The EMBO Journal vol.12 no.4 pp.1523-1532, 1993

Implications of novel guide RNA features for the mechanism of RNA editing in *Crithidia fasciculata*

Gert Jan Arts, Hans van der Spek[1], Dave Speijer, Janny van den Burg, Harry van Steeg[2], Paul Sloof and Rob Benne

E.C.Slater Institute, University of Amsterdam, Academic Medical Centre, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands and [2]National Institute of Public Health and Environmental Protection (RIVM), PO Box 1, 3720 BA, Bilthoven, The Netherlands
[1]Present address: Institut de Biologie Moleculaire des Plantes, Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France

Communicated by L.Grivell

We have determined the relative steady state concentration of the two *Crithidia fasciculata* guide (g)RNAs involved in editing the two domains of mRNAs for NADH dehydrogenase (ND) subunit 7. We found that, although there was an 8-fold difference between the molar ratio of these two gRNAs relative to the (pre)-mRNA, the two domains are edited with a very similar frequency (around 50%). Also, for the editing of a given domain, many gRNA species exist with the same 5′ end but with a different 3′ uridylation site. Approximately 20% of these short gRNAs do not contain the information required for editing a complete domain, which may explain the high incidence of partially edited RNAs. Remarkably, genomically encoded Us are missing from two sites of a few of the gRNAs involved in editing apocytochrome *b* RNA. We speculate that these species are created by editing-like events. Both the short and complete forms of the ND7 gRNAs are found in chimeric molecules, in which the gRNA is covalently linked via its 3′-terminus to an editing site of pre-edited ND7 RNA. Some features of the chimeric molecules are at odds with current models of RNA editing: (i) U residues are completely absent from the connecting sequence of a number of these molecules, (ii) the ND7 gRNAs are frequently hooked up to the wrong editing domain of ND7 RNA, although other gRNAs are not found at these positions and (iii) in some chimeric molecules the gRNA appears to be linked to the 5′ end of pre-edited RNA.
Key words: guide RNA/mitochondrion/RNA editing/ trypanosomes

Introduction

Trypanosomal mitochondrial (mt) RNAs are post-transcriptionally edited by insertion/deletion of U residues (reviewed in Simpson and Shaw, 1989; Benne, 1990; Stuart, 1991b). The information for this remarkable form of RNA processing is provided by small so-called guide (g)RNAs, which are complementary to edited RNA segments, if G:U base-pairing is allowed (Blum *et al.*, 1990). gRNAs are encoded both in the maxicircle and in the minicircle component of the trypanosome mt DNA (Koslowsky *et al.*, 1990; Pollard *et al.*, 1990; Sturm and Simpson, 1990b, 1991; Van der Spek *et al.*, 1991; Maslov and Simpson, 1992) and their involvement in the editing process is supported by the fact that both the genomic location and the potential to base-pair with edited RNA have been conserved between the maxicircle-encoded gRNAs of *Leishmania tarentolae* and *Crithidia fasciculata* (Van der Spek *et al.*, 1991). Three different gRNA regions can be distinguished on the basis of their putative functioning in the editing process: (i) an anchor sequence located in the 5′ region of the gRNA which is complementary to a target sequence on pre-edited RNA downstream of an editing domain, (ii) a middle part containing the editing information, and (iii) a 3′ terminal oligo(U) extension (Blum and Simpson, 1990; Pollard and Hajduk, 1991). The reported size heterogeneity of gRNAs (55–70 nt) is reflected in the variation in size of all three regions: anchors vary between 4 and 14 nt, the informational part varies from guiding the insertion of just one U to that of 30–40 Us at 10 sites or more, and the U-tail ranges from 5 to 24 nt [for a review see Benne (1992)]. The editing machinery is apparently able to cope with such a diverse collection of gRNAs. The question arises, however, whether or not the differences in properties between gRNAs influence the efficiency with which the corresponding domains are edited.

The precise mechanism via which gRNAs act in the editing process is unknown. After initial anchoring of the pre-edited RNA by the gRNA, editing may proceed either by biased 'mismatch repair' (Blum *et al.*, 1990, 1991) or by a 'match protection' mechanism (Decker and Sollner-Webb, 1990), which extends the base-paired region of the gRNA:mRNA duplex via the insertion and deletion of uridines into/from the pre-edited RNA (reviewed in Stuart, 1991a). The overall direction of the editing process seems to be from 3′ to 5′ for most extensively edited RNAs, whereby editing directed by the first (most 3′) gRNA creates the anchor sequence with which the next gRNA can interact (Abraham *et al.*, 1988; Bhat *et al.*, 1990; Maslov and Simpson, 1992; Souza *et al.*, 1992). For some RNAs, however, different domains exist which are independently edited (Koslowsky *et al.*, 1990; Maslov and Simpson, 1992). The characteristics of some partially edited apocytochrome *b* (CYb) RNAs from *L.tarentolae* suggested that a 3′ to 5′ polarity is also maintained within the domain of a single gRNA, editing taking place in a stepwise fashion, a single site at a time and, within a site, a single nucleotide at a time (Sturm and Simpson, 1990a). The finding of other, much less neatly ordered collections of partially and even abnormally edited transcripts, however, argued against such a strict 3′ to 5′ polarity and stepwise processivity (Abraham *et al.*, 1988; Bhat *et al.*, 1990; Decker and Sollner-Webb, 1990; Koslowsky *et al.*, 1991; Souza *et al.*, 1992). The occurrence *in vivo* (Blum *et al.*, 1991; Read *et al.*, 1992; Sturm *et al.*, 1992), and the possibility of synthesis *in vitro* (Harris and

Hajduk, 1992; Koslowsky *et al.*, 1992a), of chimeric molecules in which a gRNA is covalently linked through its U-tail to an editing site of a pre-edited RNA, provides convincing evidence for the participation of gRNAs in the editing process, not only as a source of genetic information, but also as a source of the Us themselves. The chimeric molecules may be formed in one-step RNA-mediated transesterification reactions in analogy with splicing reactions (Blum *et al.*, 1991; Cech, 1991) or, alternatively, in a two-step 'cut and paste' enzyme-mediated process (Sollner-Webb, 1992). Although direct evidence (e.g. from kinetic experiments) that these molecules are intermediates of the editing process is lacking *in vivo*, their consistent presence in mt RNA and the relative ease with which they can be produced *in vitro* suggest that they indeed play such a role.

The studies reported in this paper were initiated to determine whether or not gRNAs recycle during the editing process in *C.fasciculata*. In the course of this work we noticed a remarkable size heterogeneity of a gRNA involved in the editing of the mRNA for NADH dehydrogenase (ND) subunit 7, which could not be explained by the available data. Northern blot analysis and PCR cloning of a number of gRNAs shows that the uridylation site of a gRNA is highly variable, which results in shorter 'truncated' gRNA versions in which the U-tail may be found attached to virtually every nucleotide. The properties of these gRNAs and the chimeric molecules in which they occur are described in this paper and discussed in relation to current models of RNA editing.

