The EMBO Journal vol.7 no.8 pp.2509 – 2514, 1988

Transcripts from the frameshifted *MURF3* gene from *Crithidia fasciculata* are edited by U insertion at multiple sites

Hans van der Spek, Janny van den Burg, Alexander Croiset, Marianne van den Broek, Paul Sloof and Rob Benne

Laboratory of Biochemistry, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Communicated by P.Borst

In trypanosome mitochondria an RNA editing process is operative, which co- or post-transcriptionally alters the nucleotide sequence of transcripts by insertion and/or deletion of U residues at specific sites. To increase our understanding of the mechanism of this process we have compared the nucleotide sequence of the frameshifted mitochondrial *MURF3* gene from *Crithidia fasciculata* to that of a large number of *MURF3* cDNAs. We found cDNAs derived from transcripts edited at two different sites in the protein coding sequence: (i) at the frameshift position five extra U residues connect the two reading frames and (ii) at the 5′ terminus 22 inserted Us shift a putative initiator codon out of phase. The collection also contained cDNAs that were derived from non-edited transcripts. Partially edited sequences were not found, except in one cDNA, which contained an edited frameshift site in combination with a non-edited 5′ terminus. The analysis further showed that *MURF3* transcripts have a 3′-terminal poly(AU) extension, which varies in sequence. The implications of these results are discussed.

Key words: mitochondrion/gene expression/trypanosomes/RNA editing

Introduction

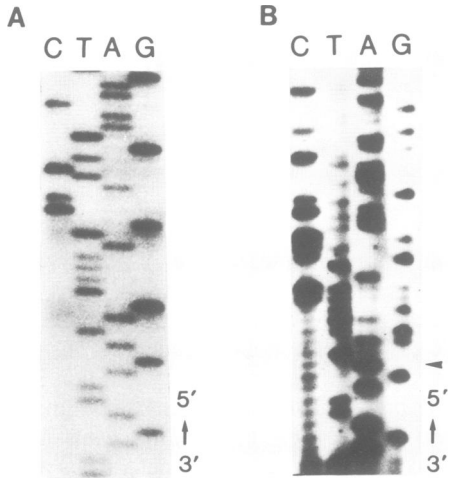
The mitochondrial (mt) DNA of trypanosomes consists of a large network of two types of circle: 10^4 mini-circles (1–3 kb, depending on the species) and 50–100 maxi-circles (20–40 kb) (reviewed in Borst and Hoeijmakers, 1979; Englund, 1981; Stuart, 1983). We are studying the organization and expression of genes residing in the maxi-circle component of *Trypanosoma brucei* and *Crithidia fasciculata* (Benne *et al.*, 1983; Hensgens *et al.*, 1984; De Vries *et al.*, 1988; Sloof *et al.*, 1985, 1987). The results obtained by us and other groups [reviewed in Benne (1985) and Simpson (1986)] clearly demonstrate that the maxi-circle is the trypanosome equivalent of mtDNA in other organisms, since 'classical' mt genes were identified. However, trypanosome mtDNAs also contain a relatively large number of genes not found in other organisms. Some of these genes are species-specific (Simpson *et al.*, 1987) and encode highly unusual amino acid sequences. Conversely, some universal mt genes appeared to be missing in trypanosomes, such as mt tRNA genes (all trypanosomes) and the gene for cytochrome oxidase (*cox*) subunit III (*T.brucei*).

The most remarkable feature of the trypanosome mt genetic system is the unconventional way in which the nucleotide sequence of transcripts at specific sites differs from the respective genomic sequence. The first example of this phenomenon was provided by four non-DNA-encoded U residues present in the major transcript of the frameshifted *coxII* gene in *T.brucei* and *C.fasciculata* (Benne *et al.*, 1986), which give the transcript a continuous reading frame. Since alternative gene versions encoding these extra nucleotides could not be found, we inferred that they are the result of a novel RNA editing process, which co- or post-transcriptionally inserts U residues at the frameshift position of the *coxII* transcript. Since then, numerous other examples of U insertion into trypanosome mt transcripts have been described. Thirty four U residues are inserted into the 5′-terminal region of the apocytochrome *b* (cyt *b*) transcript from *T.brucei* (Feagin *et al.*, 1987) and 39 extra Us appear in the cyt *b* transcript from *C.fasciculata* and *Leishmania tarentolae* (Feagin *et al.*, 1988a). The additions create in-frame AUG codons in transcripts that lack initiator triplets in the corresponding genomic sequence.

For other trypanosome mt genes similar observations were made (reviewed by Shaw *et al.*, 1988), 5′-terminal editing in general being limited to the genes without genomic AUG codons, such as the *coxIII* and *MURF2* genes. From these studies no obvious clues to a possible mechanism could be extracted, since virtually each case of RNA editing is unique in terms of number and pattern of Us inserted, without any sequence similarity in surrounding areas. Moreover, not only U insertion but also U deletion appears to occur (Benne *et al.*, 1986; Shaw *et al.*, 1988). The most spectacular example of RNA editing was recently described by Feagin and coworkers (Feagin *et al.*, 1988b), who showed that a *T.brucei* mt transcript is extensively edited by U insertion and deletion, presumably over its entire length, resulting in a nucleotide sequence with a high degree of identity to the *coxIII* gene from *C.fasciculata* and *L.tarentolae*, although the corresponding mtDNA sequences lack any obvious similarity.

