Cell, Vol. 70, 459–467, August 7, 1992, Copyright © 1992 by Cell Press

The Polarity of Editing within a Multiple gRNA-Mediated Domain Is Due to Formation of Anchors for Upstream gRNAs by Downstream Editing

Dmitri A. Maslov* and Larry Simpson*†‡
* Department of Biology
† Howard Hughes Medical Institute
‡ Molecular Biology Institute
University of California, Los Angeles
Los Angeles, California 90024-1606

Summary

Seventeen kinetoplast minicircle-encoded and nine maxicircle-encoded gRNA genes have been identified. Six overlapping minicircle-encoded gRNAs mediate editing for the 5′-pan-edited *MURF4* gene and two for the 5′-edited *COIII* gene. The pan-edited *RPS12* mRNA is edited by seven minicircle-encoded gRNAs and one maxicircle-encoded gRNA. The 3′-most gRNA in each domain forms an anchor with unedited mRNA, whereas upstream gRNAs form anchors only with edited mRNA, thereby explaining the observed 3′ to 5′ polarity of editing within an editing domain. We suggest that a role of G–U base pairs is to allow breathing of the edited mRNA–gRNA hybrid and formation of the upstream anchor hybrid.

Introduction

The editing of transcripts of maxicircle cryptogenes (Benne, 1989; Simpson and Shaw, 1989; Stuart, 1991) in the mitochondria of kinetoplastid protozoa is mediated by small transcripts known as guide RNAs (gRNAs), which contain the edited sequence information and possibly also act as a source of U residues (Blum et al., 1990, 1991; Simpson, 1990; Pollard et al., 1990). gRNAs are transcribed from both the maxicircle and the minicircle DNA molecules present in the kinetoplast DNA (kDNA) network. In Leishmania tarentolae, eight maxicircle-encoded gRNAs (Blum et al., 1990) and five minicircle-encoded gRNAs (Sturm and Simpson, 1990b, 1991) have been identified so far.

The minicircle component of L. tarentolae consists of approximately 10^4 molecules 850–880 bp in size (Simpson, 1987). Each minicircle is organized into a conserved region of 170–180 bp, which contains the origins of replication for both strands, and a variable region, which contains a single gRNA gene located approximately 150 bp from the conserved region (Sturm and Simpson, 1991). Five different minicircle sequence classes have been sequenced to date, and an estimate of approximately 10–20 total sequence classes was derived from the appearance of kDNA restriction patterns in acrylamide gels.

Studies of partially edited transcripts showed that editing generally proceeds 3′ to 5′ on the preedited mRNA, both within a single editing block (the editing mediated by a single gRNA) and within an editing domain (a region involving multiple gRNAs) (Sturm and Simpson, 1990a; Decker and Sollner-Webb, 1990; Abraham et al., 1988). The existence of unexpected editing patterns in junction regions between fully edited and unedited sequences (Sturm and Simpson, 1990a; Decker and Sollner-Webb, 1990) can be explained by a 3′ to 5′ processive mechanism (Sturm et al., 1992 [this issue of *Cell*]), but alternative hypotheses have also been proposed (Decker and Sollner-Webb, 1990; Koslowsky et al., 1991).

In this paper we characterize the remaining minicircle-encoded gRNAs and one additional maxicircle-encoded gRNA from the kinetoplast of L. tarentolae in terms of specific editing roles. We show that the observed 3′ to 5′ polarity of editing within a domain is a direct consequence of the formation of edited anchor sequences for hybridization of adjacent upstream gRNAs, and we provide a rationale for the existence of G–U base pairs in the mRNA–gRNA hybrid.

Results

Identification of Additional Minicircle Sequence Classes

Two plasmid libraries of unit-length minicircle DNA were constructed as described in Experimental Procedures. Library 1 consisted of 2500 clones of XmaI single-cut minicircles, which would allow the detection of minicircle classes present in approximately 5–10 copies per network. Library 2 consisted of approximately 25,000 clones of permuted MspI-linearized minicircles. This library would allow detection of single-copy minicircles with a 90% probability. Clones containing minicircle inserts were identified by colony hybridization using an oligonucleotide probe (*S-293*) to the conserved sequence block 3 (Ray, 1989) region of the minicircle (Figure 1). Mixed oligonucleotide probes for the variable regions of the molecules were used in a negative selection protocol to detect clones containing minicircles distinct from the 5 known sequence classes. Clones selected by this method were then screened by single-lane sequencing with the oligonucleotide primer, *S-486*, which anneals to the edge of the conserved region and extends into the variable region (Figure 1). Novel clones were completely sequenced.

As shown in Table 1, 96% of the clones in library 1 consisted of the 5 known sequence classes (Sturm and Simpson, 1991; Kidane et al., 1984; Muhich and Simpson, 1986) in different frequencies, and 4% of the clones contain novel sequence classes. The latter 100 clones comprised 11 different novel minicircle sequence classes.

Minicircle library 2 was approximately 10-fold larger than library 1. As an internal control for the completeness of this library, the initial colony hybridization for known classes was performed with a mixture of variable region probes representing only 14 of the identified 16 minicircle sequence classes. Both of these minicircle classes (X2 and X8 in Table 1) were detected among the remaining 34 clones that did not hybridize with the probe mixture. In addition, three clones were obtained that represented a new minicircle sequence class, bringing to 17 the total number of minicircle sequence classes. The complete se-

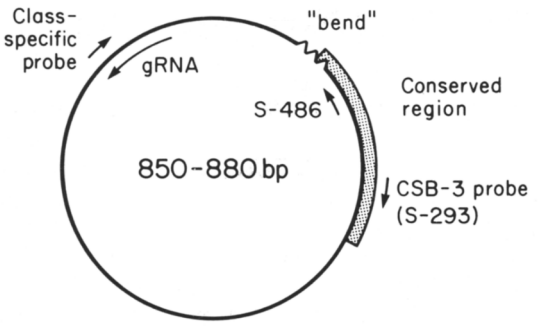

Figure 1. Diagram of kDNA Minicircle Molecule

The locations of the conserved region, the bend, and the gRNA gene, and the oligomeric sequences used for detection of an insert (*S-293*), detection of specific sequence classes, and for sequencing of a gRNA gene (*S-486*) are shown. Conserved sequence block 3 (CSB-3) is a conserved 12-mer sequence, and the arrows indicate the polarities of the probe sequences.

quences of the 12 new minicircle classes will be presented elsewhere.

