length CYb-II gene versions. Although we cannot formally exclude the possibility that additional, smaller gRNAs [CYb-III, MURF3(5')-II, ?] exist, we consider this unlikely in view of the fact that the remainder of the gRNAs, 3' of the C:A basepair, can form a perfect duplex of considerable length with the respective edited RNAs (e.g. 29 bp for CYb-II gRNA). The presence of these C:A pairs is, therefore, difficult to explain. The possibility arises that they are not seen as a mismatch by the editing machinery. It has indeed been suggested that the presence of a C:A pair in DNA−oligonucleotide duplexes and in tRNA hairpins gives only little distortion of the double helix (Hunter *et al.*, 1986; Wyatt *et al.*, 1989). Even if these observations can be extended to gRNA:mRNA duplexes, it would be hard to envisage how the editing machinery would be able to skip C:A pairs, without also missing the reverse A:C pairs, unless one of the two (g- or mRNA) is marked by some form of base modification. A:C pairs, with the A in the gRNA, should be formed occasionally during editing of a number of RNAs in *T.brucei* (e.g. twice in COIII RNA, Feagin *et al.*, 1988), and would result in U insertion into the pre-edited RNA. Another scenario, along the lines of the model of Blum *et al.* (1990), could therefore be that the C:A pairs are seen as a mismatch and that their presence would result, to some extent, in futile cleavage and ligation cycles at the site of the mismatch. As long as the editing machinery would be capable of proceeding beyond the C:A pair with reasonable efficiency, however, its presence in limited numbers should not be too deleterious to the editing process, since a C:A pair is of no consequence for the sequence of the edited RNA.

Blum and Simpson (1990) further noted the presence of a short oligo(U) sequence at the 3' end of most gRNAs of *L.tarentolae*, which is reminiscent of transcription termination signals for several RNA polymerases (Bogenhagen and Brown, 1981) and a tetrameric AAUA sequence just upstream of it. In *C.fasciculata* these sequence motifs are much less clearly present, or even absent, as outlined in Figure 3. The relevance of these motifs for gRNA functioning is obscure whereas other common sequences are not apparent in the *C.fasciculata*, *L.tarentolae* and *T.brucei* gRNA (gene) sequences. The identification of the primary, secondary and tertiary structural elements involved in gRNA functioning, therefore, awaits the development of properly working *in vitro* RNA editing assay systems.

Materials and methods

Oligonucleotide primers
The primers used are shown in Table I.

Cloning of maxicircle regions containing gRNA genes and sequencing
A schematic representation of the procedures is given in Figure 9. Restriction sites and location of the primers are indicated. A *C.fasciculata* kinetoplast (k)DNA library was constructed by digesting kDNA [isolated according to Borst *et al.*, (1979)] with restriction enzymes *Hin*dIII and/or *Eco*RI followed by ligation into the plasmid vector pUC19, cut by the appropriate enzyme(s). The sequence of the gRNA gene for MURF2-II was obtained by sequencing of the R3R4 clone with primer C67. The sequence is given in Figure 9A. This primer was originally used for COI RNA sequencing (van der Spek *et al.*, 1990).

The clone R4D4 containing the gRNA genes for MURF3(FS) and CYb-I (nomenclature according to Blum *et al.*, 1990) was isolated by screening the library with NDV cDNA. Areas containing the MURF3(FS)- and CYb-I

gRNA genes were sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977) using the primers C87 and the standard pUC 'reverse' primer. The sequence is given in Figure 9B. The other gRNA genes are located on fragments for which the sequence has been published before (Sloof *et al.*, 1987; van der Spek *et al.*, 1989).

Computer analysis
Sequence comparisons were made on a VAX computer, using programs that were designed by Mr B.F.de Vries in our laboratory.

Nucleic acid isolation
Total RNA and DNA were isolated as described in van der Spek *et al.* (1990). For RNA isolation, DNA was removed by DNase I digestion (5 μg/ml, 15 min, 37°C) according to the proteinase K/Ca^{2+} protection method (Tullis and Rubin, 1980). Mitochondrial RNA was extracted from a Renografin fraction enriched in mitochondrial vesicles that was obtained by the method described by Birkenmeyer and Ray (1986). The enrichment for mitochondrial transcripts was tested by Northern blot hybridization of total and mtRNA and amounts reproducibly to ~50 times.

Electrophoresis, blotting and hybridization
Procedures were essentially as described (Sambrook *et al.*, 1989; van der Spek *et al.*, 1990) except for the preparation of Northern blots. Glyoxylated RNA was run and transferred to Hybond filters as described. To ensure that small RNAs are retained on the filter upon hybridization, an extra UV cross-linking step as described by Blum *et al.* (1990) was included (120 mJ/cm^2).

Primers were kinased using T4 polynucleotide kinase and [γ-^{32}P]ATP and used in RNA sequence reactions or in hybridizations as described (Sambrook *et al.*, 1989). Hybridization was performed overnight in a small volume (10 ng/ml, 5 ml) followed by two washes of 5 min each at 60°C (6 × SSC, 0.1% SDS).

PCR analysis
After a denaturation step of 5 min at 95°C, DNA was amplified by 30 cycles of 1.5 min at 95°C, 2 min at 48°C and 2.5 min at 72°C in a Hybaid temperature controller. The program was ended with a general extension step of 7.5 min at 72°C. For all reactions a Perkin−Elmer Cetus PCR kit and enzymes (AmpliTaq) were used according to the manufacturer's instructions.