RNA editing in *T.brucei* is developmentally regulated and restricted for most RNAs to the procyclic life-cycle stage (= insect and culture form) when *T.brucei* contains a fully functional mitochondrion (Feagin and Stuart, 1988). This implies that the RNA present in bloodstream form *T.brucei* (that lack a mt respiratory system) is not translatable and that RNA editing plays a crucial role in regulating the expression of mt genes.

In most of the work mentioned above, transcript sequences are obtained by primer-extension analysis with total cellular or mtRNA. They represent only a very small area of the respective genes. Moreover, this approach does not always yield readable sequence ladders, due to low abundance of template and/or the simultaneous presence of differentially edited forms of a transcript. For this reason, intermediates of the editing process (if they exist) go undetected.



C

```
. K T V E Q C L P     M K I R L C *
 . K Q L N N V Y R     W R L D Y V S .   PROTEIN
                1
..AAAACAGTTGAACAATGTCTACCG A   TGA AGATTAGATTATGTTAGT..   DNA
..AAAACAGUUGAACAAUGUCUACCGUAUUUUGAUAGAUUAGAUUAUGUUAGU..   RNA Ed.
                2
. K T V E Q C L P Y F D R L D Y V S .   PROTEIN Ed.
```

Fig. 1. Sequence of *MURF3* transcripts at the frameshift position. (**A**) Sequence of an edited cDNA clone. (**B**) Sequence obtained from cellular poly(A)$^+$ RNA (see Materials and methods). The arrow indicates the ultimate nucleotide 3′ of the frameshift. (A,B) Five U residues not encoded in the *MURF3* genomic sequences are indicated by a dot. (**C**) Alignment of edited transcript (RNA$^{Ed.}$) and genomic nucleotide sequences of the frameshift region (nt 3316–3361 in Sloof *et al.*, 1987). Translation into amino acids of the genomic nucleotide sequence is given in the two separate reading frames in which homology to the *M.polymorpha* chloroplast ORF392 is found. Translation into amino acids of the edited sequence is also given (PROTEIN$^{Ed.}$). Amino acids that are identical in *M.polymorpha* chloroplast ORF392 are underlined. * = stop codon. Sequences complementary to the DNA oligonucleotide (1) and RNA oligonucleotide (2), are indicated.

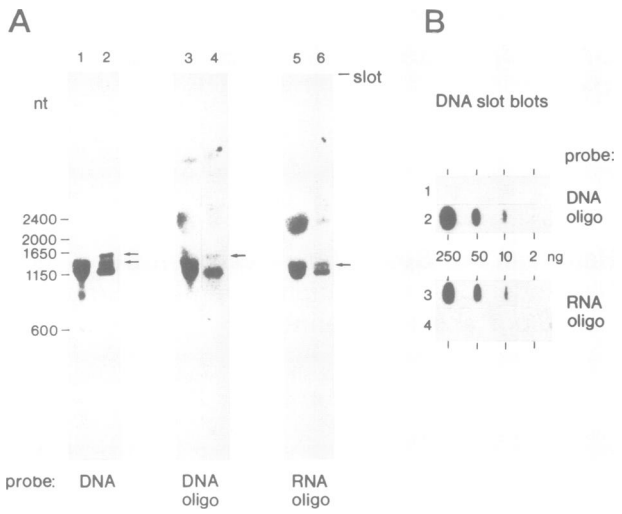


Fig. 2. Northern blot analysis of *MURF3* transcripts. (**A**) 50 μg of total cellular RNA (**lanes 1, 3** and **5**) or 5 μg of poly(A)$^+$ RNA (**lanes 2, 4** and **6**) was electrophoresed and blotted as described by Benne *et al.* (1983). Blots were hybridized with a maxi-circle fragment containing the 5′-moiety of the *MURF3* gene (nt 2866–3458, Sloof *et al.*, 1987), lanes 1 and 2; the end-labelled DNA oligonucleotide (lanes 3 and 4) and the RNA oligonucleotide (lanes 5 and 6), respectively. The conditions of the hybridization and the sequences of the oligonucleotides are given in Materials and methods (see also Figure 1). The arrows in the figure point at transcripts of 1250 and 1400 nt mentioned in the text. (**B**) Slotblots of *MURF3* genomic DNA (clone described in legend to Figure 1) **panels 2** and **4**, and *MURF3* cDNA, **panels 1** and **3**, hybridized to the DNA oligonucleotide (panels 1 and 2) and the RNA oligonucleotide (panels 3 and 4). The slots contain the amounts of DNA indicated.

