A Model for RNA Editing in Kinetoplastid Mitochondria: "Guide" RNA Molecules Transcribed from Maxicircle DNA Provide the Edited Information

Beat Blum,* Norbert Bakalar,* and Larry Simpson†
†Molecular Biology Institute
University of California
Los Angeles, California 90024

Summary
A class of small RNA molecules possibly involved in RNA editing is present in the mitochondrion of Leishmania tarentolae. These "guide" RNA (gRNA) molecules are encoded in intergenic regions of the mitochondrial maxicircle DNA and contain sequences that represent precise complementary versions of the mature mRNAs within the edited regions. In addition, the 5' portions of several gRNAs can form hybrids with mRNAs just 3' of the preedited region. A model is presented in which a partial hybrid formed between the gRNA and preedited mRNA is substrate for multiple cycles of cleavage, addition or deletion of uridylates, and religation, eventually resulting in a complete hybrid between the gRNA and the mature edited mRNA.

Introduction
RNA editing of the transcripts of mitochondrial cryptogenes in the kinetoplastid protozoa (reviewed in Simpson and Shaw, 1989; Benne, 1989) has posed an apparent paradox in relation to the central dogma of molecular genetics, which specifies that the genetic information of an organism is exclusively encoded in the nucleic acid, and this information determines the primary amino acid sequence of the proteins. In the case of RNA editing in kinetoplastid mitochondria, uridines are added to and deleted from mRNAs for at least six mitochondrial genes at precise sites within coding regions, thereby giving rise to new amino acid sequences and rendering the mRNAs translatable by uncovering cryptic reading frames and supplying translation initiation codons. The major problem relates to the origin of the precise sequence information added in the editing process, in so far as no nucleic acid templates have yet been detected containing the edited sequences, and no consensus sequence or structural signals have been uncovered that could provide the required information encoded in the nucleic acid. In this paper we show that there are small transcripts, which we call guide RNAs or gRNAs, derived from scattered intergenic regions of the Leishmania tarentolae maxicircle genome, which encode the necessary information for RNA editing in terms of a proposed posttranscriptional process. This model, which is consistent with all known facts regarding RNA editing, resolves the problem of the origin of edited sequence information in a manner that is novel but consistent with the basic tenets of molecular genetics.

Results
Short Sequences Complementary to Edited RNAs Are Encoded in the Maxicircle DNA of L. tarentolae
A computer search was performed on the known 21 kb L. tarentolae maxicircle sequence for short DNA sequences that could be transcribed into RNAs complementary to edited regions of mRNAs. We allowed G-U base pairs, which are known to exist in tRNA and rRNA molecules (Glotz et al., 1981), and did not allow gaps in the alignments. The computer search identified seven short sequences, mostly in intergenic regions of the maxicircle, that satisfied these criteria for the edited regions of cytochrome b (CYb), maxicircle unidentified reading frame 2 (MRF2), maxicircle unidentified reading frame 3-5' end (MRF3(5')), cytochrome oxidase subunit II frameshift (COII), and maxicircle unidentified reading frame III frameshift (MRF3(FS)) (Figure 1). In line with the proposed model described below, which implicates transcripts from these regions in the process of RNA editing, we shall call these DNA sequences guide RNA (gRNA) genes. It should be noted that no completely satisfactory DNA sequences complementary to edited RNA sequences have yet been found for the cytochrome oxidase subunit III (COIII) edited region. However, a set of six overlapping sequences for the COIII gene have been identified by computer analysis (data not shown), but there are several mismatches that make these less certain than the already identified gRNA sequences.

For the CYb and MURF2 genes, there are two overlapping gRNA sequences that together cover the entire edited regions. We define the 3'-most gRNA sequence (with reference to the mRNA polarity) as block I (gRNA-I) and the 5'-most gRNA sequence as block II (gRNA-II) in each case. The MRF3(5'), MURF3(FS), and the COIII edited regions appear to be covered by single gRNA sequences in each case.

As shown in the alignments in Figure 1, the length of perfect complementarity of the DNA sequences (sense strand) with edited mRNA sequences without gaps or mismatches varies from 14 nucleotides for the MRF2-I sequence to as many as 51 nucleotides for the CYb-II sequence. An additional characteristic of the CYb-I and MURF2-I sequences, and also the MURF3(5'), MURF3(FS), and COIII(FS) sequences, is an extension of the perfectly matched base pairing in a 3' direction (with reference to the mRNA polarity) downstream of the preedited region for up to 16 nucleotides (in the case of the CYb-I sequence). The existence of these DNA sequences which exhibit perfect complementarity with fully edited mRNA sequences strongly suggested that a class of mitochondrial RNAs (gRNAs) exist that are transcribed from these sequences.

