The EMBO Journal vol.9 no.1 pp.257 - 262, 1990

RNA editing in transcripts of the mitochondrial genes of the insect trypanosome *Crithidia fasciculata*

Hans Van der Spek, Dave Speijer, Gert-Jan Arts, Janny Van den Burg, Harry Van Steeg[1], Paul Sloof and Rob Benne

E.C.Slater Institute for Biochemical Research, Faculty of Medicine, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam and [1]National Institute of Public Health and Environmental Protection (RIVM), PO Box 1, 3720 BA, Bilthoven, The Netherlands

Communicated by P.Borst

With the aid of cDNA and RNA sequence analysis, we have determined to what extent transcripts of mitochondrial maxicircle genes of the insect trypanosome *Crithidia fasciculata* are altered by RNA editing, a novel mechanism of gene expression which operates via the insertion and deletion of uridine residues. Editing of cytochrome *c* oxidase (cox) subunit II and III transcripts and of maxicircle unidentified reading frame (MURF) 2 RNA is limited to a small section and results in the creation of a potential AUG translational initiation codon (coxIII, MURF2) or the removal of a frameshift (coxII). No differences with the genomic sequences were observed in the remainder of these RNAs. Surprisingly, NADH dehydrogenase subunit I transcripts were completely unedited in the coding region, implying that an AUG translational initiation codon is absent. The partial ribosomal RNA sequences determined also conform to the gene sequences. Together these results lead to the conclusion that the unusual sequences predicted by the protein and rRNA genes must indeed be present in the gene products. Editing also occurred in the poly(A) tail of RNAs from all protein genes, including those that are unedited in the coding region. The tails display a large variation in AU sequence motifs. Finally, some cDNAs contained sequences absent from both the DNA and the edited RNA. Some of these may represent intermediates in the RNA editing process. We argue, however, that long runs of T may be artefacts of cDNA synthesis.

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

Introduction

The mitochondrial (mt) genes of kinetoplastid trypanosomes are located in the maxicircle component of the mtDNA network (reviewed in Benne, 1985; Simpson, 1987). The genes thus far identified comprise ribosomal (r) RNA genes encoding small rRNAs of 9S and 12S, protein genes encoding subunits of the respiratory chain complexes such as cytochrome *c* oxidase (cox) subunits I, II and III, NADH dehydrogenase (ND) subunits I, IV and V, apocytochrome (cyt) *b* and some maxicircle unidentified reading frames (MURFs). One of these, MURF3, is homologous to a recently sequenced cDNA from the nuclearly encoded 49 kd subunit of the bovine ND complex (Fearnly *et al.*, 1989). In general, the degree of similarity between the trypanosome mt protein genes and their homologues in other organisms is low, ranging from 20 to 40% identity at the amino acid level, one of the striking features being their high cysteine content (see Benne, 1985). The rRNAs seem to be quite unusual too, both with respect to size (which is only 60% of that of the corresponding mammalian mt rRNAs) and with respect to primary and secondary structures (Sloof *et al.*, 1985; De la Cruz *et al.*, 1985a,b).

The study of the expression of these genes has produced a few surprises, the most spectacular of which was the discovery that at certain sites the uridine sequence of the RNAs is not in agreement with the genomic sequence, whereas G, A and C sequences are always co-linear (Benne *et al.*, 1986; Feagin *et al.*, 1987, 1988a; Shaw *et al.*, 1988; Van der Spek *et al.*, 1988). No DNA equivalent for these aberrant RNA segment(s) could be found, so it was inferred that they were the result of a novel process of uridine insertion and deletion, which was termed RNA editing (Benne *et al.*, 1986). The mechanistic principles that guide the editing process and the enzymology are mysterious at present, but the available evidence suggests that it is post-transcriptional and proceeds from 3′ to 5′ on an RNA to be edited (for reviews see Benne, 1989; Simpson and Shaw, 1989). In general, editing increases the translatability of the RNAs in question since potential AUG start codons are created and/or frameshifts removed, but also the poly(A) tail can be edited (Feagin *et al.*, 1988c; Benne, 1989; Simpson and Shaw, 1989). In one case (the *Trypanosoma brucei* coxIII transcript (Abraham *et al.*, 1988; Feagin *et al.*, 1988c), it was shown that an RNA can be edited over its entire length, 55% of its nucleotides being produced by RNA editing, but in general only small RNA sections have been analysed. For most transcripts, therefore, the extent to which sequences are altered by RNA editing is unknown. Consequently, the question arises whether the actual sequence of the gene products could be less unusual than that predicted by the DNA sequence.