In a more systematic analysis that aims at detecting sites of editing over the entire length of transcripts we are determining the nucleotide sequence of *C.fasciculata* mt cDNAs. For this purpose, a cDNA library was constructed which contains a large number of mt cDNA clones.

In this paper we report the results of such an analysis for the frameshifted *MURF3* gene from *C.fasciculata*. This gene (previously referred to as the *CURF2/1* gene, Sloof *et al.*, 1987; for the new nomenclature see Shaw *et al.*, 1988; Simpson *et al.*, 1987) is localized just downstream of the gene for the small subunit rRNA and is homologous to a chloroplast gene from *Marchantia polymorpha* (ORF392, Ohyama *et al.*, 1986). The analysis shows that the *MURF3* transcript can be edited at three different sites. (i) At the frameshift site five inserted U residues provide a continuous reading frame; (ii) at the 5′ terminus 22 extra U residues disconnect a putative AUG initiator codon from the rest of the gene; and (iii) in the 3′-terminal extension large runs of Us interrupt a poly(A) tail. No further editing was found.

Results

MURF3 transcripts contain five inserted U residues at the frameshift position

Two overlapping reading frames in the *C.fasciculata* maxi-circle show similarity at the amino acid level (27% overall) to a single *M.polymorpha* chloroplast gene (ORF392, Ohyama *et al.*, 1986). It is likely, therefore, that they represent a single gene and that, by analogy to the frame-shifted *coxII* gene (Benne *et al.*, 1986), a continuous reading frame is created by editing of the transcript.

Therefore, the nucleotide sequence of the frameshift area of *MURF3* transcripts was determined, both from cDNA clones and by direct sequence analysis of *MURF3* RNA, utilizing an oligonucleotide primer complementary to a region ~30 nt downstream of the frameshift. The results are shown in Figure 1A and B, respectively. Of the 11 cDNAs analysed, eight contain the sequence shown in Figure 1A, in which five U residues are present, that are not encoded in the genomic sequence and appear to be inserted. These Us correct the +1 frameshift present in genomic DNA (Figure 1C). A substantial fraction of the cDNAs (three), however, still contain the non-edited sequence (see Table I); intermediary forms (with less than five Us) were not found. Close inspection of the sequences obtained with cellular poly(A)$^+$ RNA (Figure 1B) indeed revealed that they are derived from a mixed population of transcripts. The sequences are clear up to the frameshift site (see the arrow in the figure), beyond it the edited and non-edited sequences appear to be superimposed.

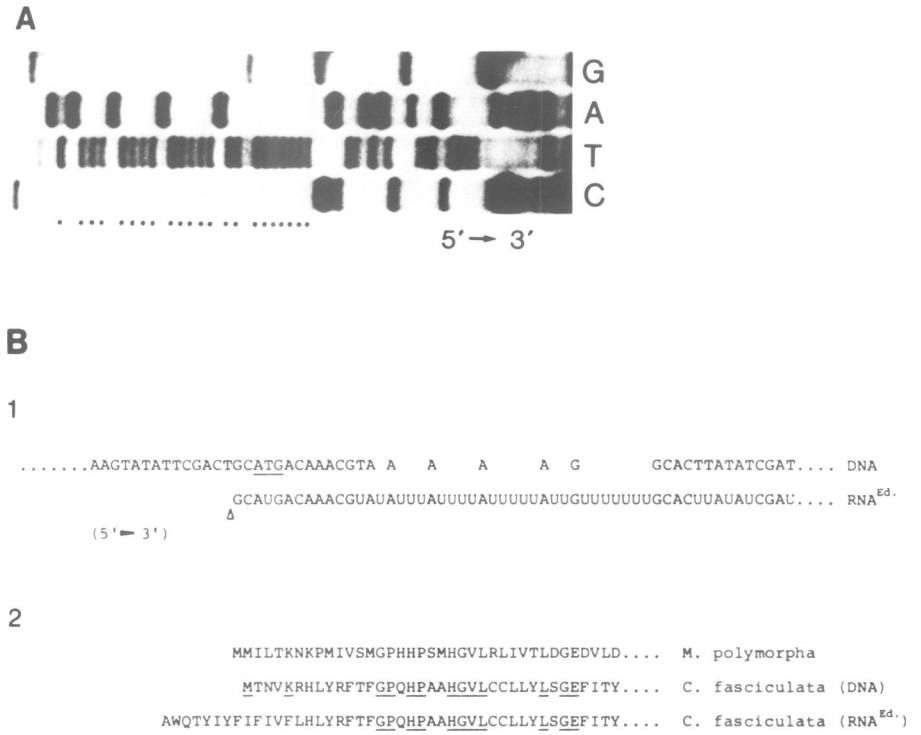


Fig. 3. Nucleotide sequence at the 5′ terminus of edited *MURF3* cDNA. (**A**) Sequence analysis of an edited cDNA. The position of 22 U residues not encoded in genomic DNA, is indicated by a dot. (**B1**) Alignment of 5′-terminal sequences of genomic DNA and the cDNA of panel A. The position at which three of the cDNAs end, is indicated, △. (**B2**) The amino acid sequence of *M.polymorpha* chloroplast ORF392 is aligned with that derived from the edited ($RNA^{Ed.}$) and non-edited (DNA) *MURF3* sequence. Identical amino acids are underlined.