The locations of the gRNA sequences are shown in the linearized maxicircle genomic map in Figure 2. The gRNAs are shown in boxes and the direction of transcrip-
Figure 1. Identification of Complementary Edited gRNA Sequences in Maxicircle DNA Alignments of edited RNA sequences of several mitochondrial mRNAs with corresponding gRNA sequences (sense strand of DNA). The added uridines in the mRNA are shown as u, and the "guide" adenine and guanine nucleotides in the gRNA are shown as a and g. Standard base pairs are indicated with solid lines, G-U base pairs with asterisks. The uridine residues that are unpaired with guide a's or g's are deleted in the mature edited mRNA are indicated as superscript u. The major 3' ends of the gRNAs are indicated by arrows, and the tentative oligo(U) transcription termination signals of the gRNAs, as deduced from this study (see Figures 5 and 6 below), are indicated by underlining. Some flanking sequences are shown both 5' and 3' of the gRNA regions, to show the limited base complementarity with edited RNA upstream and downstream of the gRNA sequence. Overlap regions of the gRNA-I and gRNA-II sequences for the Cyb and MURF2 mRNAs are indicated by brackets connected by horizontal lines. The location of the COII gRNA at the 3' end of the COII mRNA is indicated by dots connecting the two sequences. This also illustrates the fold-back hybridization model for this mRNA-gRNA hybrid. The UAG termination codon for the COII mRNA that overlaps the gRNA is indicated by a box. The maxicircle (entry LEIKPMAX in GenBank) nucleotide localizations of the 5' ends and the polarities of the gRNAs (+ or − with reference to the polarity of the 9S and 12S rRNA genes) are as follows:

Figure 2. Localization of gRNA Sequences in Genomic Map of Maxicircle The sites of the seven identified gRNA sequences are indicated in the maxicircle genomic map, which is linearized at the single EcoRI site (E). The polarities of the gRNA sequences (in boxes) and the unedited structural genes and cryptogenes are shown by arrows. The MspI (M) and HhaI (H) restriction sites are also indicated together with the size of each restriction fragment produced by the double digest, and a specific oligomer probe for each gRNA is shown below the unique restriction fragment that should exhibit hybridization with that probe if the gRNA sequence is not repeated elsewhere in the genome (see Figure 4). The probe specificities are as follows: S-200 = MURF3(5), S-198 = Cyb-II, S-199 = Cyb-I, S-202 = MURF2-II, S-201 = MURF2-I, S-197 = Cyb-I, S-218 = MURF3(F5).

tion of the gRNAs and the cryptogenes themselves indicated by arrows. All of the gRNA sequences are located within intergenic regions at a variable distance from the homologous unedited cryptogene sequence, except for the COII gRNA sequence, which is located overlapping the 3' end of the COII gene in the same polarity as the gene; the MURF2-I gRNA sequence, which is located within the COII gene but on the opposite strand; and the Cyb-II gRNA sequence, which is located overlapping the 3' end of the 9S rRNA gene but on the opposite strand.

Detection of gRNAs by Northern Blot Hybridizations of Kinetoplast RNA Oligonucleotide probes were generated for the seven identified gRNA sequences, and these probes were used with Northern blots of total mitochondrial RNA. In each case the probe hybridized with a small RNA (gRNA) that migrated ahead of tRNA in agarose and especially in acrylamide gels (Figure 3). Maxicircle transcripts of unknown function with this electrophoretic migration behavior have been described previously as contaminants of...
Figure 3. Detection of gRNAs by Northern Hybridization of Mitochondrial RNA

(A) Hybridization of 5' end-labeled oligonucleotide probes to gRNA sequences. gRNA sequences were separated on a formaldehyde-1.5% agarose gel. The indicated reference nucleotide sizes were derived from rehybridization of the filters with labeled pLt120 maxicircle DNA and from the position of tRNA seen in the stained gel.

(B) Acrylamide-7M urea gel. Electroblots were hybridized with several of the same probes used in (A). One blot was hybridized with a labeled cloned kinetoplast DNA minicircle probe, pLt19. The position of tRNA as obtained from the stained gel is indicated.