In an attempt to answer this question we have undertaken systematic cDNA analysis of a number of mt maxicircle protein genes and the two rRNA genes of the insect trypanosome *Crithidia fasciculata*. One of the protein genes in question (NDI) is of further interest, because the genomic sequences lack a canonical ATG translational initiation codon (Van der Spek *et al.*, 1989) and the corresponding region of the transcripts has not yet been analysed. Another gene (MURF2), although continuous in *C.fasciculata* (Van der Spek *et al.*, 1989), is frameshifted in *Leishmania tarentolae* (De la Cruz *et al.*, 1984).

This approach has the additional advantage that cDNAs derived from individual RNA species can be analysed, which allows the identification of sequences (if present) that do not correspond to either the DNA or the edited RNA and could be derived from intermediates of the RNA editing process.

Moreover, the site of polyadenylation and the AU sequence motif of individual trancripts can be determined.

Results

Transcript analysis

We first studied the transcription pattern of the NDI, coxII, MURF2 and coxI genes (for rRNA and coxIII gene transcription, see Sloof *et al.*, 1985, 1987). Figure 1 shows that each cDNA probe lights up a major RNA species on a Northern blot, the size of which is only slightly larger than the protein-coding area of the corresponding gene. One other major RNA species, which is ~200–400 nt larger, is found for NDI, coxII and MURF2, resulting in the characteristic overlapping transcript pattern observed for many trypanosome mt genes (Feagin *et al.*, 1985, 1986; Simpson *et al.*, 1985). The interrelation of the two RNA species is not known: both forms are retained on oligo(dT)–cellulose and must contain a poly(A) tail, and both the large and small forms of the cyt *b* and MURF3 doublets hybridize to edited oligonucleotides (Feagin *et al.*, 1987; Van der Spek *et al.*, 1988).

The 5′ end of NDI RNA is unedited

A schematic map of the genes, the transcripts of which were analysed in this study, is given in Figure 2. The gene sequences determined predict translationally defective RNAs for all but one of the protein genes: coxIII, NDI and MURF2 genes lack a canonical ATG initiation codon (Sloof *et al.*, 1987; Van der Spek *et al.*, 1989) and the cox II gene contains a gene internal frameshift (Benne *et al.*, 1986). Oligonucleotide-primed sequence analysis of the corresponding RNA sections has indeed shown that a major fraction of the cox III, cox II and MURF2 RNAs is edited, resulting in a correction of the defects (Benne *et al.*, 1986; Shaw *et al.*, 1988; see Figure 2). We have performed a similar analysis for the 5′ terminal section of NDI RNA. In this case a 20mer oligonucleotide was used, the sequence of which corresponds to a section of the gene encoding a highly conserved (Anderson *et al.*, 1981; Pritchard *et al.*, 1986) amino acid motif ~90 nt downstream of the (putative) initiation codon. This section of the *C.fasciculata* NDI RNA is therefore not likely to be edited, even if the translational start site is. No differences could be observed, however, between the resulting RNA sequences and the published genomic sequences (Figure 3). This implies that the bulk of the NDI RNA does not contain an AUG translational start codon.

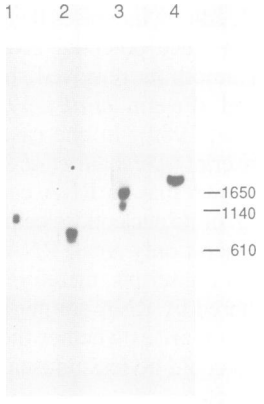

Fig. 1. Northern blot analysis of NDI (**lane 1**), coxII (**lane 2**), MURF2 (**lane 3**) and coxI (**lane 4**) transcripts. As a probe, edited (coxII and MURF2) or unedited (NDI and coxI) cDNAs have been used. The experimental details are described in Materials and methods.

cDNA sequences at (putative) editing sites

The results of the cDNA sequence analysis confirm (and extend) the direct RNA sequence determination. For coxIII, coxII and MURF2 a large, but varying, proportion of the cDNAs contain the edited sequence, although unedited cDNAs are also present in all cases (Figure 4A, see also Figure 2). All 25 NDI cDNAs, however, display the unedited sequence. In two cases (coxIII and MURF2) cDNAs were found, the sequence of which is different from both the DNA and the (edited) RNA. One coxIII cDNA (out of seven) has two extra Us deleted upstream of the putative AUG initiation codon, whereas no less than four out of seven MURF2 cDNAs have anomalous U sequences. Interestingly, in two of these, an in-frame AUG codon is created at a site different from that in the fully edited RNA (see Figure 4A). Two MURF3 cDNAs have long 5′ terminal runs of T.