Hybridization analysis with RNA blots and oligonucleotides with the sequence corresponding to either the genomic or the edited sequences ('DNA'-oligo and 'RNA'-oligo, respectively, as indicated in Figure 1C) confirmed the sequence results. The major *MURF3* transcript of 1150 nt (Figure 2A, lanes 1 and 2; see also Sloof *et al.*, 1987) hybridizes to both the DNA and the RNA oligonucleotide (lanes 3,4 and 5,6, respectively). This suggests that this band, in fact, contains both edited and non-edited transcripts of similar length. Other transcripts of ~1250 and 1400 nt also hybridize to a *MURF3* DNA fragment (arrows in lane 2). The product of 1250 nt hybridizes to the RNA oligo (lane 6) and not to the DNA oligo (lane 4), whereas the 1400 nt product shows the opposite pattern. This indicates that the 1250 nt product is edited at the frameshift position and that the 1400 nt product is not. At present, we do not know what the nature of these larger transcripts is. The total amount of edited and non-edited *MURF3* transcripts seems to be approximately equal, which implies that the non-edited form is somewhat under-represented in the cDNA library.

Some high mol. wt products give a signal with the oligonucleotides as a probe, but not when a *MURF3* DNA fragment is used. These bands are therefore most likely due to spurious hybridization caused by the large amounts (up to 40 μg) of the cytoplasmic rRNAs that are present in this area of the blot. Indeed, these signals are mainly observed with total cellular RNA and not with poly(A)$^+$ RNA (compare odd and even lanes).

Under the conditions of the experiment, the oligonucleotides hybridize specifically with up to 0.25 μg cloned DNA on slotblots. As shown in Figure 2B the DNA oligonucleotide only hybridizes to genomic DNA, the RNA oligonucleotide only to the cDNA.

Table I. Editing of *MURF3* cDNAs

Number of cDNAs	Editing at		
	5′-terminus	frameshift site	3′-terminal extension
7	+	+	+[a]
1	–	+	+
3	–	–	ND

ND, not determined.
[a]Determined for five of the seven clones.

Editing at the 5′ and 3′ termini of MURF3 transcripts

The sequence analysis was extended to those regions of the *MURF3* cDNAs that correspond to the 5′ and 3′ part of the transcripts. At the 5′ end 7 out of 11 cDNAs displayed extensive editing, a representative example of which is shown in Figure 3A. Twenty two U residues are inserted into the transcript that are not encoded in the genome, as outlined in Figure 3B1. cDNAs derived from transcripts with non-edited 5′ termini were also found (Table I). Remarkably, one of these possessed an edited frameshift site. Partially edited sequences did not occur: all edited cDNAs that extend upstream of the site of editing contain 22 U residues (see also Discussion). When present, the 22 Us disconnect a genomically encoded AUG codon, located at the expected position with respect to the *M.polymorpha* ORF392 amino acid sequence, from the remainder of the *MURF3* reading frame (Figure 3B2, see also Discussion).

3′-terminal cDNA sequences were also obtained and were found to be derived from transcripts with a 3′-terminal poly(AU) tail. A poly(AU) extension was also reported to

$U_{10}A_n$ (MURF3; this paper)

$U_{15}A_n$ (,,)

$U_{15}AUA_n$ (,,)

$U_7A_{12}U_4A_6U_3A_n$ (TURF2; Benne et al., 1986)

(5'→ 3') n = 8 - 18

Fig. 4. Nucleotide sequence of poly(AU) tails of *C.fasciculata MURF3* and *T.brucei TURF2* transcripts.