Relative Abundance of Minicircle Sequence Classes

The relative copy number of the different minicircle sequence classes can be estimated based on the occurrence of specific clones in the libraries. These values, however, are only approximate since they assume random release of minicircles from the network and equal relative cloning efficiencies of the different minicircles. The main conclusion from these data is that the previously described Lt19 and B4 minicircles (Muhich and Simpson, 1986; Kidane et al., 1984; Sturm and Simpson, 1991) represent the major sequence classes.

This question was approached more directly by probing Southern blots of MspI-digested kDNA with minicircle-specific oligomers of known specific activities. This enzyme has previously been shown to release all minicircle and maxicircle DNA from the network (Simpson, 1979). Each probe hybridized with a major minicircle band and, in some cases, a minor band. The relative amounts of the minor MspI bands that hybridized with the specific probes varied from less than 1% to as much as 18% of the major bands. We have not investigated the nature of these polymorphisms. Densitometric analysis of the hybridizing bands provided values for the relative abundances of the various minicircle sequence classes (Table 2). In agreement with the clone frequency data, the major network minicircles are the Lt19 and B4 sequence classes, which are present in 2,500 and 3,000 copies, respectively, per network of 10,000 minicircles. The copy numbers of the remaining 15 minicircle classes vary from 30 to 1050.

Identification of gRNA Genes Encoded by 12 New Minicircles

A previous analysis (Sturm and Simpson, 1991) of 5 cloned minicircles from L. tarentolae indicated that a single gRNA gene is localized in the variable region on the same strand,

Table 1. Analysis of Minicircle Plasmid Libraries

Number of Selected Clones	Number of Analyzed Clones	Minicircle Sequence Class	Number of Clones	Encoded gRNA
Library 1[a]				
Old classes 2400[b] (96%)	92[c]	Lt19	59	g*Lt19*
		B4	28	g*B4*
		D3	3	g*MURF4-III*
		D12	3	g*COIII-I*
		Lt154	0	g*RPS12-IV*
New classes 100[d] (4%)	100	X1	28	g*RPS12-III*
		X2	1	g*MURF4-VI*
		X3	3	g*MURF4-I*
		X4	9	g*COIII-II*
		X5	10	g*RPS12-I*
		X6	31	g*RPS12-V*
		X7	7	g*MURF4-V*
		X8	1	g*RPS12-II*
		X9	2	g*MURF4-II*
		X10	3	g*RPS12-VIII*
		X11	3	g*RPS12-VII*
Library 2[e]				
34[f]	34	X2	24	g*MURF4-VI*
		X8	7	g*RPS12-II*
		M1	3	g*MURF4-IV*

[a] XmaI-cut minicircle clones (2500 clones). Negative selection with 5 classes: Lt19, B4, Lt154, D3, and D12.
[b] The clones were positive for the 5 known minicircle classes.
[c] Randomly chosen clones were analyzed.
[d] The minicircle DNA clones that were negative for the 5 known sequence classes.
[e] MspI + ethidium bromide–cut minicircle clones (~25,000 clones). Negative selection with all above classes but X2 and X8.
[f] The minicircle DNA clones that were negative for 14 other sequence classes.

Table 2. The Relative Abundance of Minicircle Sequence Classes

Minicircle Sequence Class/Encoded gRNA	Copy Number per Network[a]
Lt19/g*Lt19*	2500
B4/g*B4*	2980
X3/g*MURF4-I*	220
X9/g*MURF4-II*	110
D3/g*MURF4-III*	380
M1/g*MURF4-IV*	330
X7/g*MURF4-V*	210
X2/g*MURF4-VI*	320
X5/g*RPS12-I*	370
X8/g*RPS12-II*	140
X1/g*RPS12-III*	1050
Lt154/g*RPS12-IV*	210
X6/g*RPS12-V*	500
X11/g*RPS12-VII*	90
X10/g*RPS12-VIII*	30
D12/g*COIII-I*	190
X4/g*COIII-II*	370

[a] Assuming 10^4 minicircles per network.

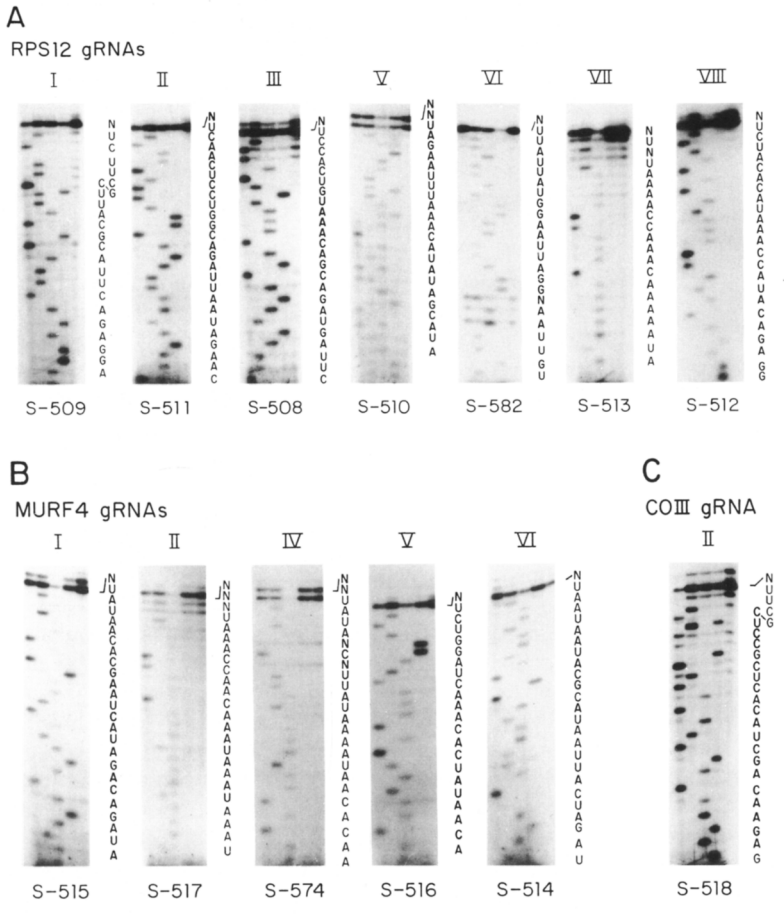

Figure 2. Primer Extension Sequencing of 5′ Ends of New gRNAs

The sequencing primers used are indicated below each lane, and the sequences obtained are beside each lane.