cDNA sequence analysis in other areas

We have determined the complete nucleotide sequence of cDNAs that are derived from coxII, coxIII, MURF2 and NDI transcripts. We have also obtained cDNA sequences that correspond to the major part of 12S and 9S rRNA and to the 5′ and 3′ ends of coxI RNA. The total extent of the collected cDNA sequences is indicated in Figure 2. We do not find (further) differences with the genomic sequence for any of the cDNAs with the exception of 5′ terminal runs of T that occurred in a large fraction of the coxI cDNA collection and in one 9S cDNA (Figure 4B, see also Figure 2). Similar runs of T are also present in MURF2 cDNAs (Figure 4A) and in *T.brucei* cyt *b* (Feagin *et al.*, 1987) and coxIII (Abraham *et al.*, 1988) cDNAs. Long runs of uridines were not found in the corresponding region of

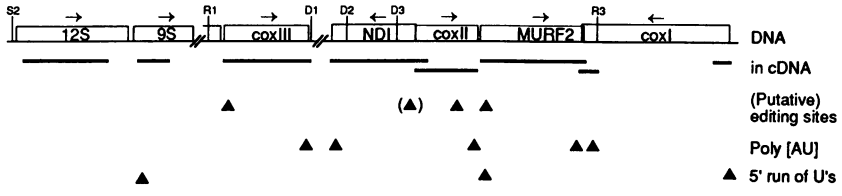

Fig. 2. Partial gene and restriction map of *C.fasciculata* maxicircle DNA segments. Putative RNA editing sites in the protein coding sequences and the poly(A) tail are indicated, as are sites at which runs of T(U) are found. The area of the sequenced portion of the cDNAs has been indicated by a thick line. For details, see Materials and methods. S, *Sal*I; R, *Eco*RI; D, *Hin*dIII, see Hoeijmakers *et al.* (1982).

9S and coxI RNA, however, as determined by reverse transcriptase-directed sequence analysis primed by the appropriate oligonucleotides (Figure 5). The sequences found precisely match those of the corresponding gene segment.

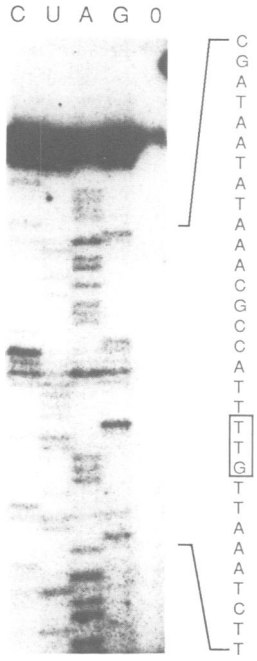

Fig. 3. 5′ end sequence analysis of NDI RNA. Primer C51 was used in the experimental procedure described in the Materials and methods. The 15 3′ nucleotides of this primer correspond to a five amino acid sequence motif (ERRIL), highly conserved in evolutionary distant organisms such as *Homo sapiens* (Anderson *et al.*, 1981) and *Paramecium* (Pritchard *et al.*, 1986). The RNA sequence fully conforms to the genomic sequence (Van der Spek *et al.*, 1989) which is partially given in the figure. A putative TTG initiation codon (see text) is boxed. This codon is part of the only sequence motif (ATTTTG) that can be converted by editing into ATG in this area of the transcript.

Editing of poly(A) tails

A substantial fraction of the cDNAs from all protein genes studied showed evidence of U insertion into the poly(A) tail of the respective RNAs, although the ratio of edited and unedited tails varies between genes (Figure 6, see also Figure 2). An enormous diversity in AU sequence motifs appears to exist, adding to those already known from other studies (Feagin *et al.*, 1988c; Van der Spek *et al.*, 1988; Benne, 1989; Simpson and Shaw, 1989). This results in a wealth of different tails, even for transcripts from the same gene. Also the site of polyadenylation can be heterogeneous, since NDI and coxII tails are hooked up to the RNA at positions that vary a few nucleotides. Finally, editing of tails can occur in RNAs that are unedited anywhere else (NDI and, presumably, coxI).

Discussion

NDI transcripts lack an AUG translational initiation codon

In this paper we report the results of sequence analysis of a large number of cDNAs derived from five mt protein genes and the mt rRNA genes of the insect trypanosome *C.fasciculata*. The analysis, as summarized in Figures 2 and 4, confirms that RNA editing occurs at the frameshift position of coxII RNA and at the 5′ end of coxIII and MURF2 transcripts. In addition, editing is found in the poly(A) tail of all protein gene-derived RNAs (Figure 6). No editing was observed at the 5′ end of coxI and coxII RNAs which contain a genomically encoded canonical (AUG) translational initiation codon. Similar observations were made for a number of AUG-containing *L.tarentolae* and *T.brucei* transcripts (Shaw *et al.*, 1988). Surprisingly, however, the 5′ end of NDI RNA, which lacks a gene-encoded AUG, is unedited (Figures 3 and 4). In principle, two explanations for this phenomenon can be envisaged:

(i) An edited AUG containing RNA is present at low concentrations, below the limits of detection of the analysis. This cannot be ruled out by the present data, particularly

A

Gene	Sequences	Correspond to	# of U's ins. /del.	sites	# of cDNA's
coxIII	-AUAUA G A CG G A G GG GUUGAG GG G G UUAUU-	DNA	0/0	0	3
	-AUAUAuGuuuuuAuuuCGuGuuAuuuuuGuuGGuG UGAGuGGuGuuuuuGuuuuuUUA U-	RNA	32/2	9	3
	-A A AuGuuuuuAuuuCGuGuuAuuuuuGuuGGuG UGAGuGGuGuuuuuGuuuuuUUA U-	?	32/4	11	1
coxII	-A G A A-	DNA	0/0	0	2
	-AuuGuAuA-	RNA	4/0	3	10
NDI	-CGAUAAUAUAAACGCCAUUUUGUUAAAUCUUGA-	DNA/RNA	0/0	0	25
MURF2	-A G GG G AA AG A A G G A CUG A-	DNA	0/0	0	2
	-AuGuuuGGuuGuuuuAAuuuAGuuuuAuuuuuAuGuuuuG A CUGuA-	RNA	28/0	10	1
	uuuAuuuuuAuGuuuuG A CUG A-	?	12/0	3	1
	uuuuuuuuuuuuAuGuuuuG A CUG A-	?	18/0	4	1
	uuuuuuuuuuuuuGuAuuCUGuA-	?	≥17/0	4	2

B

Gene	Sequences	Correspond to	n	# of cDNAs
coxI	-UUUAUUUUUUCUAAAUAUG-	DNA/RNA	0	2
	-(u)$_n$UUUUUUCUAAAUAUG-	?	3-8	6
9S	-UUAAUUUUUUUUAAAGUUU-	DNA/RNA	0	6
	-(u)$_n$UUUUUUUUAAAGUUU-	?	2	1

Fig. 4. (**A**) Compilation of cDNA sequences at editing sites. Putative translational initiation codons are underlined. Inserted Us are in lower case. The number (#) of cDNAs possessing a certain sequence has been given. (**B**) Runs of T in 9S and coxI cDNAs. The sequence of clones that have the genomic (and RNA) sequence in this area is also given.

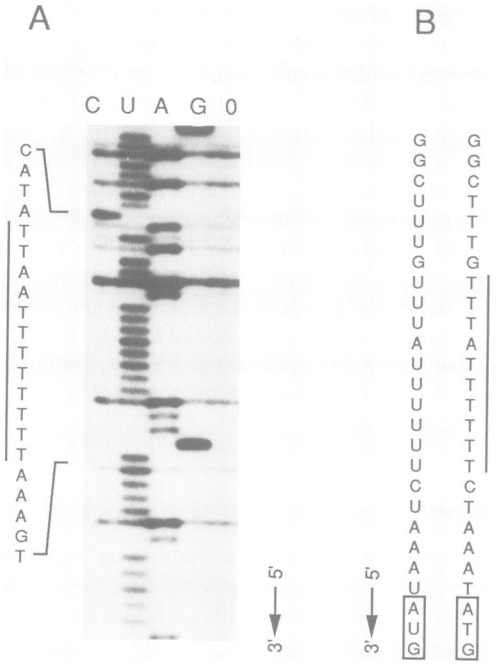

Fig. 5. Sequence analysis of 9S and coxI RNA. (**A**) Primer C67 was used in the experimental procedure described in Materials and methods. The corresponding genomic sequence is given. The line next to the sequence indicates the position of the runs of T found in the cDNAs. (**B**) The same as (A) for coxI RNA sequence analysis, with primer C57. The putative translational initiation codon is boxed.

Gene	Sequences	
coxIII	DNA	AAATATTTAATCCACACAAATTCTATTTTTAA
		---$A_{29}U A_n$
		---$U_2 A_n$
NDI	DNA	not determined
		---AACUUUUUGCA$_7$U$_6$A$_n$
		---AACUUUUUGAU$_5$A$_n$
		---AACUUUUUA$_{26}$U$_2$A$_2$U$_2$A$_3$UA$_n$
coxII	DNA	ATATAGGTATGTAATCTAAAATAAAAGGGG
		---$U_8 A_n$
		---$A_{17} U_n$
		---$U_5 A_{11} U_n$
		---$AU_2 A_2 U_n$
		---$U_3 AU_5 A_n$
		---$A_3 UA_n$
		---$U_2 A_2 UA_n$
MURF2	DNA	ACAAAGATTGTCAAAAAAAAATAAAAAAAAC
		---UA_n
coxI	DNA	AAAAGAAGAGTCACGCAAAAATCGT
		---$A_5 UA_n$