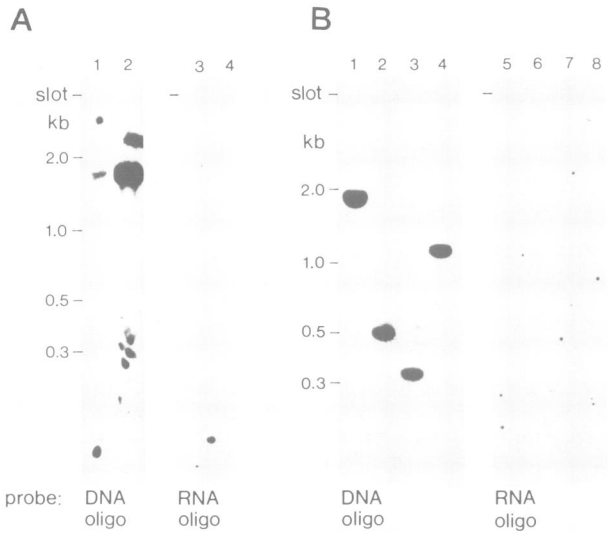


Fig. 5. Southern blot analysis of total *C.fasciculata* DNA. (**A**) *C.fasciculata* total cellular DNA was restricted with *Hin*fI and 5.0 μg (**lanes 2** and **4**) or 0.1 μg (**lanes 1** and **3**) was applied in slots of a 0.7% agarose gel in 0.5% low melting agarose. After solidification, the gels were run and blotted with the inclusion of a 20 min soaking step in 0.25 M HCl. This procedure was followed in order to make sure that network remnants consisting of mini-circles that have no *Hin*fI site, are also blotted onto the filter. (This was shown to be the case by hybridization with a mini-circle probe, not shown.) The blots were hybridized with the DNA oligo (lanes 1 and 2) and RNA oligo (lanes 3 and 4), respectively. (**B**) 5.0 μg *C.fasciculata* total cellular DNA was restricted with *Hin*fI (**lanes 1** and **5**), *Mbo*II (**lanes 2** and **6**), *Alu*I (**lanes 3** and **7**) and *Acc*I (**lanes 4** and **8**). The procedure described in A was followed. Hybridization was with the DNA oligo (lanes 1–4) and the RNA oligo (lanes 5–8). (A and B) Under the conditions of the experiment the oligonucleotides hybridize specifically to cloned DNA on slotblots, as shown in Figure 2B.

occur on a transcript of the *TURF2* gene (Benne *et al.*, 1986) now identified as the *coxIII* gene (Feagin *et al.*, 1988b). The tails not only differ between transcripts of the *TURF2* and *MURF3* genes, but also between *MURF3* transcripts, for which three types of tail were observed (as outlined in Figure 4).

Outside the regions of the *MURF3* transcript discussed so far, no further differences between cDNA and genomic sequences were detected in the one (edited) cDNA that was completely sequenced.

Edited versions of the MURF3 gene are absent

To rigorously exclude the presence of second gene versions that could encode the edited transcript-segments (Benne *et al.*, 1986; Feagin *et al.*, 1987,1988b), we have utilized the DNA and RNA oligonucleotides of Figure 2 as probes on blots of digests of large amounts of total cellular DNA from *C.fasciculata*. As shown in Figure 5 (Figure 5A, lane 4 and Figure 5B, lanes 5–8) no signal is produced on these blots with the RNA oligonucleotides under conditions that allow strong signals with the DNA oligonucleotide at the expected positions with a number of different restriction enzymes (Figure 5A, lane 2 and Figure 5B, lanes 1–4). The intensity of the signal derived from one single maxi-circle is visualized in Figure 5A, lane 1, which contains 2% of the amount of DNA applied in the other lanes (the *C.fasciculata* network contains ~50 maxi-circles). Care was taken also to blot high-mol.-wt DNA, such as network segments without sites for the respective restriction enzyme (see Figure 5 legend). The conclusion from this experiment is, therefore, that alternative *MURF3* gene versions are absent from *C.fasciculata* DNA.

Discussion

In this paper, we report the results of sequence analysis of a large number of cDNAs derived from the frameshifted *MURF3* gene. The analysis shows that *MURF3* transcripts can be altered by U insertion at three different sites: the 5′ terminus, the frameshift area and the 3′-terminal extension.

Gene versions that encode the extra Us at the frameshift region appear to be absent. In view of the now overwhelming evidence for the absence of alternative versions of trypanosome mitochondrial genes (this paper; Benne *et al.*, 1986; Feagin *et al.*, 1987, 1988b) it should be concluded that the alterations in sequence are the result of a process of U insertion (and U deletion) at specific sites of the transcripts of the maxi-circle genes that we know from sequence and transcript analysis (see Benne, 1985 and Simpson, 1986). Until further knowledge of the mechanism of this process is acquired, the term 'RNA editing' appears appropriate (Benne *et al.*, 1986).

Apart from the involvement of U residues, very few characteristics are shared in the different examples of RNA editing. The number of U residues involved and the patterns of insertion/deletion appear to be different in each case. It varies from the insertion of four Us at a gene-internal position of the *coxII* transcript (Benne *et al.*, 1986) via forms of 5′-terminal editing of intermediate complexity (*MURF3*, this paper, Shaw *et al.*, 1988; apocytochrome *b* and *MURF2*, Feagin *et al.*, 1987, 1988a; Feagin and Stuart, 1988; *L.tarentolae* and *C.fasciculata coxIII*, Shaw *et al.*, 1988) to the spectacular example of the *T.brucei coxIII* transcript, which is edited (presumably) over its entire length at >128 different sites (Feagin *et al.*, 1988b). In the latter case the paradoxical situation arises of a transcript that does not hybridize to its gene.