(C) Rehybridization of several of the stripped blots from (A) with labeled oligonucleotide probes for the edited mRNAs (Shaw et al., 1989; Shaw et al., 1988). Lane 5 was rehybridized with an oligonucleotide probe for the unedited mRNA. Lane 6 was rehybridized with the pLt19 minicircle probe. The low molecular weight bands in lanes 2 and 3 represent residual gRNA bands remaining from the initial hybridizations due to incomplete stripping from the Nytran filters. Sizes of the bands are shown in nucleotides.

mitochondrial tRNA preparations (Simpson et al., 1989). We estimate the size of these RNAs to be less than 80 nucleotides. An accurate size must be determined by direct sequence analysis, and this is in progress. A certain size heterogeneity is apparent with the CYb-I, CYb-II, MURF3(5'), and MURF2-II gRNAs in acrylamide (Figure 3B). This size heterogeneity, in the case of the CYb-II and the MURF3(5') gRNAs, has also been visualized by two-dimensional acrylamide-urea gel electrophoresis (10% acrylamide and 20% acrylamide, 7 M urea) as a series of spots running along a diagonal (A. Simpson and L. S., unpublished data). In addition, the CYb-II gRNA has been isolated by hybrid selection, labeled with 32P at the 5' end, and shown to migrate in acrylamide-urea as 12–13 bands 1 nucleotide apart, differing in relative abundance (B. B. and L. S., unpublished data).

The relative abundance of the gRNAs identified by the Northern blot analysis in Figure 3 varied from the highly abundant CYb-II and MURF3(5') gRNAs to the less abundant COII and MURF2-I gRNAs (Figure 3C). The CYb-II and MURF3(5') gRNAs, as shown on Figure 3C, have been processed to the lower molecular weight forms, but it is also possible that they represent aberrant transcript products.

A small gRNA was also identified for the COIII gene by use of an oligomer probe (S74) derived from the edited mRNA sequence, in spite of the lack of identification of a corresponding gRNA sequence in the known maxicircle DNA sequence.

As a control for the quality of the mitochondrial RNA preparation, we checked for the presence of mature edited transcripts in these blots by rehybridization with oligomer probes for the mature edited sequences. As shown in Figure 3C, the 1300 nucleotide CYb edited RNA (Feagin et al., 1988), the 1200 nucleotide MURF3 edited RNA, and the 1000 nucleotide MURF2 edited RNA were detected (Simpson et al., 1985). The COII edited probe hybridized with the expected 750 nucleotide edited RNA (Simpson et al., 1985), whereas an unedited probe (in the frameshift region) hybridized with a 1600 nucleotide RNA. It should be noted that the low molecular weight bands in lanes 2 and 3 are derived from residual gRNA bands from the first hybridization and do not represent antisense gRNAs.

Identification of gRNA Sequences as Single-Copy Elements in Maxicircle DNA by Hybridization

Purified EcoRI-linearized maxicircle DNA was digested with MspI and HhaI, which should release each predicted gRNA gene on a distinct restriction fragment (Figure 2). Each gRNA-specific oligomer hybridized with the expected maxicircle DNA fragment, as shown in Figure 4. There was no hybridization with minicircle DNA, which is present as a contaminant of the purified maxicircle preparation. These results indicate that the identified gRNA sequences are probably single-copy elements in the maxicircle genome and do not occur in minicircle DNA.

The lack of previous detection of the gRNA sequences in maxicircle DNA or RNA by hybridization (reviewed in Simpson and Shaw, 1989; Benne, 1989) results primarily from the presence of mismatches due to the use of antisense sequences as probes derived directly from edited RNA. The gRNA sequences found in this study show distinct base differences from a perfect antisense edited sequence in that a G residue can be present instead of an
A residue to pair with a U residue in the mRNA, or a U residue can be present instead of a C residue to pair with a G residue in the mRNA. Both types of substitutions are frequent enough to prevent the detection of gRNA sequences using perfect antisense edited probes. Additional reasons for the failure to detect gRNA sequences previously include the fact that in at least two cases the gRNA sequences do not cover the entire edited regions and that the gRNA sequences (and gRNAs) are quite short and do not include substantial complementary flanking sequences.

Localization of the 5′ Ends of gRNAs

Determination of the 5′ ends of the identified gRNAs was performed by primer extension analysis. As shown in Figure 5A, a definite 5′ end was identified for each gRNA, except for MURF2 gRNA-I, for which no extension product was observed. The latter result may be due to a matching of the 3′ end of the probe with the 5′ end of the gRNA. In the case of the CYb gRNA-I, there were two extension products, indicating a minor 5′ end 1 nucleotide downstream of the major 5′ end. For two of the gRNAs, CYb-II and MURF3(5′), the actual 5′ terminal sequences were obtained by adding dideoxynucleotides to primer extension assays (Figure 5B). Both sequences agreed exactly with those predicted from the maxicircle sequence.