Fig. 6. Compilation of poly(AU) tail sequence motifs. The tails are lined up underneath the corresponding genomic sequence, so that the putative polyadenylation site can be inferred. The NDI genomic sequence has not been determined, but the cDNA sequences show that different polyadenylation sites are used. For each gene, cDNAs with tails consisting exclusively of As are also found.

if one assumes that the edited sequences, for some reason, resist cloning. One could argue, for example, that the presence of poly(AU) tails without long runs of A would interfere with oligo(dT)-primed reverse transcription of a certain RNA. For most genes, however, edited cDNAs are present in the libraries in a raito to unedited versions more or less in agreement with the outcome of other types of analysis. For example, oligonucleotide-primed sequence analysis and hybridization experiments indicate that >80% of coxII RNA, but only ~50% of MURF3 RNA, is edited at the frameshift position (Benne *et al.*, 1986; Van der Spek *et al.*, 1988). These numbers nicely agree with the cDNA data (Figure 4, Van der Spek *et al.*, 1988). Furthermore, edited coxIII and MURF2 cDNAs are abundantly present, as might be expected from the results of the direct sequence analysis of the corresponding RNA segments (Feagin *et al.*, 1988b; Shaw *et al.*, 1988). It is clear, therefore, that the nature of the poly(AU) tails does not seriously interfere with the cDNA cloning of any of these transcripts. It is also interesting to note that the NDI cDNAs have edited tails, making it even more difficult to understand why a (hypothetical) NDI RNA that is edited at the 5′ end would be unclonable.

(ii) A non-canonical triplet is used as translation start codon on unedited NDI RNA. Both in *C.fasciculata* and in *T.brucei* (Hensgens *et al.*, 1984), a UUG triplet (see Figure 3) lines up with a putative AUG initiation triplet of *L.tarentolae* NDI. UUG can be used as an initiation codon both in eukaryotic and in prokaryotic translation (Kozak, 1983; Peabody, 1989), albeit with a lower efficiency.

In order to decide between these two alternatives, *in vitro* translation experiments in systems programmed with trypanosome mtRNA could be informative. Unfortunately, a heterologous system derived from reticulocytes, wheat germ or bacteria cannot be easily used, since, like mtRNAs from many organisms, trypanosome mtRNAs contain multiple UGA codons (Benne *et al.*, 1983; De la Cruz *et al.*, 1984; Hensgens *et al.*, 1984). These would be read as 'stop', unless excess suppressor tRNA is added. Moreover, these RNAs lack the specific *cis*-acting elements required for translation by bacterial or eukaryotic ribosomes. One of the long-standing goals of our research is, therefore, to develop a homologous, trypanosome-derived translation system, in which one could study the mechanism of translation initiation of NDI RNA and of other RNAs that do not seem to possess a proper initiation codon (as reviewed in Simpson and Shaw, 1989).

Editing in poly(A) tails

Poly(AU) tails are found on transcripts derived from all protein genes, including NDI and coxI. We have not found any evidence for U insertion/deletion in other areas of NDI and coxI RNAs, illustrating that editing of the poly(A) tail is not strictly coupled to editing at other sites. This could imply that the presence of Us in the tail is important for stability and/or translatability of trypanosome mtRNA. Alternatively, the poly(A) tail may be essential for recognition of a transcript by the editing machinery, a stage at which editing of the tail could occur. Editing at other sites would follow only if specific, as yet unidentified, primary and/or secondary structure motifs are present. It is not likely, though, that specific AU sequence motifs have a special function, considering the fact that almost all individual tails are different, even for RNAs from the same gene. This variability in tail apparently also involves the location of the polyadenylation site (Figure 6). Another consequence of the presence of U residues in a poly(A) tail is the potential for basepairing and secondary structure of these tails. It can be envisaged that this would result in anomalous migration in agarose gels and that differences in tail structure rather than size are, in fact, responsible for the differences in migration of the overlapping transcripts produced by trypanosome mt genes (see Figure 1).

RNA editing affects small sections of transcripts

No differences with the genomic sequences were found in the remainder of the protein gene-derived cDNA sequences or in the rRNA-derived cDNAs, which cover 68 and 75% respectively of the 9S and 12S rRNA. Although we cannot formally exclude that we have missed editing sites in the analysis, we consider this to be unlikely in view of the large number of cDNAs used to compose the sequence of each transcript (see Materials and methods). This implies that the unusual protein and rRNA sequences predicted by the genes in question are indeed transferred to the products without (much) interference from the RNA editing process. The question arises, however, whether extensively edited RNAs (so-called pan-edited transcripts; Simpson and Shaw, 1989) exist that escaped detection by the genomic DNA fragments used as probes in the cDNA screening. We consider this highly unlikely for the following reasons.