Also the extent to which the pattern of editing of homologous transcripts is conserved between different trypanosome species greatly varies. The frameshift in the *MURF3* gene (this paper, unpublished results cited by Shaw *et al.*, 1988) and that in the *coxII* gene (Benne *et al.*, 1986) is repaired in an identical fashion in different trypanosomes. Only slight species-specific differences exist in the way in which the respective apocytochrome *b* transcripts are edited in the 5′ terminus (Feagin *et al.*, 1988a). For other transcripts the differences are much larger. For example, significant differences in editing exist between the *C.fasciculata* and *L.tarentolae coxIII* transcripts (Shaw *et al.*,

```
NNNNNNNNNNNNGCCGACUACACGAUAAuuAUAuuuUAuAuuuAuu--A-----AuuGuuuuuuuACACUU... L.t. RNA (Shaw et al.,Cell 1988)
                GCAUGACAA--ACG---UAuAuuuAuuuuAuuuuuAuuGuuuuuuuGCACUU... C.f. cDNA (This paper)
[NNNNAGUAUAUUCGACUGCAUGNCAA--ACG---UAuAuuNNNNNNNNNNNNNNNNNNNNNNNNNNNN... C.f. RNA (Shaw et al.,Cell 1988)]
  (5'← 3')
```

Fig. 6. The 5'-terminal sequence of *MURF3* transcripts in *L.tarentolae* and *C.fasciculata*. The *L.tarentolae* (L.t.) and *C.fasciculata* (C.f.) RNA sequence is derived from Shaw *et al.* (1988). The inserted nucleotides are in lower case. A putative initiator AUG codon in *C.fasciculata* is underlined.

1988), which, in turn, are much less extensively edited than the *T.brucei* version (Feagin *et al.*, 1988b). Still the amino acid sequence of the respective proteins is highly similar. A species-specific RNA editing pattern is also observed for the 5' terminus of the *MURF3* gene in *C.fasciculata* (this paper) and *L.tarentolae* (Shaw *et al.*, 1988). As shown in Figure 6, 22 Us are inserted in *C.fasciculata*, and only 20 in *L.tarentolae* in a rather different manner. These differences in U insertion pattern result in slightly different protein sequences (see Figure 6).

An additional complication is provided by the fact that in both species an in-frame AUG codon is lacking in the edited transcripts. In principle, in-frame AUG codons could be provided by U insertion further upstream. We are unable to verify this possibility, however, since we cannot find edited cDNAs that extent beyond the sequences show in Figure 3B. In fact, a fairly large number (three out of seven) of cDNAs that are edited in the 5' terminus, terminate at exactly the same position (see arrow in Figure 3B), indicative of a strong stop for the reverse transcriptase during first strand cDNA synthesis.

It could also be envisaged that other than conventional initiator codons are used, as proposed for AUG-less trypanosome mt genes, before we realized how widespread RNA editing is (see Sloof *et al.*, 1987). A third alternative is provided by the occurrence of a cDNA that combines a non-edited 5' terminus to an edited frameshift site. As demonstrated in Figure 3, a genomic AUG present in the non-edited sequences lines up with one of the two consecutive AUG triplets located at the N terminus of *M.polymorpha* chloroplast ORF392. This AUG codon could, therefore, serve as initiator triplet on *MURF3* (m)RNAs which possess a non-edited 5' terminus and a repaired frameshift. If this were the case, one would have to assume, though, that RNA editing can regulate mt protein synthesis also in a negative sense, since completely edited *C.fasciculata MURF3* transcripts are untranslatable. It would also not be a valid explanation for *L.tarentolae MURF3*, since the non-edited sequences in this species do not contain an AUG triplet. It seems plausible, though, that in *Leishmania*, the AUG could be produced by editing but only in a minor fraction of the *MURF3* transcripts. This would not be detected by the sequence analysis of total RNA used by Shaw *et al.* (1988). Only sequence analysis of the N terminus of the MURF3 protein can settle this issue.

The *C.fasciculata MURF3* cDNA sequences show no further sign of editing in the protein-coding and non-coding region, except for the presence of U residues in the 3'-terminal extension. Although we have no formal proof that these poly(AU) tails are created by the insertion of U residues into a pre-existing poly(A) tail, we consider this an attractive possiblity. The presence of a poly(AU) extension has been observed before for another gene (Benne *et al.*, 1986; see Figure 4). It is not likely that the differences between the (AU) patterns observed for different genes reflect differences in RNA stability, since the sequences even differ between transcripts from the same gene. Rather, they may result from a U insertion process that is less strictly controlled than in protein-coding areas (also suggested by Feagin *et al.*, 1988b).