In the case of the COII gRNA sequence, the 5′ end of the gRNA sequence is 5 nucleotides upstream of the UAG termination codon for the COII gene (Figure 1), suggesting that processing of this gRNA would interfere with the translation capacity of the COII mRNA. In this regard, a longer exposure of the COII gRNA primer extension showed two high molecular weight extension products (data not shown), which may represent the 5′ ends of the larger low abundance species detected on Northern blots (Figure 3). The size of the largest extension product is consistent with a transcript that includes the entire COII gene in addition to the gRNA.

The localizations of the major 5′ ends of the gRNAs as derived from the primer extension experiments are indicated by arrows in Figure 1. The presence of several weak primer extension products 1–4 bases larger than the major 5′ end products is possibly due to a very low frequency of heterogeneity at the 5′ end. The fact that the major 5′ ends are located no more than 8 nucleotides from the end of the regions of perfect complementarity with the edited sequences in every case provides additional evidence that the gRNAs are involved in the process of RNA editing.

Localization of the 3′ End of CYb gRNA-II by S1 Protection and Presence of a Possible 3′ Consensus Sequence

To determine the length of the transcribed sequence present on the gRNA, S1 protection analysis was performed on the most abundant gRNA species, CYb gRNA-II (Figure 6). To avoid “breathing” of the ends and cleavage at A-U base pairs, S1 digestion was performed at 55°C. These digestion conditions revealed a single major band of 33 nucleotides (arrow) at the higher S1 nuclease concentration, which maps the 3′ end of the transcribed region of CYb gRNA-II to the last of a stretch of 5 uridine residues 55 nucleotides downstream of the identified 5′ end (Figure 7, arrow). The presence of the three to four additional protected minor bands in the S1 nuclease protection experiment (Figure 6) could reflect “nibbling” by S1 nuclease or could represent an actual 3′ heterogeneity of the gRNA. Additional evidence for 3′ heterogeneity of the CYb-II and MURF3(5′) gRNAs was provided by the two-dimensional
Figure 6. Localization of the 3' End of CYb gRNA-II by S1 Nuclease Protection Analysis

The labeled oligonucleotides were annealed with total kinetoplast RNA and digested with S1 nuclease at 150 U/ml and 500 U/ml at 5°C. The major protected band is indicated by an arrow. The sequence of CYb gRNA-II and the localization of the site corresponding to this protected band are shown in Figure 7. "Ref" = a primer extension used as a size marker.

Discussion

We have identified a class of small mitochondrial RNAs, the gRNAs, encoded in scattered intergenic regions of the L. tarentolae maxicircle, which contain the precise sequence information necessary for RNA editing of at least four of the five genes known to be edited in this organism. Seven maxicircle sequences have been identified by computer analysis and the corresponding gRNAs detected by Northern blot analysis. Six of these gRNAs were shown to have homogeneous 5' termini, and in two cases the 5' terminal sequences have been determined directly. For one gRNA, the 3' end of the transcribed region was tentatively mapped. In view of these findings, we would like to propose a model that could account for the involvement of gRNAs in RNA editing.

A Model for the Involvement of gRNAs in RNA Editing

As described schematically in Figure 8, the process is initiated when a partial hybrid is formed between the 5' portion of the gRNA and the complementary sequence on the preedited mRNA. Since in some cases the length of the 3' duplex region is quite small, protein factors may be involved also in stabilizing the initial binding of the gRNA. A putative RNAase P-like endoribonuclease (Doersen et al., 1985; Chang and Clayton, 1989; Nichols et al., 1988; Morales et al., 1989), which is part of a multienzyme editing complex that may also include a terminal uridylyltransferase (TUTase), exonuclease, and ligase, then cleaves the mRNA exactly 3' of the first base that is not paired with the gRNA, releasing a free 3' OH. Initiation of the process could also be dependent on sequence or secondary structure of the preedited mRNA, which might enable the editing complex to bind prior to the first gRNA-driven cleavage. Two alternative reactions can then occur: if the 3' terminal base at the cleavage site is not a uridine residue, a uridine is added by a 3' TUTase (= uridine addition), or if the 3' terminal base is a uridine residue, this uridine is removed by a 3' exonuclease (= uridine deletion). We assume that the known mitochondrial 3' TUTase (Bakalar et al., 1989) is involved in the uridine addition process, but it is possible, of course, that some yet undescribed 5' TUTase could add uridines to the 5' end of the downstream fragment. The uridine addition or deletion is followed by immediate religation of the two molecules, by acrylamide gel electrophoresis results mentioned above and by the results indicating a size heterogeneity of hybrid-selected 5' end-labeled CYb-II gRNA. It should be noted that a nonencoded extension of the transcribed sequence at the 3' terminus can not be excluded by our experiments.