(A) gRNAs for *RPS12*. The sequence ladder of g*RPS12-I* shows a compression as indicated, and the sequence was resolved using the genomic sequence.

(B) gRNAs for *MURF4*.

(C) *COIII*. The sequence ladder shows a compression, as indicated.

approximately 150 bp from the end of the conserved region (Figure 1). Analogous regions of the 12 new minicircle sequences were analyzed for potential gRNA sequences involved in the known editing of the *COIII*, *MURF4*, or *RPS12* mRNAs, using the local alignment program described in Experimental Procedures. In each case, a short DNA sequence was identified at the expected location and in the expected polarity that could represent a gRNA gene (see Figures 4, 5, and 6).

Identification of 12 New Minicircle-Encoded gRNAs

Oligonucleotide probes complementary to the gRNA sense strands determined from the DNA sequences were used to detect the predicted gRNAs in steady-state kRNA.

The 5′ ends of the 12 gRNAs were determined by primer extension sequencing (Figure 2). Single major 5′ ends were observed in most cases; g*MURF4-IV* and g*RPS12-V* both showed two equimolar 5′ ends differing by 1 nt. The RNA sequences agreed in all cases with those predicted from the minicircle DNA sequences.

The relative abundances of the gRNAs were measured by densitometric analysis of the Northern blots (Figure 3), with correction for the specific activities of the probes. Although it is clear that the most abundant gRNA is derived from the high copy number Lt19 minicircle, there is no direct correlation of the relative abundance of the remaining minicircle-encoded gRNAs to the copy numbers of the minicircles encoding those gRNAs. This suggests that different promoter strength and/or RNA degradation rates are responsible for gRNA steady-state levels, rather than a gene dosage effect.

Identification of Editing Roles of gRNAs

Editing of the COIII mRNA

The D12 minicircle was previously shown to contain a gRNA gene (the g*COIII-I* gene) (Sturm and Simpson, 1990b), which encodes editing information for the first 8 sites of the 5′-edited *COIII* mRNA in L. tarentolae, but no gRNA had been reported for the remaining editing sites. As shown in Figure 4, one of the new minicircle-encoded gRNAs (minicircle class X4 in library 1, the gRNA designated as g*COIII-II*) encodes the information for the remaining 9 sites of this 5′-edited mRNA. The anchor sequence for g*COIII-II* covers editing sites 5–8 of the g*COIII-I* gRNA and is complementary to the edited mRNA sequence, whereas the anchor sequence for the g*COIII-I* gRNA is complementary to the unedited mRNA sequence (Sturm and Simpson, 1990b).

Editing of the MURF4 mRNA

Editing of the 5′ portion of the L. tarentolae *MURF4* mRNA involves 46 sites with 106 U additions and 4 sites with 5 U deletions (Bhat et al., 1990, 1991). A single minicircle-

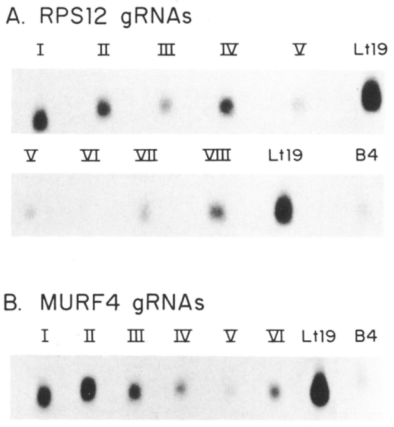

Figure 3. Northern Blot Analysis of gRNAs for *RPS12* and *MURF4*

Equal amounts of kRNA were loaded in each lane of formaldehyde–agarose gels. The blots were hybridized with equal amounts of the indicated labeled oligomeric probes, and the bands were quantitated by densitometry. The Lt19 and B4 probes were included as internal references to calculate relative abundances.

encoded gRNA (D3 minicircle) was previously shown to edit a portion of the *MURF4* mRNA (Bhat et al., 1990). As shown in Figure 5, six overlapping minicircle-encoded gRNAs mediate editing of this entire region. The 3′-most gRNA (g*MURF4-I*) forms an anchor duplex with the unedited mRNA sequence, but all the other gRNAs form anchors with edited mRNA sequences ranging in size from 10 to 17 bp.

It is of interest that the 5′-most pair of uridine deletions in the mRNA appear to be guided by the nonencoded 3′-oligo(U) tail (Blum and Simpson, 1990) of g*MURF4-VI* rather than by minicircle-encoded guide nucleotides as in all other known examples.

The 9 nt unedited mRNA sequence between editing sites 30 and 31 may represent a domain connection sequence. Since the precise 3′ end of g*MURF4-IV* is not known, it is possible that the actual gRNA sequence is several nucleotides shorter than shown in Figure 5 and that g*MURF4-V* forms a 9 bp anchor with the unedited domain connection sequence, thereby initiating editing of the short 5′ domain. However, additional evidence in the form of partially edited mRNAs must be obtained to confirm this hypothesis.

Editing of the RPS12 mRNA

RPS12 is a pan-edited cryptogene comprising three editing domains joined by two short unedited domain-connection sequences (DCS-I and DCS-II) (Maslov et al., 1992). As shown in Figure 6, editing of domain I is mediated by a single gRNA (g*RPS12-I*), which forms an anchor hybrid with the unedited mRNA sequence and terminates the gRNA-edited mRNA hybrid at DCS-I.

The first gRNA of domain II (g*RPS12-II*) forms an anchor with the unedited DCS-I sequence and creates an anchor by editing for the second gRNA (g*RPS12-III*), which then creates an anchor for the third gRNA (g*RPS12-IV*), which in turn creates an anchor for the fourth gRNA (g*RPS12-V*). This gRNA has been shown previously to be derived from the Lt154 minicircle (Maslov et al., 1992). The g*RPS12-V*-edited mRNA hybrid terminates at the unedited DCS-II sequence.