Although the mechanism of RNA editing, at present, remains a mystery, the main characteristics can now be specified: (i) RNA editing occurs in purine-rich sequences, that have no obvious similarity. (ii) Except in the 3'-terminal extension, RNA editing results in very precise, (gene- and sometimes species-) specific insertion/deletion of U residues. (iii) The ratio of edited/non-edited transcripts varies between genes. The frameshift area from the *coxII* transcript occurs predominantly in the edited form (Benne *et al.*, 1986), whereas the amounts of edited and non-edited RNA are more equal for the *MURF3* (this paper) or the apocytochrome *b* transcripts (Feagin *et al.*, 1987). Whether this is a consequence of regulation of gene expression at the level of RNA editing is an open question. (iv) Editing may occur at multiple sites in a transcript. For the *MURF3* gene, 'hybrid' transcripts are found which combine a 5' non-edited region to 3' edited areas. Also a *TURF2* (now *coxIII*) cDNA was found that showed similar characteristics (Benne *et al.*, 1986).

The occurrence of transcripts that are edited at the 3' terminus, but not at the 5' end, could imply that RNA editing is a post-transcriptional process, which proceeds in a 3'→5' direction (also suggested by Feagin *et al.*, 1988b). If this is true, mitochondrial ribosomes (that move from 5'→3') are not involved. It also argues against the involvement of virus-like dsRNA replicative intermediates, since these would not produce RNAs that are only partially edited.

Instead, one could envisage a (nucleo)protein (complex), an 'editosome', that moves upstream and inserts/deletes U residues at specific sites of the primary transcript. Breaks may be introduced into the RNA at specific stages of this process, since we find an otherwise unexplainably high proportion of truncated cDNAs that terminate in the edited regions: two in the frameshift area, four in the 5'-terminal region, but none in the remaining 1000 or so nucleotides of the protein coding sequence downstream of the 5' editing site. In all these cDNAs U insertion is precise and complete up to the 5'-ultimate nucleotide. This finding, again, is indicative of a 3'→5' movement of the editing machinery. We found no cDNAs with runs of Us that exceeded the number of Us present in the mature transcript, as was reported for the *T.brucei* apocytochrome *b* gene (Feagin *et al.*, 1987).

The factors that govern the insertion/deletion pattern in each case remain obscure. Even in 'simple' examples of RNA editing, consensus primary and secondary structures

appear to be absent around insertion/deletion sites. For more intricate RNA editing cases such a model would be clearly inadequate and naive. Even in the most extensively edited transcripts (the *coxIII* transcript in *T.brucei*), however, the order of Gs, As and Cs is identical to that of the corresponding 'gene'. Except for the Us, the nucleotide sequence of the RNA is dictated by a DNA template. Further investigations into the mechanism of RNA editing should take this into account.

Materials and methods

Cell culture, RNA and DNA isolation

C.fasciculata was grown in culture as described by Kleisen *et al.* (1975). Total cellular DNA and mitochondrial DNA were isolated according to Borst and Fase-Fowler (1979). Total cellular RNA was isolated using the hot-phenol method and subsequently enriched for poly(A)$^+$ RNA by two cycles of oligo(dT) cellulose as described by Hoeijmakers *et al.* (1981). Plasmid and M13 RF DNA were prepared according to Birnboim and Doly (1979).

Gel electrophoresis

Agarose gel electrophoresis and blot analysis of RNA and DNA fragments were performed as described by Benne *et al.* (1983).

Cloning and sequencing of the MURF3 gene

The cloning and sequencing procedure of the *C.fasciculata MURF3* gene (previously called *CURF2/1*) has been described before (Sloof *et al.*, 1987). In the experiments described in this paper M13 and pUC clones were used that contain maxi-circle segments on which (part of) this gene is localized.

cDNA construction and cloning

C.fasciculata cDNA libraries were constructed via the procedure of Gubler and Hoffman (1983), as modified in a protocol provided by the manufacturers (Amersham, UK) of a cDNA synthesis kit that was used. In short, first strand synthesis was primed with oligo(dT) (library A) or with oligonucleotides (library B), the nucleotide sequence of which is derived from *MURF3* genomic sequences (see below). After second strand synthesis with *Escherichia coli* DNA polymerase and RNase H-treated hybrids of the first strand and the RNA template, the cDNAs were blunted with T4 DNA polymerase and directly cloned into the *Hin*cII site of pUC19. Transformation of *E.coli* DH5 yielded libraries of $5-10 \times 10^6$ (library A) and 2.5×10^4 (library B) recombinant clones, respectively. These libraries were screened with a DNA fragment that contains the 5'-terminal moiety of the *MURF3* gene (nt 2866–3458, Sloof *et al.*, 1987); 17 different *MURF3* derived cDNA clones were picked up.

Hybridization and sequence analysis

The following oligonucleotides were used:

	5'→3'	Nucleotide coordinates in Sloof *et al.* (1987):
C35	ATCTAATCTTCATCGGTAGA	3333–3352
C41	CTAATCTATCAAAATACGGT	see below
C34	CATAAGGATAGCAAATGTTC	3374–3393
C27	TGCAAATGAGCAACCTGG	4088–4105

C27 and C34 were used in the construction of library B; C35 is complementary to the frameshift region of non-edited *MURF3* transcripts ('DNA' oligo), C41 is complementary to the corresponding region of edited transcripts ('RNA' oligo), C35 and C41 were 5'-end labelled and 5×10^7 c.p.m. was used in hybridization experiments with blots of RNA and DNA fragments (blotted onto nitrocellulose) as described by Winter *et al.* (1982).