Although S1 protection data have not yet been obtained for the 3' ends of other gRNAs, a similar stretch of 3-5 encoded uridines can be found on all seven gRNA sequences just downstream of the adenosine guiding the last uridine addition into the preedited region of the mRNA (Figure 7). This oligo(U) sequence is reminiscent of transcription termination signals for several DNA-dependent RNA polymerases (Bogenhagen and Brown, 1981). In addition, a tetrameric consensus sequence (AAUA) occurs just upstream of the oligo(U) sequence (Figure 7). The localization of the putative 3' end of the transcribed region of CYb gRNA-II and possibly of all seven identified gRNAs is within a few nucleotides of the end of the region of perfect complementarity with the edited sequence.

Table 1: Possible 3' Ends of Transcribed Regions of All Seven gRNAs with Conserved Consensus Sequence

Alignment of gRNA sequences at 3' ends, showing the conserved AAUA consensus sequence and the putative terminal oligo(U) sequence. The 3' ends of the transcribed portions of the gRNAs were tentatively assigned to be at the same relative location (arrow) as that of CYb gRNA-II, which was determined by S1 protection analysis (Figure 6), assuming that the oligo(U) sequences are encoded and not due to posttranscriptional addition. Matching nucleotides are boxed and the consensus sequence is shown below the alignments.
Figure 8. Schematic Diagram of Model for Involvement of gRNAs in RNA Editing

(A) U Addition. Addition of several uridylates to the preedited region of the CYb mRNA as guided by the CYb gRNA-I. A single complete cycle is given in detail, and then several additional cycles of addition are shown until the final mature edited mRNA sequence is obtained. The arrow above the sequence indicates the site of the initial cleavage. Solid lines indicate A-U or G-C base pairs and asterisks indicate G-U base pairs. The dotted lines indicate possible upstream base pairs at the editing junction. Added uridines are shown as u and the guide adenine and guanine nucleotides in the gRNA as a and g.

(B) U Deletion. Deletion of several uridylates from the preedited region of the MURF2 mRNA as guided by the MURF2 gRNA-II. The initial two cycles of uridylate deletion are presented in detail, and then the final two cycles are shown yielding the mature edited mRNA. The circled u's indicate uridines that are mismatched to the gRNA sequence and will be deleted. Note that the mRNA sequence shown includes the single added u residue from block I editing.

Our evidence has not excluded the possibility that the gRNAs may have a heterogeneous 3' oligo(U) extension due to activity of the mitochondrial 3' TUTase (Bakalara et al., 1989). This must be determined by direct sequencing of the gRNAs. The presence of oligo(U) could aid in increasing the stability of the initial hybrid formed between the 3' terminus of the gRNA and the preedited mRNA by forming an additional hybrid with the GA-rich preedited region, with the nonhybridizing "guide" gRNA sequence in the form of a loopout. By analogy with the RNA splicing paradigm (Konarska and Sharp, 1987; Reed et al., 1988) it is also possible that the gRNAs are present in the form of ribonucleoprotein particles.

Evidence for the gRNA Editing Model

The main evidence involves the precise complementarity of the gRNA sequences with the edited mRNA sequences which had been held together by secondary structure and/or by interaction with protein factors in the editing complex. In the case of uridine addition, the initial hybrid between the gRNA and the mRNA is elongated by at least 1 bp. In the case of uridine deletion, one to several rounds, depending on the number of uridines to be deleted, are required to elongate the hybrid by at least 1 bp. Multiple cycles of this process will lead, in a 3' to 5' progressive fashion, to fully edited mRNA, at which point further cleavage is excluded by the absence of a mismatch structure in the complete hybrid formed between the gRNA and the mRNA. The gRNA–mRNA hybrid is then opened by, for example, the binding of a ribosome to the newly created AUG codon, and the gRNA is liberated to hybridize to another preedited mRNA. This would allow a positive translational control of the extent of editing by regulation of the concentration of free gRNA. Since the gRNA might also hybridize to a mature edited mRNA and be removed from reinitiating the editing process, this could provide a negative feedback mechanism to regulate the level of RNA editing.