Results

The frequency of editing of the ND7 domains does not correlate with gRNA concentration

ND7 mRNA in *C.fasciculata* contains two editing domains, one situated at the 5′ end, the [5′] domain, and the other at an RNA-internal position, the frameshift [FS] domain (Van der Spek *et al.*, 1988, see Figure 6A). These domains are separated by a stretch of 172 nt and are edited by separate gRNAs by insertion of 22 and 5 Us, respectively (Van der Spek *et al.*, 1988). ND7 RNA, therefore, provides the simplest system in which the regulation and coordination (if any) of editing of one RNA by multiple gRNAs can be studied. In order to determine the frequency at which the two ND7 domains are edited, the primer extension method first described by Shaw *et al.* (1989) was applied (Figure 1). In this procedure, oligonucleotide primers complementary to a region immediately downstream of the editing domains are extended with reverse transcriptase in a mixture from which dGTP has been omitted and in which ddGTP is present. Extended products end at a position corresponding to the first C in the RNA, which in both cases is beyond the editing domain. This gives rise to a difference in length of 22 and 5 nt, respectively, between products derived from edited and unedited [5′] and [FS] domains. Quantification of the signals found at these positions allows calculation of the frequency of editing of these two domains in mitochondrial ND7 RNA. The results given in Figure 1C show that the frequency of editing of the two domains is very similar (54% versus 51% for the [5′] and [FS] domains, respectively). These numbers are in good agreement with those obtained from Northern blot analysis (Van der Spek *et al.*, 1988).

We next determined the steady-state concentration of all

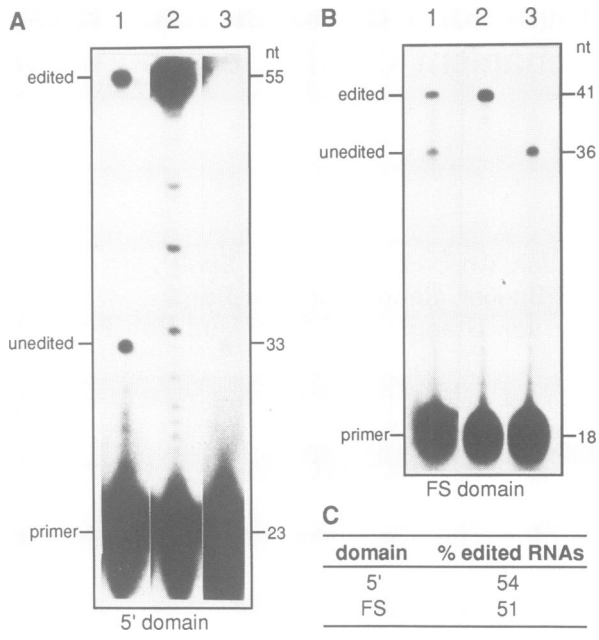
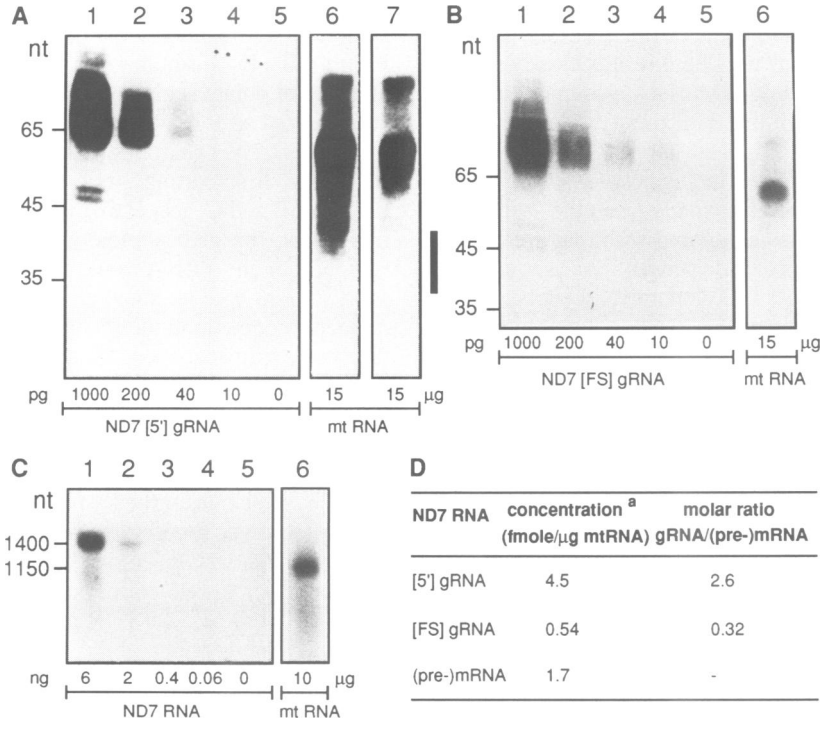


Fig. 1. Extent of editing of ND7 domains. (**A** and **B**) Primer extension in the presence of ddGTP and absence of dGTP. Reactions were performed with 2 μg of mt RNA (lanes 1) or 10 ng of *in vitro* synthesized edited (lanes 2) or unedited (lanes 3) ND7 (pre-)mRNA, respectively. The differences in size of the extension products of primer C42 (A) and primer C30 (B) correspond to the absence or presence of 22 and 5 inserted Us in the [5′] and [FS] editing domain respectively. (**C**) Frequency of editing derived from the relative intensities of the edited and unedited extension products of lanes 1 (A and B). The numbers given are an average obtained from six different mt RNA preparations, the variation being 54 ± 13% for the [5′] domain and 51 ± 12% for the [FS] domain. The values for the RNA preparation used in the figure are 65% and 60%, respectively.

the RNA components involved. For this purpose, we quantified the signals which had been obtained by hybridizing blotted mt RNA to (i) oligonucleotide probes complementary to the 5′ terminal region of the gRNAs (Figure 2, lanes A6 and B6), and (ii) to an ND7 cDNA probe (lane C6), using a PhosphorImager apparatus. Subsequently the signals were compared with the signals obtained with known quantities of gRNA and mRNA synthesized *in vitro* (Figure 2A–C, lanes 1–5). The calculated steady-state concentrations and the molar ratio at which the gRNAs were present with respect to ND7 RNA (unedited plus edited) given in Figure 1D show that the [5′] gRNA was present at an ~8-fold higher concentration than the [FS] gRNA, in spite of the similar frequency of editing of the two domains.

The uridylation site of gRNAs is highly variable

The hybridization signal of ND7 [5′] gRNA and to a lesser extent that of the [FS] gRNA suggest a considerable size heterogeneity (Figure 2). This appears to be a general property of *C.fasciculata* gRNAs since MURF2-II and CYb-II gRNAs also display these characteristics (Figure 3). The size of some of the hybridizing gRNA species is smaller than that predicted on the basis of the corresponding maxicircle gene sequences, particularly for the ND7 [5′] and CYb-II gRNAs (see Table I). This discrepancy is even more pronounced if one assumes that gRNAs have a 3′ terminal non-genomically encoded oligo(U) tail (Blum and Simpson, 1990; Pollard and Hajduk, 1991). The gRNA size heterogeneity is not caused by random degradation or



D

ND7 RNA	concentration [a] (fmole/µg mtRNA)	molar ratio gRNA/(pre-)mRNA
[5'] gRNA	4.5	2.6
[FS] gRNA	0.54	0.32
(pre-)mRNA	1.7	-