The remaining two minicircle-encoded gRNAs (g*RPS12-VII* and g*RPS12-VIII*) mediate editing of the 5′ portion of domain III. No minicircle-encoded gRNA was found that could mediate editing of the 3′ portion of domain III.

However, a putative gRNA gene for the 3′ portion of domain III (g*RPS12-VI*) was indicated by computer search to be located in the maxicircle genome just downstream of the *ND5* gene (LEIKPMAX nucleotides 17,020–16,975). The localization of this gRNA gene within the maxicircle DNA was demonstrated by Southern blot analysis (data not shown), and the existence of the gRNA in steady-state mitochondrial RNA was verified by direct 5′-end sequencing (see Figure 2). This is the first example of a pan-edited gene that has both minicircle- and maxicircle-encoded gRNAs. g*RPS12-VI* forms an anchor with the unedited DCS-II sequence, whereas the second and third gRNAs (minicircle-encoded g*RPS12-VII* and g*RPS12-VIII*) for this domain each forms anchors with edited mRNA sequences. The 5′-most gRNA–mRNA duplex ends 3 nt after the final editing event.

Role of Noncanonical Base Pairs in gRNA–mRNA Hybrids

The gRNA-edited mRNA anchor duplex regions in general consist almost entirely of the standard G–C and A–U base pairs; the relatively weak G–U wobble base pairs are mainly found outside the anchor regions, and A–C base pairs (Van der Spek et al., 1990) are very rare. As shown in Table 3 for the *MURF4* and *RPS12* genes, the relative

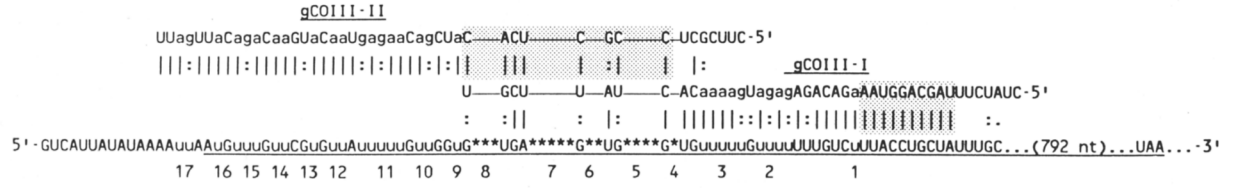

Figure 4. Two Overlapping gRNAs Mediate Editing of the *COIII* mRNA

The 5′ end of the edited *COIII* mRNA is shown, with the editing sites numbered. Added uridines are shown as lowercase u's and the deleted uridines as asterisks. The open reading frame of the edited mRNA is indicated by underlining. The complementary g*COIII-I* (Sturm and Simpson, 1990b) and g*COIII-II* sequences are shown with G–C and A–U base pairs indicated by parallel lines and G–U base pairs by colons. Guide nucleotides in the gRNA sequences are indicated as lowercase a's and g's. The 3′ anchor duplex regions for each gRNA–mRNA hybrid are indicated by stippling.

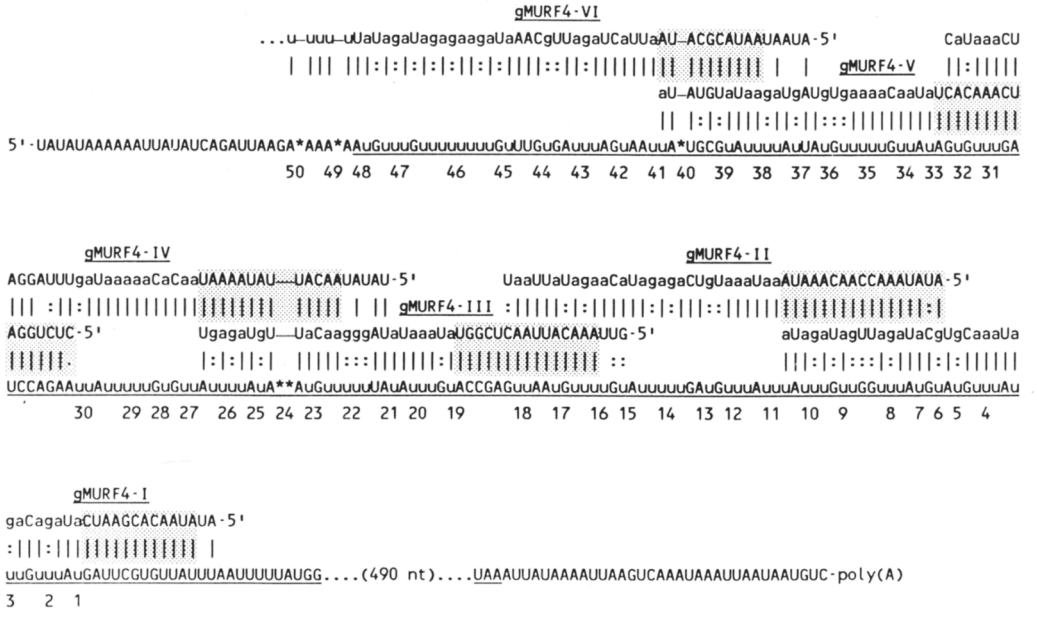

Figure 5. Six Overlapping gRNAs Mediate Editing of the *MURF4* mRNA

The 5′ portion of the mature edited *MURF4* mRNA is shown with editing sites numbered. See the Figure 4 legend for other symbols. The UAA termination codon is indicated by underlining.