Hybridization was performed in sealed bags in 6 ml hybridization mix ($6 \times$ SSC, 0.1% SDS, 0.2% Ficoll, 0.2% bovine serum albumin, 0.2% polyvinyl pyrrolidone, 150 μg/ml salmom sperm DNA). Filters were washed in 2×50 ml $6 \times$ SSC at room temperature followed by 1.5 min washes with 2×50 ml $6 \times$ SSC at 52°C (C41) or 54°C (C35).

C34 is complementary to a region of the *MURF3* transcript located ~30 nt downstream of the frameshift area. This oligonucleotide was used in primer-extension assays with reverse transcriptase (1 U) and *C.fasciculata* poly(A)$^+$ RNA (5 μg) as template, essentially as described by Tabak *et al.* (1984). This oligonucleotide was also used in sequence analysis of the frameshift area of cDNA clones with double stranded pUC19 recombinant DNAs as template. The DNA was prepared as described by Zhang *et al.* (1988) and used in the dideoxy sequencing procedure of Sanger *et al.* (1977).

A complete sequence analysis of an (edited) *MURF3* cDNA was also obtained by utilizing oligonucleotide C27, and the forward and reverse primer of the pUC 19 vector in the analysis.

Acknowledgements

We thank Professors P.Borst and H.F.Tabak for their continuing interest and comments on the manuscript. We acknowledge the contribution of Mr P.Smits during the initial phase of this work. We are greatly indebted to Mr M.de Haan and Professor L.A.Grivell of the Department of Molecular Cell Biology of the University of Amsterdam and Dr H.v.Steeg of the RIVM, Bilthoven for generous gifts of synthetic oligonucleotides.

References

Benne,R. (1985) *Trends Genet.*, **1**, 117–121.
Benne,R., De Vries,B.F., Van Den Burg,J. and Klaver,B. (1983) *Nucleic Acids Res.*, **11**, 6929–6941.
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.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.
Borst,P. and Fase-Fowler,F. (1979) *Biochim. Biophys. Acta*, **565**, 1–12.
Borst,P. and Hoeijmakers,J.H.J. (1979) *Plasmid*, **2**, 20–40.
De Vries,B.F., Mulder,E., Brakenhoff,J.P.J., Sloof,P. and Benne,R. (1988) *Mol. Biochem. Parasitol.*, **27**, 71–82.
Englund,P.T. (1981) In Levandowsky,M. and Hynter,S.H. (eds), *Biochemistry and Physiology of Protozoa*. Academic Press, New York, Vol. 4, pp. 334–383.
Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *Cell*, **49**, 337–345.
Feagin,J.E. and Stuart,K. (1988) *Mol. Cell. Biol.*, **8**, 1259–1265.
Feagin,J.E., Shaw,J.M., Simpson,L. and Stuart,K. (1988a) *Proc. Natl. Acad. Sci. USA*, **85**, 539–543.
Feagin,J.E., Abraham,J.M. and Stuart,K. (1988b) *Cell*, **53**, 413–422.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Hensgens,L.A.M., Brakenhoff,J., De Vries,B.F., Sloof,P., Tromp,M.C., Van Boom,J.M. and Benne,R. (1984) *Nucleic Acids Res.*, **12**, 7327–7344.
Hoeijmakers,J.H.J., Snijders,A., Janssen,J.W.G. and Borst,P. (1981) *Plasmid*, **5**, 329–350.
Kleisen,C.M., Borst,P. and Weyers,P.J. (1975) *Biochim. Biophys. Acta*, **390**, 155–167.
Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,T., Sano,S., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aoata,S.-I., Inokuchi,H. and Ozeki,H. (1986) *Nature*, **322**, 572–574.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Shaw,J.M., Feagin,J.E., Stuart,K. and Simpson,L. (1988) *Cell*, **53**, 401–411.
Simpson,L. (1986) *Int. Rev. Cytol.*, **99**, 119–179.
Simpson,L., Neckelmann,N., De la Cruz,V., Simpson,A.M., Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *J. Biol. Chem.*, **262**, 6182–6196.
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.
Sloof,P., Van Den Burg,J., Voogd,A. and Benne,R. (1987) *Nucleic Acids Res.*, **15**, 51–65.
Stuart,K. (1983) *Mol. Biochem. Parasitol.*, **9**, 93–104.
Tabak,H.F., Van der Horst,G., Osinga,K.A. and Arnberg,A.C. (1984) *Cell*, **39**, 623–629.
Winter,G., Fersht,A.R., Wilkinson,A.J., Zoller,M. and Smith,M. (1982) *Nature*, **299**, 756–758.
Zhang,H., Scholl,R., Browse,J. and Sommerville,C. (1988) *Nucleic Acids Res.*, **16**, 1220.

Received on May 26, 1988