Fig. 2. Quantification of ND7 (pre-)mRNA and both ND7 gRNAs. (**A**) ND7 [5′] gRNA. A Northern blot of a 14% polyacrylamide gel was obtained as described in Materials and methods and probed for ND7 [5′] gRNA. Lanes 1–5 contain 15 µg of total RNA of *C.fasciculata* mixed with the indicated amounts of *in vitro* synthesized ND7 [5′] gRNA. Lanes 1–6 are hybridized with oligo C100, lane 7 with C80, recognizing 5′ and 3′ sequences of this gRNA, respectively. (**B**) ND7 [FS] gRNA. Northern blot, essentially as A, but in lanes 1–5 ND7 [FS] gRNA was added. Lanes 1–6 were probed with C99 recognizing 5′ sequences of ND7 [FS] gRNA. (**C**) ND7 (pre-)mRNA. A Northern blot of a glyoxal agarose gel was probed with ND7 cDNA. In lanes 1–5 indicated amounts of synthetic ND7 pre-mRNA were added to 10 µg of total RNA. The size difference with mt ND7 RNA is caused by the fact that a cDNA clone with a 200 nt 5′ terminal extension was used for *in vitro* RNA production. The intensities of the signals were analysed as described in Materials and methods. (**D**) Amounts and molar ratio to ND7 RNA (edited plus unedited) of the ND7 gRNAs. [a], the numbers given are an average obtained from four different mt RNA preparations for the [5′] gRNA and the (pre-)mRNA and two RNA preparations for the [FS] gRNA. The variation was 4.5 ± 3.4 fmol/µg mt RNA for the [5′] gRNA, 0.54 ± 0.17 fmol/µg for the [FS] gRNA and 1.7 ± 0.6 fmol/µg for the (pre-)mRNA. For the calculation of the molar ratios, the average values were used.

heterogeneous 5′ termini, given that all gRNAs tested via primer-extension analysis have fairly discrete 5′ ends, although in some cases microheterogeneity of maximally 2 or 3 nt is observed (ND7 [5′], CYb-II, Van der Spek *et al.*, 1991; MURF2-II, results not shown). For most gRNAs, the heterogeneity also appears to exceed greatly the size variation reported to exist for the oligo(U) tail (5–24 nt, Blum and Simpson, 1990; Pollard and Hajduk, 1991).

In order to shed more light on this phenomenon we prepared cDNA of a number of gRNAs with the aid of an oligo(dA) primer and amplified this by PCR with gRNA-specific oligonucleotides. The amplified material was subsequently cloned and sequenced; a summary of the sequences obtained is shown in Figure 4. The most striking aspect of this analysis is the fact that for three of the four gRNAs studied, the U-tail is found to be attached to many different, apparently random, positions. Only the ND7 [FS] gRNA has a clearly preferred oligo(U) attachment point. The variation in uridylation site is not an artefact of the PCR procedure used, since in control experiments with ND7 [5′] gRNA synthesized *in vitro* as a substrate, all of the clones displayed the U-attachment site of the input RNA (see legend to Figure 4 for details).

In order to determine the concentration of the gRNAs with incomplete editing information more precisely, we performed an experiment with an oligonucleotide complementary to the information domain of ND7 [5′] gRNA (oligonucleotide C 80). The results are shown in Figure 2A, lane 7. Indeed, the smallest gRNA species hybridizing to the 5′ oligonucleotide (lane 6, indicated by a bar in the figure) do not hybridize to oligonucleotide C 80. Quantification of these results with the procedures described for the experiment shown in Figure 2 demonstrates that the 'full-length' ND7 [5′] gRNA species constitute

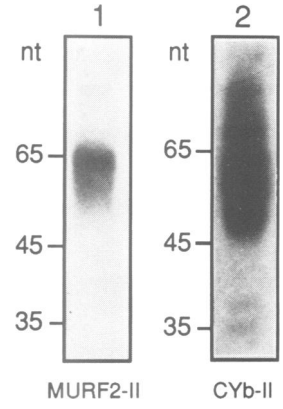


Fig. 3. Size heterogeneity of MURF2-II and CYb-II gRNAs. Northern blots of a 14% polyacrylamide gel were hybridized with oligonucleotides C104 (lane 1) and C106 (lane 2), recognizing 5′ sequences of MURF2-II gRNA and CYb-II gRNA, respectively.

~80% of the total amount of gRNAs present. Similar calculations with similar results could be made for MURF2-II and CYb-II gRNAs (not shown). These results clearly show that a sizeable portion of the gRNAs lack a segment of the informational part, although the PCR procedures followed in the experiments shown in Figure 4 appear to have created some bias in favour of smaller molecules.

An unexpected observation was made when the sequence of CYb-II gRNA clones was compared with that predicted by the corresponding maxicircle genes. As shown in Figure 5, the longest gRNAs lack genomically encoded Us at two different sites. This opens the exciting possibility that gRNAs themselves are also subject to some form of editing (see Discussion).

Chimeric molecules

We next wanted to determine to what extent the shorter gRNAs occur in so-called chimeric molecules, in which a gRNA is covalently linked to an editing site of pre-edited RNA through its U-tail. We therefore designed the PCR experiments outlined in Figure 6A with a common 3′ oligonucleotide primer downstream of the FS editing site and either a [5′] or a [FS] gRNA-specific oligonucleotide as 5′ primer. In this way we could also check the possible formation of chimeric molecules at unusual positions, e.g. outside the editing domains. The sequence of a representative sample of the chimeric molecules we cloned is shown in Figure 6B–E, a summary of the location of the gRNA attachment points is given in Figure 6A. The results can be summarized as follows: (i) in the majority of the chimeric molecules (80%) we found the gRNA ([5′] or [FS]) hooked up to the first editing site of its corresponding domain (Figure 6B and C, respectively); (ii) in virtually all of these molecules, the gRNA moiety is not full-length in the 3′ region, i.e. the gRNA could be derived from one of the shorter forms with an 'internal' uridylation site; (iii) chimeric molecules containing the [5′] gRNA have only a few Us in the connecting sequence, insufficient for the editing of the site which would require at least seven Us, and sometimes the Us are completely missing. The [FS] chimeric molecules have more Us (~10). Also in this case, however, we found chimeric molecules without Us; (iv) a number of unexpected chimeric molecules turned up in which either the [5′] or the [FS] gRNA was found to be hooked up to a position in, or close to, the wrong editing domain (Figure 6D and E, respectively). In two chimeric molecules, clones 25.2 and 26.2 in Figure 6E, the [FS] gRNA was found linked to a position 30 nt upstream of the (unedited) [5′] editing domain of ND7 RNA. Primer-extension analysis of unedited mt ND7 RNA revealed a strong stop at this position indicative of the existence of ND7 RNA species with a 5′ end at this position (see Figure 7A, arrow). Such strong stops were not observed with synthetic ND7 RNA, produced *in vitro*, as a template (Figure 7B).

A discussion of the possible way in which the chimeric molecules may be formed and the relevance of these molecules for the mechanism of RNA editing will be presented in the next section. They probably do not represent non-specific ligation products, since chimeric molecules between the ND7 pre-edited RNA and other gRNAs

Table I. Size variation of gRNAs

gRNA	Gene encoded (nt)	Northern (nt)
ND7[FS]	28	75–50
ND7[5′]	44	80–32
MURF2-II	52	83–54
CYb-II	53	85–37