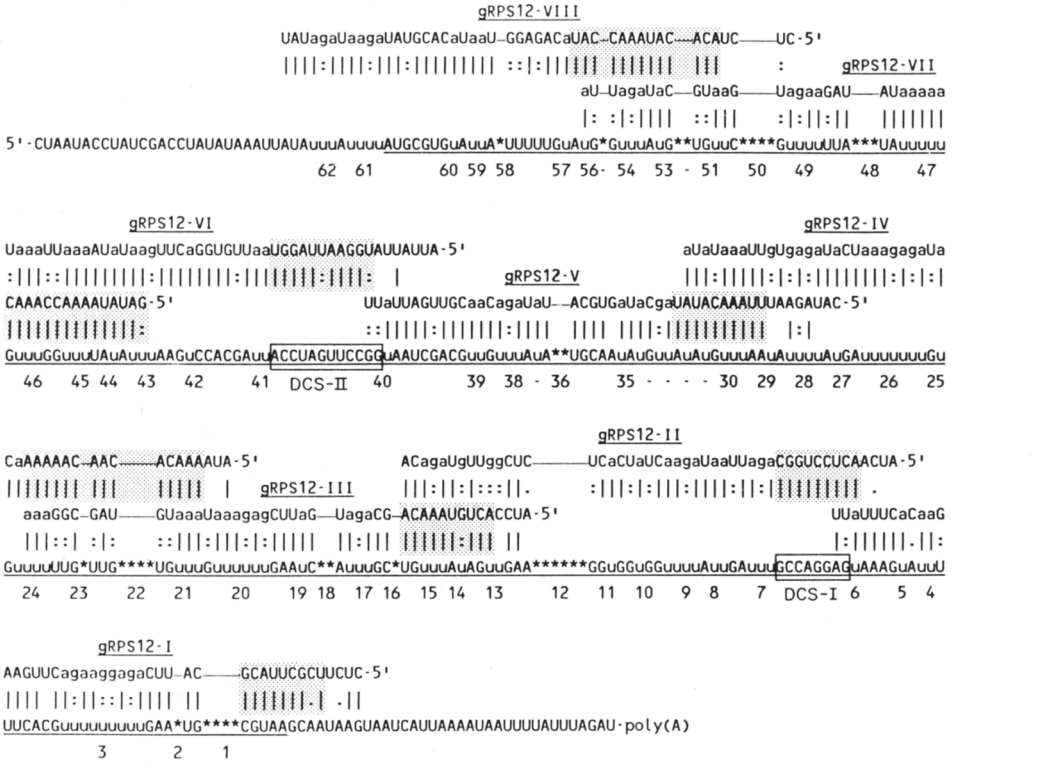

Figure 6. Eight Overlapping gRNAs Mediate Editing of the *RPS12* mRNA

The entire mature edited *RPS12* mRNA sequence is shown, with the open reading frame underlined and the editing sites numbered. DCS-I and DCS-II, domain connection sequences I and II. See the Figure 4 legend for other symbols.

Table 3. Free Energy of gRNA Anchor versus Free Energy of Overlapping Downstream gRNA-Edited mRNA Duplex

gRNA Pairs[a]	$\Delta G_{anchor}/\Delta G_{3'duplex}$[b] (kcal/mol)	Ratio
RPS12 gRNAs		
III/II	−7.7/−4.1	1.9
IV/III	−13.7/−6.6	2.1
V/IV	−6.3/−2.9	2.3
VII/VI	−13.6/−7.7	1.8
VIII/VII	−15.0/−5.5	2.7
MURF4 gRNAs		
II/I	−16.3/−9.0	1.8
III/II	−18.3/−2.6	7.0
IV/III	−9.0/−7.0	1.3
V/IV	−21.2/−11.8	1.8
VI/V	−9.9/−5.4	1.8
COIII gRNAs		
II/I	−9.9/−3.4	2.9

For each pair of consecutive gRNAs, the free energy of anchor formation for the upstream gRNA (without the very 5′ end, which is often unpaired) was compared with the free energy of duplex formation between the overlapping 3′ part of the downstream gRNA and the edited mRNA.

[a] For each set of overlapping gRNAs, the upstream anchor forming gRNA is on the left, and the gRNA to be displaced is on the right.

[b] The free energies of duplex formation were determined using the GCG FOLD program with the default base match energy table.

thermodynamic stabilities of anchor duplexes are consistently greater than those of the corresponding mRNA–downstream gRNA duplexes in the same region. This would allow breathing of the end of the edited mRNA–gRNA duplex and strand invasion of the 5′ end of the incoming upstream gRNA to form a more stable hybrid and thereby initiate another editing block. We suggest that this is a functional explanation for the existence of G–U base pairs in mRNA–gRNA duplex structures.

A U–C mismatch is present between editing sites 3 and 4 of the *RPS12* mRNA. The helix-destabilizing effect of this mismatch could be minimized by the presence of a water molecule (Holbrook et al., 1991). Two examples of A–C base pairs between the guiding portion of a gRNA and the edited mRNA occur in L. tarentolae *RPS12* mRNA. In both these cases and in the two previously reported A–C base pairs in Crithidia fasciculata gRNA–mRNA hybrids (Van der Spek et al., 1990), the C residue is located on the gRNA and the A residue on the mRNA. This asymmetry is probably functional, since when an A residue on the gRNA encounters a C residue on the preedited mRNA, it does not form an A–C base pair but rather directs an insertion of U into the mRNA. C is, in general, a rare nucleotide in preedited mRNA sequences.

Discussion

The kDNA network in L. tarentolae contains approximately 10^4 minicircles comprising at least 17 different sequence classes. The copy numbers of the different minicircle sequence classes present in these libraries were estimated by the frequencies of clones obtained and by quantitative Southern blot analysis. It is clear that minicircles of different sequence classes are present in widely different abundances and that some classes are present at an extremely low copy number. No correlation was observed between the relative abundance of gRNAs and the relative copy number of the template minicircle.

The 17 identified minicircle sequence classes probably exhaust the minicircle-encoded gRNA sequence complexity in this strain. In addition to minicircle-encoded gRNAs, 9 maxicircle-encoded gRNAs have also been identified so far, bringing the total known gRNA genomic complement in this strain of Leishmania to 26 (Table 4).

Two minicircle-encoded gRNAs (*Lt19* and *B4*) have not yet been assigned to edited genes. It cannot be excluded that these gRNAs mediate editing of portions of the transcripts from G-rich regions G1–G5, which have not yet been investigated. It can also not yet be excluded that additional maxicircle cryptogenes exist, perhaps within the divergent region (Muhich et al., 1985), which would require additional gRNAs, nor that additional maxicircle-encoded gRNA genes exist. At least one additional maxicircle-encoded gRNA has been identified in the form of a gRNA–mRNA chimeric molecule in which the putative gRNA (*M150*) is covalently linked to a misedited *COIII* mRNA (Sturm et al., 1992).