References

Abraham,J.M., Feagin,J.E. and Stuart,K. (1988) *Cell*, **55**, 267−272.
Benne,R. (1985) *Trends Genet.* **1**, 117−121.
Benne,R. (1989) *Biochim. Biophys. Acta*, **1007**, 131−139.
Benne,R. (1990) *Trends Genet.* **6**, 177−181.
Benne,R., van den Burg,J., Brakenhoff,J.P.J., Sloof,P., van Boom,J.H. and Tromp,M.C. (1986) *Cell*, **46**, 819−826.
Bhat,G.J., Koslowsky,D.J., Feagin,J.E., Smiley,B.L. and Stuart,K. (1990) *Cell*, **61**, 885−894.
Birkenmeyer,L. and Ray,D.S. (1986) *J. Biol. Chem.*, **261**, 2363−2368.
Blum,B. and Simpson,L. (1990) *Cell*, **62**, 391−397.
Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189−198.
Bogenhagen,D.F. and Brown,D.D. (1981) *Cell*, **24**, 261−270.
Borst,P. and Hoeijmakers,J.H.J. (1979) *Plasmid*, **2**, 20−40.
Cattaneo,R., Kaelin,K., Baczko,K. and Billeter,M.A. (1989) *Cell*, **56**, 759−764.
Chen,S.H., Habib,G., Yang,C.Y., Gu,Z.W., Lee,B.R., Weng,S.A., Silberman,S.R., Cai,S.J., Deslypere,J.P., Rosseneu,M., Gotto,A.M., Li,W.H. and Chan,L. (1987) *Science*, **238**, 363−366.
Covello,P.S. and Gray,M.W. (1989) *Nature*, **341**, 662−666.
Decker,C.J. and Sollner-Webb,B. (1990) *Cell*, **61**, 1001−1011.
Feagin,J.E., Jasmer,D.J. and Stuart,K. (1987) *Cell*, **49**, 337−345.

Feagin,J.E., Abraham,J.M. and Stuart,K. (1988) *Cell*, **53**, 413–422.

Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.H. and Grienenberger,J.M. (1989) *Nature*, **341**, 660–666.

Gualberto,J.M., Weil,J.H. and Grienenberger,J.M. (1990) *Nucleic Acids Res*, **18**, 3771–3776.

Harris,M.E., Moore,D.R., Hajduk,S.L. (1990) *J. Biol. Chem.*, **265**, 11368–11376.

Hensgens,L.A.M., Brakenhoff,J., de Vries,B.F., Sloof,P., Tromp,M.C., van Boom,J.H. and Benne,R. (1984) *Nucleic Acids Res.*, **12**, 7327–7344.

Hunter,W.N., Brown,T., Anand,N.N. and Kennard,O. (1986) *Nature*, **320**, 552–555.

Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L. Feagin,J.E. and Stuart,K. (1990) *Cell* **62**, 901–911.

Mahendran,R., Spottswood,M.R. and Miller,D.L. (1991) *Nature*, **349**, 434–436.

Maizels,N.M. and Weiner,A. (1988) *Nature*, **334**, 469–470.

Pollard,V.W., Rohrer,S.P., Michelotti,E.F., Hancock,K. and Hajduk,S.L. (1990) *Cell*, **63**, 783–790.

Powell,L.M., Wallis,S.C., Pease,R.J., Edwards,Y.H., Knott,T.J. and Scott,J. (1987) *Cell*, **50**, 831–840.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual*, second edition, Cold Spring Harbor Laboratory Press, Cold Spring, Harbor, NY.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schuster,W., Wissinger,B., Unseld,M. and Brennicke,A. (1990) *EMBO J.*, **9**, 263–269.

Simpson,L. (1986) *Int. Rev. Cytol.*, **99**, 119–179.

Simpson,L. and Shaw,J.M. (1989) *Cell*, **57**, 355–366.

Shaw,J.M., Feagin,J.E., Stuart,K. and Simpson,L. (1988) *Cell*, **53**, 401–411.

Sloof,P., van den Burg,J., Voogd,A. and Benne,R. (1987) *Nucleic Acids Res*, **15**, 51–65.

Stuart,K. (1989) *Exp. Parasitol.*, **68**, 486–490.

Sturm,N.R., Simpson,L. (1990a) *Cell*, **61**, 871–878.

Sturm,N.R., Simpson,L. (1990b) *Cell*, **61**, 879–884.

Thomas,S.M., Lamb,R.A. and Paterson,R.G. (1988) *Cell*, **54**, 891–902.

Tullis,R.H. and Rubin,H. (1980) *Anal. Biochem.*, **107**, 260–264.

van der Spek,H., van den Burg,J., Croiset,A., van den Broek,M., Sloof,P. and Benne,R. (1988) *EMBO J.*, **7**, 2509–2514.

van der Spek,H., Arts,G.J., van den Burg,J., Sloof,P. and Benne,R. (1989) *Nucleic Acids Res.*, **17**, 4876.

van der Spek,H., Speijer,D., Arts,G.J., van den Burg,J., van Steeg,H., Sloof,P. and Benne,R. (1990) *EMBO J.*, **9**, 257–262.

Vidal,S., Curran,J., Kolakofsky,D. (1990) *EMBO J.*, **9**, 2017–2022.

Volloch,V., Schweitzer,B. and Rits,S. (1990) *Nature*, **343**, 482–484.

Wyatt,J.R., Puglisi,J.D. and Tinoco Jr,I. (1989) *BioEssays*, **11**, 100–106.