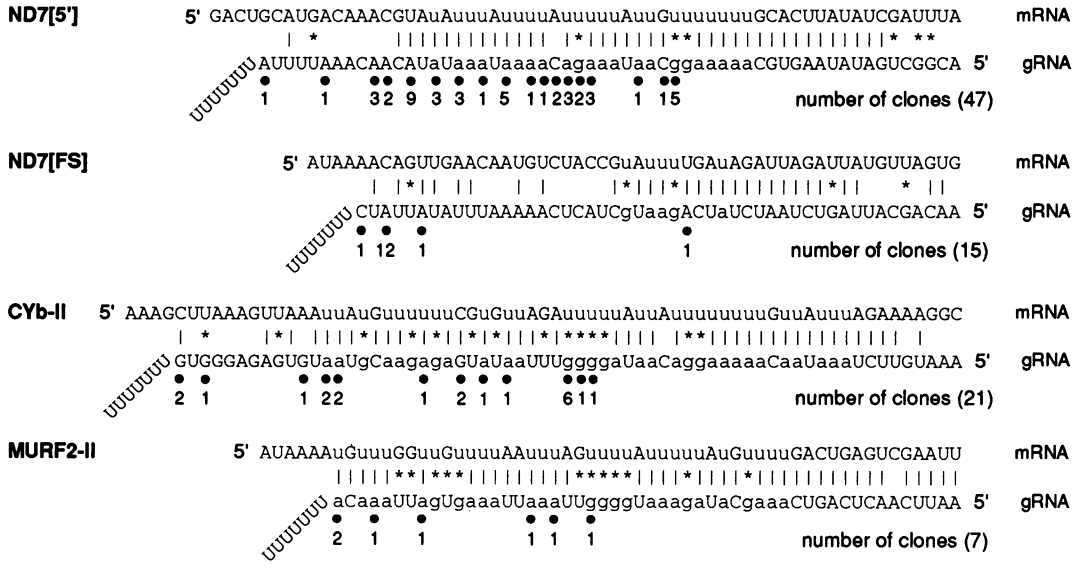


Fig. 4. Variable uridylation sites of four gRNAs. Alignments of edited mRNA sequences (upper line) and the appropriate gRNA sequences (lower line) are shown. PCR amplification was performed with a 3′ T_4A_{20} primer and with 5′ primers specific for the following gRNAs: ND7 [5′], C91; ND7 [FS], C90; CYb-II, C108; and MURF2-II, C107. cDNA clones were found with different uridylation sites, indicated with dots below the gRNA sequences. The number of clones found for each site and the total number are indicated. In control experiments, *in vitro* synthesized ND7 [5′] gRNA was subjected to the same amplification procedure. The length of the U(T)-tail of the cDNA clone used for the *in vitro* synthesis was 15 nt. The synthetic gRNA displayed a variation in length of the U-tail very similar to that of natural ND7 [5′] gRNA (compare lanes 1 and 7, Figure 1), most likely due to sliding and stuttering of the T7 polymerase. All of the amplified cDNA clones, however, had the original uridylation site (= nt 48 of the gRNA).

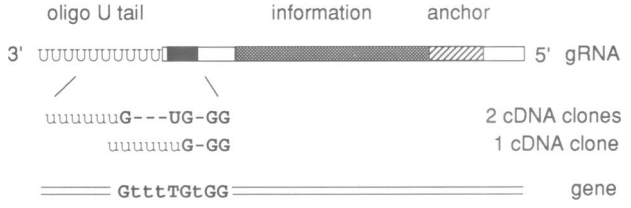


Fig. 5. 3′ end sequence heterogeneity of CYb-II gRNAs. The anchor and the information region of the gRNA are indicated with different hatches, the oligo(U) tail with a stretch of Us. The 3′ end sequence of a few CYb-II gRNA cDNA clones is compared with the CYb-II gRNA gene sequence. The remainder of the gRNA sequence is identical to that of the gene.

(MURF2-II and CYb-II) and between ND7 gRNAs and other pre-edited RNAs (ND8) could not be detected in numerous control experiments.

Discussion

gRNA concentration and frequency of editing

In this paper we report on some novel features of gRNAs that are relevant to their possible functioning in the RNA editing process. We have determined the relative steady-state concentration of the two gRNAs involved in the editing of the two ND7 domains, and we showed that these gRNAs are present in relatively high amounts, comparable to those of their target pre-edited RNAs (Figure 2). Since gRNAs and (pre-)mRNA appear to be equally stable during the preparation of mitochondria, we assume that the ratio in which they are present in mt RNA, reflects the *in vivo* situation. Catalysts and their substrates are usually not present in equimolar amounts, the most likely scenario therefore is that the gRNAs operate in a more or less stoichiometric fashion, recycling only a few times at most. As far as we are aware, our results provide the first concrete data on gRNA concentration, since little information can be extracted from the studies on editing in other trypanosome species. It has been reported that some gRNAs in *Trypanosoma brucei* are constitutively present, irrespective of the extent of editing of their target RNAs in different life-cycle stages (Koslowsky *et al.*, 1992b). In *L.tarentolae*, the hybridization signals obtained with different gRNAs involved in the editing of the pan-edited MURF4 and RPS12 RNAs vary substantially, suggesting a role for some of them in regulating the editing process (Maslov and Simpson, 1992). In both these reports, however, the hybridization signals were not quantified, so the amounts of gRNA that are actually present are unknown.

Assuming that the ND7 gRNAs do indeed act in a virtually stoichiometric fashion, their concentration could be an important factor in regulating the extent of editing of the corresponding domains. Different gRNAs may be used, however, with different efficiencies depending, for example, on the stability of the anchor duplex. For the ND7 gRNAs, these are clearly not the only contributing factors, considering the fact that in spite of the 8-fold lower concentration of the [FS] gRNA and the lower thermodynamic stability of the anchor duplex, the extent of editing of the [FS] ND7 domain is not greatly dissimilar to that of the [5′] domain. Analysis of ND7 cDNAs (Van der Spek *et al.*, 1988) and chimeric molecules (Figure 6, this paper) revealed that all (16) RNAs in which the [5′] domain is edited, or is being edited, have an edited [FS] domain. This strict coupling of the editing of these two domains is unexpected in view of the large region of (172) unedited nucleotides between them. This might suggest that an overall 3′ to 5′ direction of the editing process is also a characteristic of less extensively edited RNAs for which gRNAs are used that do not create each other's anchor sequences. Although the mechanism which prevents independent editing of the 5′ domain is unclear, such a scheme could explain the lack of correlation between gRNA concentration and frequency of editing of the ND7 domains.

3′ end heterogeneity of gRNAs

We also report on the variation in the U-tail attachment site of gRNAs, yet another unexpected source of gRNA size heterogeneity. Of the four gRNAs in *C.fasciculata* that we have looked at, three show essentially random uridylation sites. Only the ND7 [FS] gRNA has a clearly preferred U-tail attachment point (see Figure 4). Truncated gRNA moieties have been found in chimeric molecules in *L.tarentolae* (Blum *et al.*, 1991) and *T.brucei* (Koslowsky *et al.*, 1991; Read *et al.*, 1992), but, particularly in *L.tarentolae*, both the extent of truncation and the incidence in which they occurred was much less pronounced. Our results in fact provide the first evidence that these shorter gRNA molecules indeed comprise a substantial fraction (~20%) of the free mt gRNA population. As shown in the experiments of Figures 2 and 3 and Table I, they are clearly visible on Northern blots and are not artefactually produced in the PCR procedure. In principle, (truncated) gRNA moieties of chimeric molecules that possess a stretch of Us in the connecting sequence, could give rise to some of the sequences in our gRNA clone collection, since the cDNA synthesis step was primed with oligo(dA). Chimeric molecules are not an important contributing factor, however, because the PCR results were essentially the same when, instead of total mt RNA, a small RNA fraction isolated from agarose gels was used as a source of gRNA. A special case of gRNA 3′ end sequence heterogeneity is provided by the CYb-II gRNA species, the sequence of which is shown in Figure 5. In two of the clones four genomically encoded Us appear to be missing, reminiscent of U-deletion events in mRNAs. Also in this case, trivial artefactual explanations for their existence can be excluded, given the fact that these sequences were found in a number of independently picked clones from different PCR experiments.