(i) Pan-edited genes are G rich in the RNA strand (Simpson *et al.*, 1987; Simpson and Shaw, 1989), e.g. the *T.brucei* coxIII gene has a G content of 40%. The overall G content of the genes studied in this work ranged from 10% (the rRNA genes) to 15–20% (the protein genes).

(ii) In all protein genes, regions of a high degree of evolutionary conservation are present (otherwise the genes could not have been identified, see Hensgens *et al.*, 1984; Sloof *et al.*, 1985, 1987; Van der Spek *et al.*, 1988). RNA sequences corresponding to these sections are not likely to be extensively edited (and should hybridize to the probes used) even if other parts of the RNA are substantially altered.

(iii) Partially edited RNAs are found for pan-edited genes (Abraham *et al.*, 1988; Feagin *et al.*, 1988c; Simpson and Shaw, 1989) which contain large, unedited segments at the 5′ end. This produces a continuum of transcripts of increasing size that still hybridize to unedited probes. Such transcripts are not visible on the Northern blots of Figure 1, nor are they represented in the cDNA collection.

(iv) Editing results, in general, in a 2-fold increase in size of the corresponding RNA segments. In mtRNA preparations, the major EtBr-stainable RNAs migrate at the positions corresponding to the size predicted by the 9S and 12S rRNA genes. These RNAs hybridize to maxicircle probes (Hoeijmakers *et al.*, 1981). Also, the coding sequences of the mt protein genes are, at most, only slightly smaller than those in other organisms.

Together, these points strongly argue against the possibility of pan-editing of these genes.

The mechanism of RNA editing

The consensus view on the mechanism of RNA editing shared by most workers in the field (see Benne, 1989; Simpson and Shaw, 1989) can be summarized as follows: a multi-component RNA editing enzyme works its way over the RNA substrate in a 3′ to 5′ direction. In the course of this process, endonucleolytic breaks are introduced at editing sites, and uridine residues are inserted and/or deleted by the combined action of an uridilyl-transferase and an exonuclease. RNA ligase then joins the two RNA moieties. The most intriguing questions of how the editing sites are selected and how the precise number of Us, inserted/deleted is determined remain to be answered. Views range from the involvement of a minus strand RNA template (Maizels and Weiner, 1988) to a role for *cis*-acting sequences on the edited RNA itself (Abraham *et al.*, 1988; Feagin *et al.*, 1988c; Benne, 1989; Simpson and Shaw, 1989).

In the analysis that led to this scheme a number of cDNAs were found that are partially edited, others contain U sequences that are different from both the DNA and the edited RNA sequence (Abraham *et al.*, 1988; Feagin *et al.*, 1988c; Benne, 1989; Simpson and Shaw, 1989). Sequences of a similar nature are also found in this analysis (see Figure 4). Two MURF2 cDNAs have long 5′ terminal runs of T. Similar runs of T have been found in *T.brucei* cyt *b* (Feagin *et al.*, 1987) and coxIII (Abraham *et al.*, 1988; Feagin *et al.*, 1988c) cDNAs. It has been proposed (Feagin *et al.*, 1987, 1988c; Abraham *et al.*, 1988; Simpson and Shaw, 1989) that runs of U, the number of which may largely exceed the number present in the final sequence, are added to RNAs at editing sites by uridilyl transferases at an intermediary stage of the editing process. In this view, the correct number of Us is achieved by an exonuclease-mediated trimming process. We have found runs of T, however, also to occur at the 5′ terminus of six out of eight CoxI and one out of seven 9S cDNAs that otherwise conform to the mtDNA sequence (Figure 4B). Moreover, direct sequence analysis of the corresponding RNA segments showed such runs to be absent from the bulk of the RNA without any evidence of editing in surrounding areas (Figure 5). It is likely, therefore, that they arise during cDNA synthesis from oligo(dT)-mediated back-priming on A-rich segments of the first cDNA strand upon partial digestion of the RNA in the RNA–cDNA hybrid by RNase H (see Van der Spek *et al.*, 1988). The regions in the template strand of the coxI and 9S genes that line up with the oligo(U) runs of the cDNAs indeed contain runs of A; A_6TA_3 in the coxI gene and $A_8T_2A_2$ in the 9S gene. Similar A-rich segments serve as priming sites for first strand 9S and 12S cDNA synthesis. Evidently, edited regions such as that found at the 5′ end of MURF2 and other RNAs also produce A-rich cDNA strands. It remains to be established, therefore, whether long runs of U occur in RNA editing intermediates. The construction of cDNA libraries without the use of oligo(dT) should provide a more definite answer.