The precise overlapping of gRNAs is strikingly economical and provides a simple explanation for the observed 3′ to 5′ polarity of editing within an editing domain: the 3′-most

Table 4. The Genomic Complexity of gRNA in L. tarentolae UC Strain

Assigned gRNAs[a]			Putative gRNAs	
	Number of gRNAs Encoded by			
Cryptogene	Maxicircle DNA	Minicircle DNA	Putative Cryptogene	Expected Number of gRNAs
COII	1	–	*G1*	12
ND7	2	–	*G2*	12
CYb	2	–	*G3*	4
MURF2	2	–	*G4*	8
RPS12	1	7	*G5*	7
COIII	–	2		
MURF4	–	6		
Total	8	15		43

[a] Three unassigned gRNAs were identified: g*Lt19* and g*B4* (minicircle encoded) and g*M150* (maxicircle encoded).

gRNA forms an anchor with the unedited mRNA sequence and then creates an mRNA anchor sequence by editing for the adjacent upstream gRNA. Each gRNA then in turn creates an anchor for the adjacent upstream gRNA. The existence of unexpected editing patterns that do not show a 3′ to 5′ polarity at junction regions in partially edited mRNAs can also be explained in terms of this model for pan editing (Sturm et al., 1992).

Our analysis of gRNA-mediated pan editing in L. tarentolae also indicates that one role of the relatively weak G–U base pairs in the gRNA–mRNA interaction is to destabilize the hybrid formed by the gRNA and the edited mRNA and allow formation of the adjacent upstream anchor hybrid with the next gRNA. This does not eliminate the possibility of the existence of an RNA helicase that could destabilize the gRNA-edited mRNA hybrid, especially in the case of the 5′ terminal editing block in a pan-edited gene or in the case of a single block of editing that is mediated by a single gRNA.

Pan editing in L. tarentolae is not limited to minicircle-encoded gRNAs, as a single maxicircle-encoded gRNA (g*RPS12-VI*) was found to initiate editing of domain III of the *RPS12* mRNA. The presence of this gRNA may be indicative of the existence of a process involving translocation of gRNA genes between the maxicircle and minicircle genomes. In addition, this gRNA has a lower relative steady-state abundance than the other 7 minicircle-encoded *RPS12* gRNAs, perhaps suggesting a rate-limiting or possible regulatory role in the overall editing process of the *RPS12* mRNA. In the case of *MURF4* editing, in which all of the gRNAs are minicircle encoded, it is of some interest that the three late-acting gRNAs (*IV–VI*), and in particular g*MURF4-V*, are of lower relative abundance than the three early-acting gRNAs (*I–III*), again perhaps suggesting a rate-limiting role in the editing process for certain late-acting gRNAs.

The intergenic G-rich regions, *G1–G6*, were hypothesized to represent pan-edited cryptogenes, by analogy with the G-rich pan-edited *ND7*, *COIII*, and *MURF4* cryptogenes in Trypanosoma brucei (Simpson and Shaw, 1989). This hypothesis has been verified for the *G6* region; transcripts of this region in L. tarentolae are pan edited to yield an mRNA encoding the amino acid sequence of a homolog of the ribosomal protein S12 family (Maslov et al., 1992). Attempts to clone and sequence mature edited transcripts from *G1–G5* in L. tarentolae, however, have proved unsuccessful (Souza et al., 1992; E. Grusyzynski, B. Niner, D. A. M., N. Sturm, and L. S., unpublished data). Partially 3′-edited transcripts from the *G1* and *G5* regions were detected that appeared to represent unexpected editing patterns, but no evidence for a mature edited species was obtained. We estimate that an additional 43 gRNAs would be required for complete productive editing of transcripts from the *G1–G5* regions (Table 4). The fact that all 12 newly discovered minicircle-encoded gRNAs mediate editing of only the *COIII*, *MURF4*, and *RPS12* (*G6*) mRNAs makes it unlikely that the 43 predicted gRNAs for the *G1–G5* putative cryptogenes actually exist in this strain, unless they all are maxicircle encoded and have eluded our analysis.

We speculate that in general the presence of gRNA genes on low copy number minicircles that are segregated randomly to daughter mitochondria (Simpson et al., 1974; Ryan et al., 1988; Birkenmeyer and Ray, 1986; Ray, 1987) may prove to have significant genetic consequences in terms of long-term stability of the editing system and the possible production of "pseudo-cryptogenes" whose transcripts do not undergo productive editing. The UC strain of L. tarentolae was isolated from its gecko host in Algeria in 1939 (Parrot and Foley, 1939) and maintained in culture ever since (Trager, 1957). We suggest that the protein products of cryptogenes *G1–G5* may not be required for life in the culture or insect phase of the life cycle. The recent identification in T. brucei of a productively pan-edited mRNA from the *G1* (*CR1*) region (Souza et al., 1992), which is only edited in the bloodstream stage, encoding a component of NADH dehydrogenase, is consistent with this hypothesis. In the absence of selective pressure during prolonged growth of the UC strain of L. tarentolae in culture, minicircles encoding the gRNA genes involved in the editing of transcripts from cryptogenes *G1–G5* might have been lost by unequal segregation during network division without loss of viability. Consistent with this hypothesis is a previous observation that the kinetoplast minicircle DNA from strains of L. tarentolae that have been recently isolated from geckos in southern France is uniformly more complex in terms of restriction patterns in acrylamide gels than minicircle DNA from the UC strain (Gomez-Eichelmann et al., 1988), indicating a larger minicircle-encoded gRNA repertoire in the former.

Experimental Procedures

Cell Cultivation, Isolation of Mitochondrial Fraction, and Isolation of Nucleic Acids

L. tarentolae (UC strain) cells were cultivated in brain heart infusion medium (Difco) supplemented with 10 μg/ml hemin (Simpson and Braly, 1970). kDNA networks were purified by centrifugation through a cesium–chloride step gradient as described previously (Simpson, 1979). kRNA was extracted from mitochondria isolated from mid–log phase cells by flotation in Renografin density gradients (Braly et al., 1974; Simpson and Simpson, 1978).