How do these deviating 3′ end sequences of gRNA arise? Alternative gene versions, encoding overlapping gRNAs, have been shown to exist for some gRNAs in *T.brucei* (Pollard *et al.*, 1990; Koslowsky *et al.*, 1992b), but extensive Southern blot analysis of mt DNA of *C.fasciculata* with oligonucleotides complementary to CYb-II gRNA as a probe reveal that only one CYb-II gRNA gene is present (results not shown; Van der Spek *et al.*, 1991). This suggests that the deviating sequences are generated co- or post-transcriptionally. Very little is known about gRNA transcription in trypanosome mitochondria (for a review see Benne, 1992). In principle, truncated gRNAs could arise in some form of sloppy transcription termination/uridylation process. Similarly, the polyadenylation site of mRNAs is also found to be variable (Van der Spek *et al.*, 1988, 1990; Souza *et al.*, 1992). Clear consensus sequences that could serve as signals for the enzymatic machinery indeed appear to be absent. Truncation could also be the result of the combined action of nucleases and terminal uridylyl

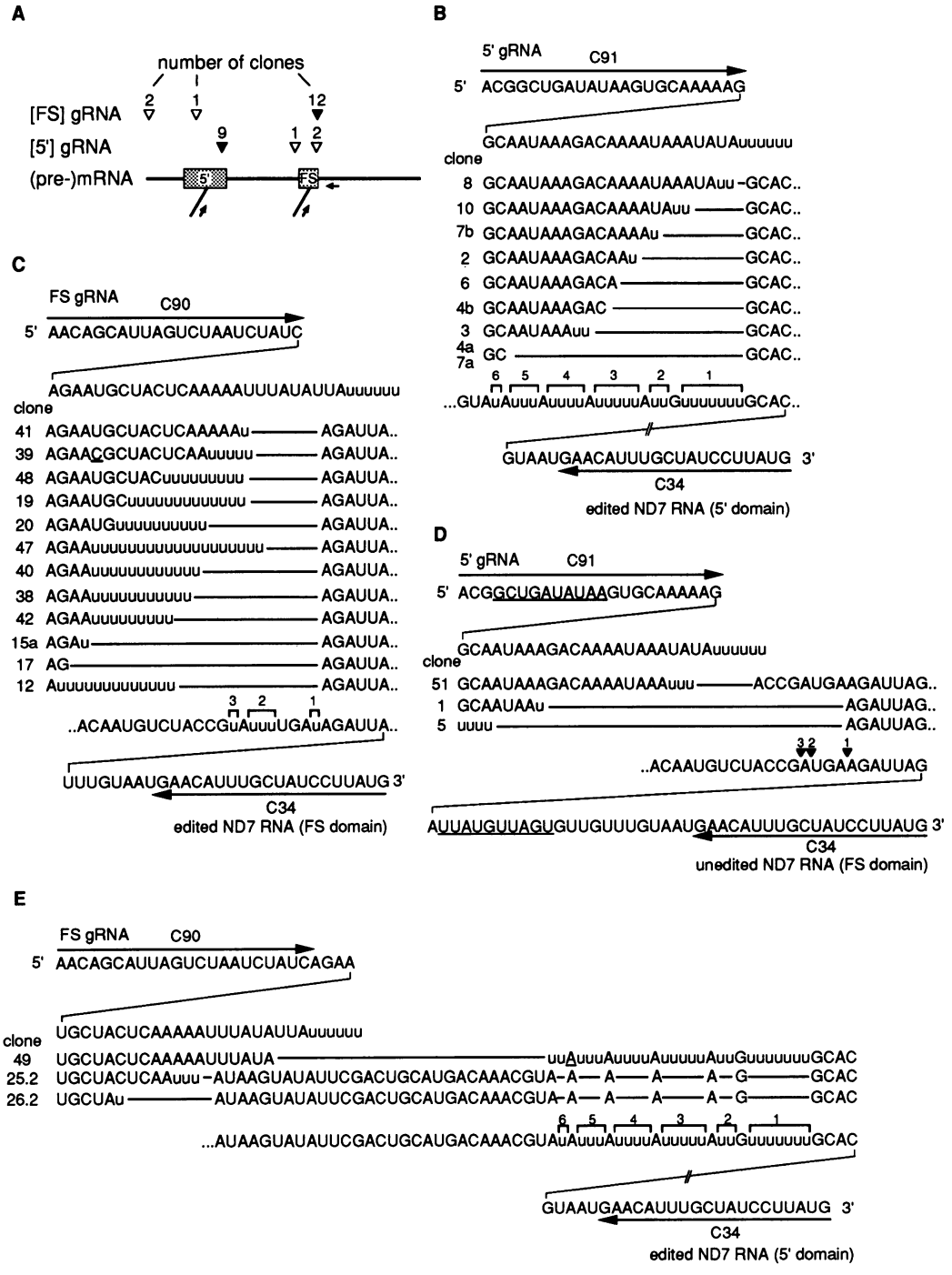


Fig. 6. Chimeric ND7 gRNA−mRNA molecules. (**A**) Schematic representation of the collection of chimeric molecules. Arrows mark the positions of the primers used in the PCR amplification and the number of clones analysed is given. The attachment point of the ND7 gRNAs to ND7 (pre)-mRNA is indicated with triangles. In nine clones ND7 [5′] gRNA is linked to the first editing site of the [5′] domain of ND7 pre)-RNA (stippled); in 12 clones ND7 [FS] gRNA is linked to the first editing site of the [FS] domain. These are 'expected' chimeras and are indicated with black triangles. No chimeras were found in which the gRNAs were linked to other editing sites of the corresponding domains. In 20−30% of the chimeric molecules the ND7 gRNAs were hooked up at or near the *wrong* editing domain: unexpected chimeras, open triangles. (**B−E**) Sequences of chimeric molecules. For comparison, the upper line shows the gRNA sequence, the bottom line the (un)edited (pre-)mRNA sequence. The position of the PCR primers is indicated. U residues derived from the oligo(U) tail of gRNAs or inserted into the edited mRNA are shown in lower case. Horizontal lines connect the gRNA- and the pre-mRNA moieties. (B and C) Expected chimeric molecules with ND7 [5′] gRNA and ND7 [FS] gRNA, respectively. In some [FS] chimeric molecules the 5′ end nucleotides are missing, due to the poor quality of one of the oligonucleotide preparations used. Such 5′ truncated nucleotides were no longer present following TLC purification of the oligonucleotide. At one position of clone 39 (underlined) a C was found substituting the expected U. This is probably due to misincorporation during the PCR. (D) Unexpected chimeric molecules with [5′] ND7 gRNA hooked up at or near the ND7 [FS] editing domain. The nucleotides which could form an alternative anchor duplex are underlined. The (unedited) editing sites of the [FS] domain are indicated. (E) Unexpected chimeric molecules with ND7 [FS] gRNA hooked up at or near the ND7 [5′] editing domain. In clone 49, ambiguity exists with respect to the last genomically encoded gRNA nucleotide. It could also be the A that is underlined.