The other deviating sequences seem to be produced by the insertion/deletion of Us at sites that are unedited in the bulk of the RNA. This phenomenon has also been observed for other transcripts (reviewed in Simpson and Shaw, 1989). They could indeed be derived from intermediary stage RNAs that arise during the editing process and are later corrected into the mature sequence. Alternatively, they could be side products of a (slightly) imprecise RNA editing machinery, and in some cases even be translatable (MURF2, coxIII). An *in vitro* RNA editing system, in which a precursor–product relation between these products and the mature RNAs can be established, is required to settle this point.

Materials and methods

Cell culture, nucleic acid isolation

Crithidia 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).

Table I.

Oligonucleotides	Gene	Coordinates and references
C67 TTAACTATTCCTAAACATAG	9S	1559–1540[a]
C26 ACCGCACAAACAAAACAC	coxIII	4409–4392[a]
C28 CGCAAATTCAACGCACAC	coxIII	4667–4650[a]
C29 TATACCCAATAATACATGAC	cox III	5003–4984[a]
C51 AGCGCTAAAATCCGTCGCTC	NDI	106–125[b]
C26 CAACTAATATGTCAACGAT	NDI	158–176[b]
C21 AGGCTGTAAAATCTC	coxII	373–259[b]
C22 CGCCTATGTCTTTTG	coxII	436–450[b]
C23 CTTAGGTATCAAGGTA	coxII	696–711[b]
C37 GTCGTATGTATAAC	MURF2	969–982[b]
C46 GTTATACATACGACAAAGTC	MURF2	982–963[b]
C45 TCCAGTCACCTAACAACTTT	MURF2	1233–1252[b]
C49 CAAATAAACAGCGAATTACA	MURF2	1368–1349[b]
C58 ATGATATATACACGGAGTTG	MURF2	1632–1651[b]
C50 AATACAAAGATAAAGAACAC	MURF2	1723–1704[b]
C57 AATAGCGAATAAACATATCC	coxI	110–91 with respect to A in ATG)

[a]Sloof *et al.* (1987).
[b]Van der Spek *et al.* (1989).

Mitochondrial DNA sequence analysis

The cloning and sequencing of *C.fasciculata* mtDNA fragments has been described: a 5.3 kb *SalI–Hin*dIII fragment (S2-D1; Hoeijmakers *et al.*, 1982) containing the rRNA (Sloof *et al.*, 1985) and MURF3 and coxIII genes (Sloof *et al.*, 1987); a 2.0 kb *Hin*dIII–*Eco*RI (D3-R3; Hoeijmakers *et al.*, 1982) fragment containing ~210 5′ terminal nucleotides of the NDI gene, the coxII and MURF2 genes and ~150 3′ terminal nucleotides of the coxI gene (Van der Spek *et al.*, 1989). The partial gene and restriction maps of these fragments are shown in Figure 2. Oligonucleotide C57 was used to determine the sequence of ~120 5′ terminal nucleotides of the coxI gene with a cloned R3-R4 (Hoeijmakers *et al.*, 1982) fragment.

cDNA cloning and sequence analysis

The construction of a *C.fasciculata* cDNA library, using oligo(dT) to prime first-strand synthesis, has been described in Van der Spek *et al.* (1988). Clones were obtained by screening with the corresponding genomic fragments: R1-D1 (Figure 2), for coxIII cDNA; R6-E2 (E = *Hae*III; Hoeijmakers *et al.*, 1982), for 9S and 12S cDNAs; and D3-R3 for NDI, coxII, MURF2 and coxI cDNAs. Nucleotide sequence analysis was performed using primer extension as described in Sanger *et al.* (1977), making use of commercially available pUC primers. In addition, the oligonucleotides in Table I were synthesized and used.

The area of the cDNAs sequenced is indicated in Figure 2. It is composed from the analysis of a large number of different cDNAs. Between 200 and 450 nt of (single strand) 5′ and 3′ end sequence are determined for numerous clones. One edited NDI, coxII, coxIII and MURF2 cDNA and one unedited NDI cDNA have been completely sequenced on one strand. Putative edited regions were sequenced on two strands in a large number of independently picked cDNAs, as indicated in Figure 2. Sequences obtained were compared to that of *C.fasciculata* genomic fragments using a VAX computer. In one case (NDI) genomic sequences from the *T.brucei* maxicircle (Hensgens *et al.*, 1984) were used, since the 3′ terminal moiety of the NDI gene is not sequenced in *C.fasciculata*. The *C.fasciculata* cDNA and *T.brucei* genomic NDI sequences encode highly similar proteins of the same length (313 amino acids). One hundred and ninety of these amino acids are identical, 39 represent a conservative substitution. The sequences determined of the 12S and 9S cDNAs correspond to nt 33–918 of 12S rRNA (which is 1150 nt long; Sloof *et al.*, 1985) and to nt 20–440 of the 611 nt long 9S rRNA. Four independently picked 12S and seven 9S clones were sequenced. The 120 5′ terminal and 150 3′ terminal nucleotides of coxI cDNA have been determined in eight clones.