Our evidence has not excluded the possibility that the gRNAs may have a heterogeneous 3' oligo(U) extension due to activity of the mitochondrial 3' TUTase (Bakalara et al., 1989). This must be determined by direct sequencing of the gRNAs. The presence of oligo(U) could aid in increasing the stability of the initial hybrid formed between the 3' terminus of the gRNA and the preedited mRNA by forming an additional hybrid with the GA-rich preedited region, with the nonhybridizing "guide" gRNA sequence in the form of a loopout. By analogy with the RNA splicing paradigm (Konarska and Sharp, 1987; Reed et al., 1988) it is also possible that the gRNAs are present in the form of ribonucleoprotein particles.
(assuming G-U base pairs). There is some additional circumstantial evidence for our posttranscriptional model in that several of the enzyme activities predicted by the model have already been shown to exist. A 3′ TUTase and an RNA ligase have been shown to be localized within the mitochondrion of L. tarentolae (Bakalara et al., 1989). The TUTase adds up to 8 uridylic acid in a non sequence specific manner to the 3′ OH of an RNA molecule, and this represents the maximum number of uridine residues added as a consequence of in vivo editing (Simpson and Shaw, 1989). We propose initially, as the simplest hypothesis, that only a single uridine is added (or deleted) at each site prior to religation. However, the real situation may be more complex, and there may be an equilibrium between addition of uridines, deletion of uridines, and rapid religation at a single site.

Another activity predicted by the model is a ribonuclease that can cut 3′ of the first unpaired base in the mRNA. RNAase P-like nucleases have been shown to be involved in processing of mitochondrial RNA in mammalian cells and yeast (Doersen et al., 1985; Chang and Clayton, 1989; Nichols et al., 1988; Morales et al., 1989). We have preliminary evidence for the presence of a sequence-specific ribonuclease in detergent lysates of mitochondria from L. tarentolae that can be detected by an accumulation of cleaved products after inhibition of the mitochondrial TUTase with heparin (N. B., A. Simpson, and L. S., unpublished data).

The model also predicts the existence of a 3′ exonuclease activity, for which there is yet no direct evidence. A uridine-specific 3′ exonuclease could explain the specific deletion of uridine residues in several mature edited mRNAs and in the partially edited COIII mRNAs detected by PCR amplification (N. Sturm and L. S., unpublished data). However, a low frequency of the partially edited COIII mRNAs contain specific deletions of purines, and this would suggest a nonspecific exonucleolytic activity. This must be resolved by further work. The putative exonuclease activity could also serve a proofreading function for the addition of excessive uridine residues by the TUTase.

The apparent 3′ to 5′ polarity of the in vivo editing process (Simpson and Shaw, 1989) is readily explained by the model, in that the consecutive corrections of the mismatched hybrids between the preedited mRNA and the gRNA occur in a 3′ to 5′ direction on the mRNA. The appearance of partially edited RNAs with extra uridines at normal sites and at sites that are not normally edited (Abraham et al., 1988; N. Sturm and L. S., unpublished data) and with deletions of purine residues (N. Sturm and L. S., unpublished data) can also be explained if we assume that incorrect gRNAs can occasionally form functional mismatch hybrids, and that the partially edited RNAs are eventually reprocessed with the correct gRNA, unless deletion of purines has occurred. It is also possible that such a complex process is inherently imprecise, leading to a low frequency of imprecisely edited molecules that may be either reprocessed or degraded.

In view of our yet imprecise knowledge of the 3′ ends of the gRNAs, the model also thereby lacks preciseness in this region. The single example of a deletion of encoded uridine residues occurs at the 5′ end of the edited region of block II of the MURF2 mRNA and can be explained by an absence of complementary nucleotides in the gRNA sequence at that point (Figures 1 and 8B), which is at the tentative 3′ end of the gRNA as inferred from the S1 nuclease protection experiment in Figure 6 and the conserved alignment sequence in Figure 7. However, the model would in a similar way predict the deletion or addition of several additional uridine residues immediately upstream of the edited regions of several additional genes (Figure 1), a prediction for which there is apparently no mRNA sequence evidence. In the case of the CYb-II editing region, however, a reexamination of the published edited mRNA sequencing ladders (Figure 1 in Feagin et al., 1988) shows that a uridine residue that is predicted to be deleted by 3′-terminal gRNA–mRNA mismatch hybridization (circled U in Figure 1) is indeed deleted in at least half of the CYb mRNA molecules, giving rise to a precisely predicted pattern of U-A double bands in the sequencing ladder extending 5′ of the deletion, a pattern that is produced by two overlapping ladders 1 nucleotide out of register. This uridine deletion would have no effect on the function of the mRNA since the AUG translation initiation codon has already been created by downstream editing. The immediately upstream pair of uridines in the mRNA is apparently not deleted, and this remains to be explained. The fact that only a portion of the CYb mRNAs contain a deleted uridine further suggests that the observed 3′ end heterogeneity of the transcribed region indicated by the S1 nuclease protection experiment in Figure 6 may be real and not a breathing artifact.