transferase (TUTase), enzymes which have been shown to be present in extracts from *L.tarentolae* mitochondria (Bakalara *et al.*, 1989; Harris *et al.*, 1992; Simpson *et al.*, 1992). It is attractive to assume, however, that both the truncation and the U-deletion at the 3′ end of a gRNA are generated during editing. Since gRNA and pre-mRNA appear to be equal partners in the editing process (Blum *et al.*, 1991), it is likely that unexpected sequences can arise not only in the pre-mRNA moiety (Abraham *et al.*, 1988; Decker and Sollner-Webb, 1990; Sturm and Simpson, 1990a; Sturm *et al.*, 1992) but also in the gRNA. If a chimeric molecule is cleaved, either enzymically or via RNA-mediated transesterification, not at the expected site adjacent (5′) to the anchor duplex (Blum *et al.*, 1991), but rather at a gRNA internal site, and if a U-tail is added to the newly generated gRNA 3′ end by TUTase, a truncated gRNA is created. These secondary sites of cleavage could be selected because they are adjacent to small stretches of intramolecular base-pairing between gRNA nucleotides, serving as 'false' anchors. The other half of the molecule resulting from aberrant cleavage of a chimeric molecule would consist of the 3′ end fragment of the gRNA linked to an editing site of (pre-) mRNA. Such molecules have not been observed, so far, which may indicate that they are rapidly processed or degraded. Since it would be difficult to demonstrate their existence unequivocally with a PCR-type approach, we have not further pursued this point.

Similar events could give rise to the deletion of Us that are internal to the gRNA. As outlined in Figure 8, false anchors can be easily formed by intramolecular base-pairing in CYb-II gRNA, adjacent to a site from which three Us are missing. Cutting at this site (arrow in the figure), followed by deletion of Us and religation analogous to the events proposed to happen when Us are deleted from a pre-mRNA (Blum *et al.*, 1991), would result in the deletion of three Us at this site. The deletion of the single U at the other site would result from similar intramolecular base-pairing between other parts of this gRNA (not shown). Our current work with *in vitro* systems for chimeric molecule formation is aimed at determining whether these mechanisms are indeed operative.

The mechanism of the editing process

The chimeric ND7 molecules that we have detected show some unexpected and, in some cases, novel features. First of all, their formation is not an artefact of the lengthy procedure via which mitochondria are isolated, since PCR experiments with total RNA extracted from living *C.fasciculata* essentially gave the same type of chimeric molecules. In all chimeric molecules of *C.fasciculata* in which a gRNA is hooked up to its corresponding domain, the first editing site is used (producing the 'expected' chimeras, see Figure 6B and C). In chimeric molecules found in other trypanosomes (Blum *et al.*, 1991; Read *et al.*, 1992), the first editing site is also frequently used, but other attachment points are found too and sometimes even preferred. For example, in chimeric molecules for cytochrome oxidase (cox) 3 RNA in *L.tarentolae*, the second editing site is preferred. It remains unclear why different sites predominate in different chimeric molecules. When a gRNA is not hooked up to the first editing site the downstream sites are mostly correctly edited in *L.tarentolae* (Blum *et al.*, 1991), but in chimeric molecules in *T.brucei*,

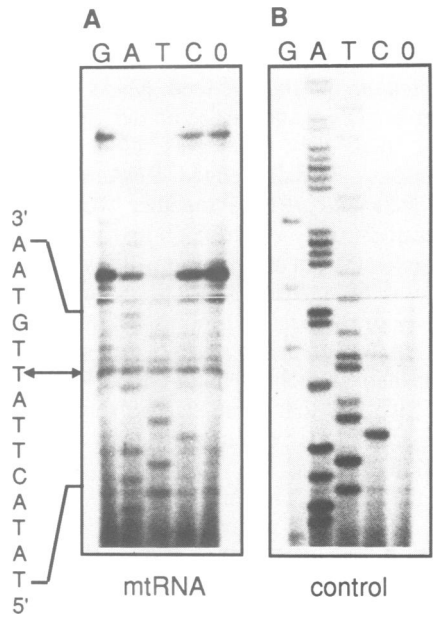


Fig. 7. The 5′ end sequence of unedited ND7 RNA. The sequence was determined with primer C109, as described in Materials and methods. (**A**) 2 μg of mtRNA; (**B**) 10 ng of synthetic unedited ND7 RNA. The strong stop corresponding to the gRNA attachment site of two clones in Figure 6E (25.2 and 26.2) is indicated with an arrow.

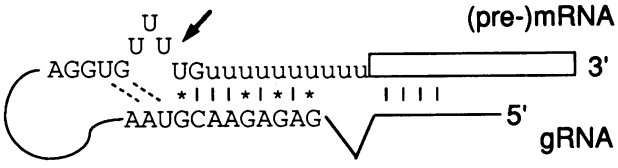


Fig. 8. A possible mechanism to explain the generation of gRNAs with 3′ end heterogeneity. For details, see text.

this is frequently not the case (Read *et al.*, 1992). All this has contributed to differences in opinion with respect to the mechanism of the editing process, which underlines the problems related to the construction of models using molecules obtained via PCR as a source of information. It should be emphasized, therefore, that the functional relevance of chimeric molecules visualized by PCR, including those reported in this paper, must be further assessed in an efficient *in vitro* editing system.

In *C.fasciculata* Us are frequently absent from the connecting sequence between the gRNA and the pre-mRNA part, particularly for the ND7 [5′] chimeras. In this respect, chimeric molecules of *C.fasciculata* look more like those produced *in vitro* with a *T.brucei* mt extract (Koslowsky *et al.*, 1992a). The frequency with which they are found is not dissimilar to that of the molecules that do conform to current models, which suggests that also *in vivo*, Us are not required *per se* in the reactions that produce chimeric molecules and can be added at a later stage of the editing process.

As discussed above, truncated gRNAs are abundant in the chimeric molecules. If we assume that the truncated gRNAs themselves participate in the editing process [and this appears likely from the results of the *in vitro* systems in which a short (25 nt) synthetic gRNA proved efficient in chimer formation, see Harris and Hajduk, 1992], those that do not possess the

complete information for an editing domain would give rise to a partially edited transcript. As in other trypanosomes, in *C.fasciculata* partially edited RNAs are occasionally picked up in cDNA analysis (Van der Spek *et al.*, 1990), indicating that they are present at relatively high concentrations. Partially edited RNAs produced by the different truncated gRNAs together would resemble the collection of partially edited CYb and cox 3 RNAs found in *L.tarentolae* (Sturm and Simpson, 1990a) that has inspired the proposal of a strict 3′ to 5′ polarity and stepwise processivity of the editing process within the domain of a single gRNA. Given the completely different way in which, at least in theory, such partially edited RNAs can also be produced, we feel that some caution should be exerted in extracting from their existence any clues as to the direction of the editing process.