Oligonucleotides C67, C51 and C57 were used in primer extension assays with AMV reverse transcriptase (1 U) and *C.fasciculata* total cellular RNA (15–50 μg) essentially as described in Tabak *et al.* (1984), to obtain sequences of 9S, NDI and coxI RNAs respectively.

Acknowledgements

We are grateful to Professors P.Borst and H.F.Tabak for stimulating discussions, their continuing interest and comments on the manuscript. We thank Ms D.Evans for her contribution in the 9S and 12S cDNA sequence analysis and Ms G.E.E.van Noppen and Mrs J.E.Vlugt-van Daalen for expert handling of the manuscript.

References

Abraham,J.M,. Feagin,J.E. and Stuart,K. (1988) *Cell*, **55**, 267–272.
Anderson,S., Bankier,A.T., Barrell,B.G., De Bruyn,M.H.L., Coulson, A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F., Schreier,P.H., Smith,A.J.H., Staden,R. and Young,I.G. (1981) *Nature*, **290**, 457–465.
Benne,R. (1985) *Trends Genet.*, **1**, 117–121.
Benne,R. (1989) *Biochim. Biophys. Acta*, **1007**, 131–139.
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.
De la Cruz,V., Neckelmann,N. and Simpson,L. (1984) *J. Biol. Chem.*, **259**, 15136–15147.
De la Cruz,V.F., Lake,J.A., Simpson,A.M. and Simpson,L. (1985a) *Proc. Natl. Acad. Sci. USA*, **82**, 1401–1405.
De la Cruz,V., Simpson,A., Lake,J.A. and Simpson,L. (1985b) *Nucleic Acids Res.*, **13**, 2337–2356.
Feagin,J.E., Jasmer,D.P. and Stuart,K. (1985) *Nucleic Acids Res.*, **13**, 4577–4596.
Feagin,J.E., Jasmer,D.P. and Stuart,K. (1986) *Mol. Biochem. Parasitol.*, **20**, 207–214.
Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *Cell*, **49**, 337–345.
Feagin,J.E., Shaw,J.M., Simpson,L. and Stuart,K. (1988a) *Proc. Natl. Acad. Sci. USA*, **85**, 539–543.
Feagin,J.E. and Stuart,K. (1988b) *Mol. Cell. Biol.*, **8**, 1259–1265.
Feagin,J.E., Abraham,J.M. and Stuart,K. (1988c) *Cell*, **53**, 413–422.
Fearnly,I.M., Runswick,M.J. and Walker,J.E. (1989) *EMBO J.*, **8**, 665–672.
Hensgens,L.A.M., Brakenhoff,J., De Vries,B.F., Sloof,P., Tromp,M., Van Boom,J.H. and Benne,R. (1984) *Nucleic Acids Res.*, **12**, 7327–7344.
Hoeijmakers,J.H.J., Schoutsen,B. and Borst,P. (1982) *Plasmid*, **7**, 199–209.
Hoeijmakers,J.H.J., Snijders,A., Jansen,J.W.G. and Borst,P. (1981) *Plasmid*, **5**, 329–350.
Kleisen,C.M., Borst,P. and Weijers,P.J. (1975) *Biochim. Biophys. Acta*, **390**, 155–167.
Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.
Maizels,N. and Weiner,A. (1988) *Nature*, **334**, 469–470.
Peabody,D. (1989) *J. Biol. Chem.*, **264**, 5031–5035.
Pritchard,A.E., Seilhamer,J.J. and Cummings,D.J. (1986) *Gene*, **44**, 243–253.
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. (1987) *Annu. Rev. Microbiol.*, **41**, 363–382.
Simspon,L. and Shaw,J. (1989) *Cell*, **57**, 355–366.
Simpson,A.M., Neckelmann,N., De la Cruz,V., Mishick,M.L. and Simpson,L. (1985) *Nucleic Acids Res.*, **17**, 5977–5993.
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.
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.
Tabak,H.F., Van der Horst,G., Osinga,K.A. and Arnberg,A.C. (1984) *Cell*, **39**, 623–629.
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.

Received on September 7, 1989; revised on October 26, 1989