Another published observation may also be relevant to the gRNA editing model. Specific extension products were seen in edited mRNA sequencing ladders 7 nucleotides downstream of the MURF3 (FS) editing site and 5 nucleotides downstream of the COII editing site (Figure 1 in Shaw et al., 1989). These products actually localize to the initiation of the gRNA–mRNA hybrids in each case, suggesting that the hybrid assumed to be present in the kRNA preparation may cause some interference with reverse transcriptase reading of the mRNA template.

The Involvement of Two gRNAs for a Single Edited Region

In the case of the genes with two gRNAs, CYb and MURF2, there are two possibilities. Either the 5′ end of the gRNA-II forms a hybrid with already partially edited mRNA and proceeds to guide the editing of the 5′ portion of the mRNA, or block II and block I editing processes proceed independently. In this regard, the initial hybrid that can be formed between the gRNA-II molecules for both genes has a longer perfect base pairing stretch with downstream unedited block I mRNA sequence than with edited block I sequence in each case (Figure 1). This would predict a kinetics of editing in which partially edited molecules exist that are edited in the upstream block II region and not in the downstream block I region. This is a testable hypothesis.

The apparent differences in relative abundances of the two gRNAs for CYb and MURF2 may be related to the
block I–block II type of editing proposed for these mRNAs. The longer lengths of the initial 3' hybrids occurring with block I gRNAs may be important for binding of the editing complex and initiation of the process, whereas the shorter lengths of the 3' block II gRNA–mRNA hybrids may be sufficient for elongation of editing, and the higher relative abundance of block II gRNAs may aid in the formation of these less stable hybrids.

The COII mRNA May Be Edited As a Precursor by Fold-Back of the gRNA Sequence Present in the 3' Portion of the mRNA

The COII gRNA represents an apparent special situation in that processing of the gRNA from the precursor COII mRNA would remove the translation termination codon of the mRNA. We propose that editing of the COII mRNA possibly occurs on the 1600 nucleotide transcript, which contains the COII gene and the gRNA sequence at the 3' end, by means of a fold-back of the gRNA sequence (shown in Figure 1). This would increase the efficiency of editing in spite of the relatively short length of the complementary sequence in this case and is consistent with the relatively low abundance of the processed lower molecular weight COII gRNA. In addition, the edited mRNA has a length of 750 nucleotides (Figure 3), suggesting that RNA processing is closely linked to RNA editing in this case. The 750 nucleotide COII edited mRNA contains the COII gRNA sequence, as shown by the appearance of a 750 nucleotide band in the Northern blot in Figure 3A, which was hybridized with a gRNA probe. This is also consistent with the fold-back model.

Edited Genes for Which gRNA Sequences Have Not Yet Been Found

The only edited gene in L. tarentolae for which completely satisfactory gRNA sequences have not yet been identified is the 5' edited COII mRNA. The COII situation clearly requires additional work, but appears to be similar to the others in that a probe for complementary edited sequence hybridized with a small transcript in Northern blots. Additional edited RNAs probably exist in L. tarentolae, such as the MURF4 gene (K. Stuart, personal communication) and the G-rich intergenic regions (Simpson and Shaw, 1989), and these remain to be analyzed in terms of corresponding gRNA sequences. In addition, the edited cryptogenes of the related kinetoplastids, Crithidia fasciculata and Trypanosoma brucei, remain to be analyzed in this manner. Clearly, the pan-edited genes of T. brucei represent the greatest challenge for analysis in terms of the gRNA model. It is of course also possible that pan-editing involves a different mechanism than that proposed here for internal editing and 5' editing in L. tarentolae.

Testing the gRNA Model

The gRNA model for RNA editing is possibly testable indirectly by the phylogenetic method involving compensatory base substitutions used to test the existence of duplex regions in ribosomal RNA (Noller, 1984), since in some cases different patterns of editing give rise to identical amino acid sequences in different species. The model could also be directly testable with an in vitro RNA editing system. Work on such a system is in progress (N. B., A. Simpson, E. Gruszynski, and L. S., unpublished data). A rigorous genetic test of the model must await the development of a selectable transformation and gene replacement system for kinetoplastids and kinetoplastid mitochondria (Bollettato and Croc, 1980; D. Wirth, personal communication).