Some of the other chimeric molecules convey a similar message. In 20% of chimeric molecules, the gRNA is hooked up at (or close to) the 'wrong' editing domain and in two of the molecules the [FS] gRNA appears to be linked to the 5′ end of a pre-mRNA species (see Figure 6D and E). It has recently been proposed that unexpected chimeric molecules could be explained by assuming that false anchors are formed between the region of the pre-mRNA immediately downstream of the nucleotide to which the gRNA is linked and some segment of the gRNA (Sturm *et al.*, 1992). Although a reasonably stable duplex between the [5′] gRNA and a sequence 9−18 nt downstream of the [FS] domain can be formed (underlined in Figure 6D), alternative anchors are absent from the region immediately adjacent (3′) to the linkage sites. Base-pairing of low thermodynamic stability is possible between the [FS] gRNA and the region around the [5′] editing domain but anchors of comparable low strength can be formed with other, non-cognate, gRNAs such as CYb-II and MURF2-II gRNA (not shown). We have, however, been unable to detect chimeric molecules in which CYb-II and MURF2-II gRNAs were linked to ND7 pre-mRNA. The apparent transcript specificity of our unexpected ND7 chimeras can be explained, nevertheless, if one assumes that chimeras can be formed at a distance from the site of anchoring. For the *C.fasciculata* ND7 RNA, this would mean that gRNAs anchored to their own domain could interact with the other domain at a low but detectable frequency, 172 nt upstream or downstream. In some chimeric molecules found in *T.brucei* mt RNA and particularly in those produced *in vitro*, the linkage point is indeed found at a considerable distance from the proposed anchors (occasionally more than 50 nt, see Koslowsky *et al.*, 1992a). Although the unexpected ND7 chimeras clearly do not represent useful intermediates of editing, their existence may be indicative of a great flexibility of the participating RNA molecules, allowing chimeric molecules to be formed at some distance from the anchor duplex. Such a flexibility would be in line with models that do not invoke a strict 3′ to 5′ polarity of editing within the domain of a gRNA (Decker and Sollner-Webb, 1990; Koslowsky *et al.*, 1991; Read *et al.*, 1992).

The observation that in some chimeric molecules a gRNA is hooked up to the 5′ end of a pre-edited RNA is intriguing, particularly in the light of the unpublished results from the Sollner-Webb laboratory, who claim that apparently 'normal' chimeric molecules can be produced by the combined actions of mung bean nuclease and RNA ligase (Sollner-Webb, 1992). This might suggest that chimeric molecules can also be formed via end-to-end ligation of gRNA and pre-mRNA. Further identification of the components involved and elucidation of their mechanism of action should reveal whether RNA editing in trypanosome mitochondria is indeed carried out via an enzymic 'cut and paste' process and not via a sequence of concerted, RNA-mediated, transesterification reactions.

Materials and methods

Oligonucleotides

The oligonucleotides used in the experiments reported in this work are listed in Table II.

Cell culture and nucleic acid isolation

Cultures of *C.fasciculata* were grown as described by Kleisen *et al.* (1975). Total and mt RNA was isolated as described in Van der Spek *et al.* (1991).

Table II.

Oligonucleotides (5′→3′)		RNA	gRNA coordinates[a]	Gene coordinates and reference[b]
C30	ATTACAAACAACACTAAC	ND7 pre-edited		3373−3356[1]
C34	CATAAGGATAGCAAATGTTC	ND7 pre-edited		3393−3374[1]
C42	CCAAAAGTAAATCGATATAAGTG	ND7 pre-edited		3191−3169[1]
C80	TATTTATTTTGTCTTTATTGCCTTTTTG	ND7 [5′]gRNA	44−17	110−137[1]
C81	TTACGTTCTCTCATATTAAACCCCTATT	CYb-II gRNA	53−26	2088−2115[1]
C90	AACAGCATTAGTCTAATCTATC	ND7 [FS]gRNA	1−22	59−80[2]
C91	ACGGCTGATATAAGTGCAAAAAG	ND7 [5′]gRNA	1−23	153−131[1]
C99	CATTCTGATAGATTAGACTAATGCTGTT	ND7 [FS]gRNA	28−1	86−59[2]
C100	ATTGCCTTTTTGCACTTATATCAGCCGT	ND7 [5′]gRNA	28−1	126−153[1]
C104	CCCCATTTCTATGCTTTGACTGAGT	MURF2-II gRNA	31−7	91−67[2]
C105	TGTTTAATCACTTTAATTTAACCCCA	MURF2-II gRNA	52−27	112−87[2]
C106	ATTGTCCTTTTTGTTATTTAGAACATTT	CYb-II gRNA	28−1	2113−2140[1]
C107	ATTCAACTCAGTCAAAGCATAG	MURF2-II gRNA	2−23	62−83[2]
C108	AAATGTTCTAAATAACAAAAAGGAC	CYb-II gRNA	1−25	2140−2116[1]
C109	TAAATCGATATAAGTGCCTTTTT T_4A_{20}	ND7 pre-edited		3184−3162[1]

[a]The most 5′ nucleotide of a gRNA is taken as number 1 (see Figure 4). [b] [1]Sloof *et al.* (1987), [2]Van der Spek *et al.* (1991).

Routinely, the enrichment of mt RNA for ribosomal 9S and 12S RNAs was ~50- to 100-fold, as judged by Northern blot analysis. Small mt RNA of 30–80 nt was isolated from a 1% low melting point agarose gel using the hot phenol method (Sambrook *et al.*, 1989). Plasmid DNA was prepared according to Birnboim and Doly (1979).

In vitro RNA synthesis

The *in vitro* T7 transcription reactions were performed essentially as described in Sambrook *et al.* (1989). Plasmids containing gRNA sequences were linearized with *Dra*I; the *Dra*I restriction site was included in the T_4A_{20} PCR primer used for gRNA cloning. The gRNAs produced contain two G-residues at the 5′ end derived from the pAC2 vector and a varying number of U-residues at the 3′ end. Vector DNA was removed by a DNase I digestion (5 μg/ml, 15 min, 37°C) according to Tullis and Rubin (1980).

Electrophoresis, blotting, hybridization and quantification of RNA

Electrophoresis of glyoxylated RNA on 1.75% agarose gels and blotting was performed as described in Van der Spek *et al.* (1990). Size analysis of gRNAs was performed as follows: samples with 15 μg of mtRNA or various amounts of *in vitro* synthesized ND7 gRNA as indicated, mixed with 15 μg total RNA of *C.fasciculata*, were heated for 3 min at 80°C before loading onto a 14% polyacrylamide gel in 7 M urea, 90 mM Tris–borate, 2.5 mM EDTA. The RNA was blotted with 25 mM sodium phosphate buffer, pH 6.5, for 75 min at 2 mA/cm^2 in a cooled Biometra fast blot apparatus onto Hybond N nylon filter (Amersham). The filters were UV cross-linked to prevent loss of small RNAs using a Stratagene UV Stratalinker 1800 (254 nm, 120 mJ/cm^2). *In vitro* synthesized RNAs of known size (65, 45 and 35 nt) were labelled and used as low molecular weight markers. For hybridization purposes oligonucleotides were phosphorylated at the 5′ end using T4 polynucleotide kinase (Pharmacia) and [γ-^{32}P]ATP (Amersham). The blots were prehybridized for 1–2 h and hybridized (14–18 h) in hybridization mixture (Sambrook *et al.*, 1989) at 30°C in a minimal volume (6–16 ml), washed for three consecutive periods of 2 min in 6×SSC, 0.1% SDS at room temperature, elevated temperature and room temperature, respectively. The elevated temperature is 6°C beneath the melting temperature calculated according to Itakura *et al.* (1984): C80, 60°C; C81, 68°C; C99, 65°C; C100, 65°C; C104, 66°C; C105, 60°C; C106, 62°C. Blots were exposed overnight or longer with Kodak X-OMAT-S or -AR using intensifying screens. The ND7 cDNA probe was nick-translated and used in hybridization (14–18 h) in 0.25 M sodium phosphate buffer pH 7.2, 7% SDS, 1 mM EDTA. Filters were washed three times for 20 min in 3×SSC, 0.1% SDS at 65°C.