Oligonucleotides

Oligonucleotides for hybridization and primer extension assays were synthesized by standard phosphoramidite methods. The following oligonucleotides were used in this study.

Nonspecific Minicircle Probes

S-293: ACTAGGGGTTGGTGTAA.
S-486: CGCCCCTACCTAAGGGACCT.

Minicircle Class-Specific Probes

The corresponding minicircle class and the functional name of the gRNA are given in parentheses.

S-252: ACACTCTATGATTTCTCTAT (class Lt154/g*RPS12-IV*).
S-255: GAATAGTGTTTTCATCTCTC (class D12/g*COIII-I*).
S-303: TATTTTTCACTTCAACCACA (class Lt19/g*Lt19*).
S-312: ACTCTACAATGTTCCCTATA (class D3/g*MURF4-III*).
S-313: AATGTTTGTTATATTTCTCT (class B4/g*B4*).
S-508: TTTCCGCTACATTTATTTCT (class X1/g*RPS12-III*).
S-514: ATATCTATCTCTTCTATTTG (class X2/g*MURF4-VI*).
S-515: TATCTATCAATCTATGCACG (class X3/g*MURF4-I*).
S-518: AATCAATGTCTGTTCATGTT (class X4/g*COIII-II*).
S-509: AATAAAGTGTTCTTCAAGTC (class X5/g*RPS12-I*).
S-510: AATAATCAACGTTGTCTATA (class X6/g*RPS12-V*).
S-516: TATACATATTCTACTACACT (class X7/g*MURF4-V*).
S-511: TGTCTACAACCGAGAGTGAT (class X8/g*RPS12-II*).
S-517: ATTAATATCTTGTATCTATG (class X9/g*MURF4-II*).

S-512: ATATCTATTCTATACGTGTA (class X10/g*RPS12-VIII*).
S-513: TAATCTATGCATTCATCTTC (class X11/g*RPS12-VII*).
S-574: GTATTTGATCCTAAACTATT (class M1/g*MURF4-IV*).

Probe for Maxicircle-Encoded gRNA gRPS12-VI

The nucleotide positions in GenBank entry LEIKPMAX are given in parentheses.
S-582: ATTTAATTTTATATTCAAGT (16975–16994).

Minicircle Libraries

For library 1, unit-length linearized minicircles were released from networks by treatment with XmaI, an enzyme having a recognition site within the conserved region. In practice, however, some network DNA remained uncut by this enzyme even after prolonged incubation.

To generate a more comprehensive library, minicircles were linearized by digestion with MspI in the presence of ethidium bromide. There are several sites for MspI on each minicircle, and the dye restricts the cleavage to any one of these, thereby producing permuted linears (Parker et al., 1977). A dye titration showed that the relative proportion of permuted linears increased while the overall yield of released molecules decreased, and a dye concentration of 70 μg/ml was chosen for library production. Digestion of the gel-isolated permuted linears yielded the same pattern as total kDNA, indicating that the permuted linears represented a random subpopulation of the total minicircular DNA.

Minicircles were cleaved by XmaI or MspI+ethidium bromide isolated from low melting temperature agarose gels and ligated into the Bluescript SK vector, which was cleaved with XmaI or AccI, respectively, and treated with calf thymus phosphatase (Sambrook et al., 1989). MAX Efficiency DH5(alpha)-competent cells were used for the transformation according to the manufacturer's protocol. Cell lifting and colony hybridization with kinased oligomer probes were performed using standard protocols.

DNA Sequencing

Plasmid DNA was extracted by a minilysate boiling method (Sambrook et al., 1989). Double-stranded templates were sequenced using the Sequenase version 2.0 DNA sequencing kit (US Biochemical Company) according to the manufacturer's protocol.

Computer Analysis

The Genetics Computer Group program package run on the UCLA Life Sciences VAX computer was used for the analysis of nucleic acid sequences. The BESTFIT program was used with a modified SWGAP-DNA.CMP weight matrix, which differentially scores G–U and A–C base pairs (Blum et al., 1990; Van der Spek et al., 1990) as well as the canonical base pairs.

Northern Analysis

kRNA (5 μg) was separated in formaldehyde–agarose gels and blotted onto nylon filters as described previously (Blum et al., 1990; Shaw et al., 1988). Blots were probed with the oligonucleotides 5′ end labeled with T4 kinase and [γ-^{32}P]ATP (the specific activities were approximately 10^8 cpm/μg). Labeled primers were purified using Bio-Gel P-4 (Bio-Rad) spin columns.

Mapping of 5′ Ends of gRNAs

kRNA (5 μg) and the end-labeled oligonucleotide primer (0.3 ng) were sequentially annealed at 65°C, 42°C, and 25°C for 5 min each. Elongation–termination reactions were performed using a 3:1 molar ratio of dideoxynucleotides and deoxynucleotides and 30–40 U of AMV reverse transcriptase at 42°C for 30 min. Extension products were analyzed on 8% sequencing gels.

Acknowledgments

This research was supported in part by a grant to L. S. from the National Institutes of Health (AI09102). We would like to thank members of the Simpson lab for advice and discussion.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received April 30, 1992.

References

Abraham, J. M., Feagin, J. E., and Stuart, K. (1988). Characterization of cytochrome c oxidase III transcripts that are edited only in the 3′ region. Cell *55*, 267–272.

Benne, R. (1989). RNA editing in trypanosome mitochondria. Biochim. Biophys. Acta *1007*, 131–139.

Bhat, G. J., Koslowsky, D. J., Feagin, J. E., Smiley, B. L., and Stuart, K. (1990). An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell *61*, 885–894.

Bhat, G. J., Myler, P. J., and Stuart, K. (1991). The two ATPase 6 mRNAs of *Leishmania tarentolae* differ at their 3′ ends. Mol. Biochem. Parasitol. *48*, 139–150.

Birkenmeyer, L., and Ray, D. (1986). Replication of kinetoplast DNA in isolated kinetoplasts from *Crithidia fasciculata*. J. Biol. Chem. *261*, 2362–2368.