Conclusions

The concept of a split gene is not novel, but the concept of a split gene in which one portion produces an RNA that aids in the modification of the RNA sequence from the first portion to produce a translatable mRNA sequence is novel. There are other examples of widely separated genetic units that coevolve, such as those regions of RNA genes that are separated by hundreds of nucleotides, but which base-pair in the mature molecule (Glotz et al., 1991). However, the gRNA sequences must coevolve with the preedited sequences of the gene with reference to the functional amino acid sequences, not merely with reference to maintenance of a double helical region, as in a molecule such as rRNA. This implies a strict selection at the protein level that maintains the integrity of the gRNA corrections in the face of mutational pressure on the gene itself, which may be substantial in mitochondria (Wilson et al., 1985).

The correction of sequences of RNA molecules by this proposed mismatch hybridization model has several evolutionary implications. It opens the possibility of a generalized modification of the genetic information of an RNA molecule and has a possible relevance to the maintenance of functional sequence information in the "RNA world" that is hypothesized to have existed prior to the evolution of DNA (Gilbert, 1986).

The presence of split genes and the editing process opens the door to another level of translational regulation of mitochondrial bioynthesis by controlling the abundance of translatable mRNAs. In fact, a developmentally controlled regulation of RNA editing and thereby mitochondrial translation was shown to occur for two cryptogenes (Cyb and COII) during the life cycle of T. brucei (Feagin and Stuart, 1988). The mechanism of this regulation is unknown.

There are many questions yet to be answered regarding details of the gRNA model of RNA editing, but we suggest that this model provides a novel solution to the problem of the source of the edited sequence information in a manner consistent with the central tenets of molecular genetics.

Experimental Procedures

Cell Culture and Mitochondrial Isolation

L. tarentolae (UC strain) cells were grown as described previously (Simpson and Braly, 1970). Cells at mid-log phase were harvested and used immediately for isolation of mitochondria. Stationary phase cells were used for isolation of kinetoplast DNA. The kinetoplast mitochondrial fraction was isolated by Renografin density gradient centrifugation as described previously (Simpson and Simpson, 1978).

Nucleic Acid Isolation

Kinetoplast DNA and EcoRI-linearized maxicircle DNA was isolated...
from total cells as described previously (Simpson, 1979). Kinetoplast RNA was isolated from purified mitochondrial fractions as described previously (Simpson and Simpson, 1979).

Oligonucleotide Synthesis and Labeling

Oligonucleotides were synthesized on an ABS 381A synthesizer using standard phosphoramidite chemistry and were purified either by the OPEC column method (Applied Biosystems) or by TLC chromatography of the detritylated oligonucleotide. Oligonucleotides were 5' end-labeled using T4 polynucleotide kinase (Bethesda Research Laboratories) and [γ-32P]ATP (>7000 Ci/mmol) as described by Maniatis et al. (1982) or 3' end-labeled with [32P]cordonycin (5000 Cpm/mol; New England Nuclear) using terminal denucleotidyl transferase as recommended by the supplier (Bethesda Research Laboratories).

The following oligonucleotides were used in this study (nucleotide localizations of the sequences in the maxicircle sequence, GenBank entry LEIKPMAX, are given in parentheses).

### Probes for gRNAs

- S-107 (CYb-I) 5'-CTTTAATCTCAATTCA (16761-16798)
- S-198 (CYb-II) 5'-CTGATTTAACTTTTCTTAC (2250-2270)
- S-230 (CYb-II) 5'-ATAATTTAATTTTAAATATATTTTCTCA (16754-16763)
- S-42 (COII unedited) 5'-CCAGGTTCTCTACTTTAACTCCT (10009-10019)
- S-43 (COII edited) 5'-CCAGGTATACGTTCTACTTTAAC (10009-10019)

### Probes for mRNAs

- S-72 (CYb edited) 5'-ATAACAAAAATCTAACACGAAAAAACATATT (10161-10181)
- S-99 (COII) 5'-ATTAGATTATATTACA (10137-10142)
- S-200 (MURF3) 5'-CTGCTATTAAATATTATTAT C (367-385)
- S-207 (MURF3) 5'-GGTAAATAAACTATACATCTCATATTTTAAATATTATTTTTC (347-367)
- S-218 (MURF; edited) 5'-ATAAAATAATGATAGACTGG (16754-16733)
- S-219 (MURF; edited) 5'-ATAAAATAATGATAGACTGG (16732-16741)
- S-201 (MURF-II) 5'-CTATTATTTATACACCTGCT (13118-13109)

### References


Note Added in Proof

Recent direct sequence analysis of CYb gRNA-II supports the idea of nonencoded oligo(U) extensions at the 3′ termini of gRNAs, as discussed in the text (B. B. and L. S., unpublished data).