For quantification, blots were exposed for 4–24 h to a phosphor screen and intensities of signals were analysed with a Molecular Dynamics PhosphorImager apparatus using ImageQuant version 3.0. The amount of a mt RNA species was determined by comparison of their hybridization signals with those obtained with known amounts of *in vitro* synthesized RNA.

gRNA cloning

For cDNA synthesis reactions 20 pmol T_4A_{20} primer was used with 250 ng of mt RNA or 2 μg of total RNA from *C.fasciculata*. After a denaturation step of 2 min at 70°C, 2.25 units of AMV reverse transcriptase (Promega) were added and cDNA was synthesized in a Hybaid Thermal Reactor in three 10 min periods at 28, 35 and 42°C, respectively, followed by a 5 min inactivation step at 95°C. For the PCR reactions, Perkin Elmer Cetus GeneAmp (AmpliTaq) and 20 pmol of various primers were used (C90, C91, C107 and C108), specific for the 5′ end of ND7 [FS], ND7 [5′], MURF2-II and CYb-II gRNA, respectively. After a denaturation step of 5 min at 95°C, 30 cycles were performed, each consisting of 1 min at 95°C, 1.5 min at 40°C and 2 min at 72°C, followed by an 8 min final extension step at 72°C. The PCR products were cloned in the *Stu*I site of the pAC2 vector, which is derived from pUC19 by insertion into the *Bam*HI site of the following DNA fragment: 5′-GATCTCGATCCCGCGAAA-TTAATACGACTCACTATaggcctG-3′, the T7 promoter is underlined, the *Stu*I site is in lower case.

Cloning of chimeric molecules

Chimeric molecules were cloned essentially as gRNAs with a few modifications. For cDNA synthesis, 20 pmol of the C34 was used, which is complementary to ND7 RNA a few nucleotides downstream of the [FS] editing domain. After a denaturation step at 70°C for 5 min, enzyme was added and the incubation continued at 42°C for 35 min, followed by a 5 min inactivation step at 95°C. In the amplification reactions primers C90, C91, C107 or C108 were used in a PCR protocol similar to that used for gRNA amplification. The PCR products were also cloned in the *Stu*I site of pAC2. Colonies were lifted and the filters were screened with oligonucleotide C30 or C42 as a probe.

Sequencing and primer extension reactions

Clones were sequenced according to Sanger *et al.* (1977) with a Pharmacia T7 Sequencing kit and standard pUC forward and reverse primers. For the determination of the percentage of editing of both ND7 domains, primer extension reactions were performed essentially as described by Feagin *et al.* (1988) and Van der Spek *et al.* (1991). However, in our experiments 2 μg mt RNA or 10 ng *in vitro* synthesized RNA was used in combination with AMV reverse transcriptase (1 U), 1 ng 5′ labelled primer C30 or C42 and 125 μM ddGTP, 100 μM dATP, dTTP and dCTP. The intensities of the corresponding extension products were determined using a PhosphorImager apparatus.

Acknowledgements

We thank Professors Drs P.Borst and H.F.Tabak for their continuing interest and stimulating discussions. We are indebted to Ms A.de Haan for expert technical assistance, to Mrs E.Vlugt for skilful handling of the manuscript and to Ms W.van Noppen for invaluable editorial assistance. This work was supported by the Netherlands Foundation of Chemical Research (SON), which is subsidized by the Netherlands Foundation for Scientific Research (NWO).

References

Abraham,J.M., Feagin,J.E. and Stuart,K. (1988) *Cell*, **55**, 267–272.
Bakalara,N., Simpson,A.M and Simpson,L. (1989) *J. Biol. Chem.*, **264**, 18679–18686.
Benne,R. (1990) *Trends Genet.*, **6**, 177–181.
Benne,R. (1992) *Mol. Biol. Rep.*, **16**, 217–227.
Bhat,G.J., Koslowsky,D.J., Feagin,J.E., Smiley,B.L. and Stuart,K. (1990) *Cell*, **61**, 885–894.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.
Blum,B. and Simpson,L. (1990) *Cell*, **62**, 391–397.
Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
Blum,B., Sturm,N.R., Simpson,A.M. and Simpson,L. (1991) *Cell*, **65**, 543–550.
Cech,T.R. (1991) *Cell*, **64**, 667–669.
Decker,C.J. and Sollner-Webb,B. (1990) *Cell*, **61**, 1001–1011.
Feagin,J.E., Shaw,J.M., Simpson,L. and Stuart,K. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 539–543.
Harris,M.E. and Hajduk,S.L. (1992) *Cell*, **68**, 1–20.
Harris,M.E., Decker,C., Sollner-Webb,B. and Hajduk,S. (1992) *Mol. Cell. Biol.*, **12**, 2591–2598.
Itakura,K., Rossi,J.J. and Wallace,R. (1984) *Annu. Rev. Biochem.*, **53**, 323–356.
Kleisen,C.M., Borst,P. and Weijers,P. (1975) *Biochim. Biophys. Acta*, **390**, 155–167.
Koslowsky,D.J., Bhat,G.J., Perrolaz,A.L., Feagin,J.E. and Stuart,K. (1990) *Cell*, **62**, 901–911.
Koslowsky,D.J., Bhat,G.J., Read,L.K. and Stuart,K. (1991) *Cell*, **67**, 537–546.
Koslowsky,D.J., Göringer,H.U., Morales,T.H. and Stuart,K. (1992a) *Nature*, **356**, 807–809.
Koslowsky,D.J., Riley,G.R., Feagin,J.E. and Stuart,K. (1992b) *Mol. Cell. Biol.*, **12**, 2043–2049.
Maslov,D.A. and Simpson,L. (1992) *Cell*, **70**, 459–467.
Pollard,V.W. and Hajduk,S.L. (1991) *Mol. Cell. Biol.*, **11**, 1668–1675.
Pollard,V.W., Rohrer,S.P., Michelotti,E.F., Hancock,K. and Hajduk,S.L. (1990) *Cell*, **63**, 783–790.
Read,L.R., Corell,R.A. and Stuart,K. (1992) *Nucleic Acids Res.*, **20**, 2341–2347.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*. 2nd 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.
Shaw,J.M., Campbell,D. and Simpson,L. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6220–6224.
Simpson,A.M., Bakalara,N. and Simpson,L. (1992) *J. Biol. Chem.*, **267**, 6782–6788.
Simpson,L. and Shaw,J.M. (1989) *Cell*, **57**, 355–366.
Sloof,P., van den Burg,J., Voogd,A. and Benne,R. (1987) *Nucleic Acids Res.*, **15**, 51–65.
Sollner-Webb,B. (1992) *Nature*, **356**, 743–744.
Souza,A., Myler,P.J. and Stuart,K. (1992) *Mol. Cell. Biol.*, **12**, 2100–2107.
Stuart,K. (1991a) *Annu. Rev. Microbiol.*, **45**, 327–344.
Stuart,K. (1991b) *Trends Biochem. Sci.*, **16**, 68–72.

Sturm,N.R. and Simpson,L. (1990a) *Cell*, **61**, 871–878.
Sturm,N.R. and Simpson,L. (1990b) *Cell*, **61**, 879–884.
Sturm,N.R. and Simpson,L. (1991) *Nucleic Acids Res.*, **19**, 6277–6281.
Sturm,N.R., Maslov,D.A., Blum,B. and Simpson,L. (1992) *Cell*, **70**, 469–476.
Tullis,R.H. and Rubin,H. (1980) *Anal. Biochem.*, **107**, 260–264.
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.
Van der Spek,H., Arts,G.J., Zwaal,R.R., Van den Burg,J., Sloof,P. and Benne,R. (1991) *EMBO J.*, **10**, 1217–1224.
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.

Received on November 17, 1992; revised on January 12, 1993