Blum, B., and Simpson, L. (1990). Guide RNAs in kinetoplastid mitochondria have a nonencoded 3′ oligo(U) tail involved in recognition of the preedited region. Cell *62*, 391–397.

Blum, B., Bakalara, N., and Simpson, L. (1990). A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell *60*, 189–198.

Blum, B., Sturm, N. R., Simpson, A. M., and Simpson, L. (1991). Chimeric gRNA–mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. Cell *65*, 543–550.

Braly, P., Simpson, L., and Kretzer, F. (1974). Isolation of kinetoplast-mitochondrial complexes from *Leishmania tarentolae*. J. Protozool. *21*, 782–790.

Decker, C. J., and Sollner-Webb, B. (1990). RNA editing involves indiscriminate U changes throughout precisely defined editing domains. Cell *61*, 1001–1011.

Gomez-Eichelmann, M. C., Holz, G., Jr., Beach, D., Simpson, A. M., and Simpson, L. (1988). Comparison of several lizard *Leishmania* species and strains in terms of kinetoplast minicircle and maxicircle DNA sequences, nuclear chromosomes, and membrane lipids. Mol. Biochem. Parasitol. *27*, 143–158.

Holbrook, S. R., Cheong, C., Tinoco, I., and Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating a track of non Watson-Crick base pairs. Nature *353*, 579–581.

Kidane, G., Hughes, D., and Simpson, L. (1984). Sequence heterogeneity and anomalous electrophoretic mobility of kinetoplast minicircle DNA in *Leishmania tarentolae*. Gene *27*, 265–277.

Koslowsky, D. J., Bhat, G. J., Read, L. K., and Stuart, K. (1991). Cycles of progressive realignment of gRNA with mRNA in RNA editing. Cell *67*, 537–546.

Maslov, D. A., Sturm, N. R., Niner, B. M., Gruszynski, E. S., Peris, M., and Simpson, L. (1992). An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. Mol. Cell. Biol. *12*, 56–67.

Muhich, M., and Simpson, L. (1986). Specific cleavage of kinetoplast minicircle DNA from *Leishmania tarentolae* and identification of several additional minicircle sequence classes. Nucl. Acids Res. *14*, 5531–5556.

Muhich, M., Neckelmann, N., and Simpson, L. (1985). The divergent region of the *Leishmania tarentolae* kinetoplast maxicircle DNA contains a diverse set of repetitive sequences. Nucl. Acids Res. *13*, 3241–3260.

Parker, R. C., Watson, R. M., and Vinograd, J. (1977). Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc. Natl. Acad. Sci. USA *74*, 851–855.

Parrot, L., and Foley, H. (1939). Sur la frequence de la leishmaniose du gecko dans le Sud oranais. Arch. Inst. Pasteur Alger. *17*, 231–232.

Pollard, V. W., Rohrer, S. P., Michelotti, E. F., Hancock, K., and Hajduk, S. L. (1990). Organization of minicircle genes for guide RNAs in Trypanosoma brucei. Cell *63*, 783–790.

Ray, D. (1987). Kinetoplast DNA minicircles: high copy number mitochondrial plasmids. Plasmid *17*, 177–190.

Ray, D. (1989). Conserved sequence blocks in kinetoplast DNA minicircles from diverse species of trypanosomes. Mol. Cell. Biol. *9*, 1365–1367.

Ryan, K., Shapiro, T., Rauch, C., and Englund, P. (1988). Replication of kinetoplast DNA in trypanosomes. Annu. Rev. Microbiol. *42*, 339–358.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Shaw, J. M., Feagin, J. E., Stuart, K., and Simpson, L. (1988). Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conserved amino acid sequences and AUG initiation codons. Cell *53*, 401–411.

Simpson, L. (1979). Isolation of maxicircle component of kinetoplast DNA from hemoflagellate protozoa. Proc. Natl. Acad. Sci. USA *76*, 1585–1588.

Simpson, L. (1987). The mitochondrial genome of kinetoplastid protozoa: genomic organization, transcription, replication and evolution. Annu. Rev. Microbiol. *41*, 363–382.

Simpson, L. (1990). RNA editing—a novel genetic phenomenon. Science *250*, 512–513.

Simpson, L., and Braly, P. (1970). Synchronization of *Leishmania tarentolae* by hydroxyurea. J. Protozool. *17*, 511–517.

Simpson, L., and Shaw, J. (1989). RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. Cell *57*, 355–366.

Simpson, L., and Simpson, A. M. (1978). Kinetoplast RNA of Leishmania tarentolae. Cell *14*, 169–178.

Simpson, L., Simpson, A., and Wesley, R. (1974). Replication of the kinetoplast DNA of *Leishmania tarentolae* and *Crithidia fasciculata*. Biochim. Biophys. Acta *349*, 161–172.

Souza, A. E., Myler, P. J., and Stuart, K. (1992). Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited, developmentally regulated, and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. Mol. Cell. Biol. *12*, 2100–2107.

Stuart, K. (1991). RNA editing in trypanosomatid mitochondria. Annu. Rev. Microbiol. *45*, 327–344.

Sturm, N. R., and Simpson, L. (1990a). Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from Leishmania tarentolae mitochondria: RNA editing intermediates. Cell *61*, 871–878.

Sturm, N. R., and Simpson, L. (1990b). Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. Cell *61*, 879–884.

Sturm, N. R., and Simpson, L. (1991). *Leishmania tarentolae* minicircles of different sequence classes encode single guide RNAs located in the variable region approximately 150 bp from the conserved region. Nucl. Acids Res. *19*, 6277–6281.

Sturm, N. R., Maslov, D. A., Blum, B., and Simpson, L. (1992). Generation of unexpected editing patterns in Leishmania tarentolae mitochondrial mRNAs: misediting produced by misguiding. Cell *70*, this issue.

Trager, W. (1957). Nutrition of a hemoflagellate (*Leishmania tarentolae*) having an interchangeable requirement for choline or pyridoxal. J. Protozool. *4*, 269–276.

Van der Spek, H., Speijer, D., Arts, G.-J., Van den Burg, J., Van Steeg, H., Sloof, P., and Benne, R. (1990). RNA editing in transcripts of the mitochondrial genes of the insect trypanosome *Crithidia fasciculata*. EMBO J. *9*, 257–262.