The coding sequence of several mitochondrial mRNAs of the kinetoplastid protozoa is created through the insertion and deletion of specific uridylates. The editing reactions are required to be highly specific in order to ensure that functional open reading frames are created in edited mRNAs and that potentially deleterious modification of normally nonedited sequence does not occur. Selection-amplification and mutagenesis were previously used to identify the optimal sequence requirements for in vitro editing. There is, however, a minority of natural editing sites with suboptimal sequence. Several cis-acting elements, obtained from an in vitro selection, are described here that are able to compensate for a suboptimal editing site. An A + U sequence element within the 5'-untranslated region of cytochrome b mRNA from Leishmania tarentolae is also demonstrated to function as a cis-acting guide RNA and is postulated to compensate for a suboptimal editing site in vivo. Two proteins within an enriched editing extract are UV-cross-linked to two different in vitro selected editing substrates more efficiently than poorly edited RNAs. The results suggest that these proteins contribute to the specificity of the editing reaction.

The kinetoplastid uridylate insertion and deletion editing reactions have been proposed to be initiated by an endonuclease cleavage at an editing site that produces a 5' mRNA fragment with a 3'-hydroxyl and a 3'-fragment with a 5'-monophosphate. U additions or deletions at the 3' (gRNA)1 that binds to its cognate mRNA immediately downstream of an editing site is contained within a guide RNA modification of RNA. Part of the genetic information specifying this mechanism has been obtained in vitro and would complete a cycle of editing that is then reiterated at the 5' end of the mRNA from Leishmania tarentolae. The A+U element during the editing reaction. Several RNAs are UV-cross-linked to two different in vitro selected editing substrates more efficiently than poorly edited RNAs. The results suggest that these proteins contribute to the specificity of the editing reaction.

Received for publication, July 23, 2003, and in revised form, September 30, 2003
Published, JBC Papers in Press, October 6, 2003, DOI 10.1074/jbc.M307997200

Lisa M. Oppegard‡‡‡, Matthew Hillestad‡‡, Ryan T. McCarthy‡‡, Raj D. Patel‡‡, and Gregory J. Connell
From the Department of Pharmacology, Medical School, University of Minnesota, Minneapolis, Minnesota 55455

This paper is available on line at http://www.jbc.org

‡‡‡ Supported by an American Heart Association postdoctoral fellowship.
‡‡‡‡ Present address: Dept. of Molecular and Cellular Biology, University of California, Berkeley, CA 94720.
To whom correspondence should be addressed: Dept. of Pharmacology, Medical School, University of Minnesota, 6-120 Jackson Hall, 321 Church St. SE. Minneapolis, MN 55455. Tel.: 612-624-3132; Fax: 612-625-5408. E-mail: geconnel@lenti.med.umn.edu.

The abbreviations used are: gRNA, guide RNA; RT, reverse transcriptase; DMS, dimethyl sulfate; CMCT, 1-cyclohexyl-3-(2-morpholinioethyl) carbodiimide metho-p-toluene sulfonate.

© 2003 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in U.S.A.
Although many of these features are consistent with a phylogenetic analysis of the sequence flanking an editing site (23), there is a minority of natural editing sites that clearly have a suboptimal sequence.

Several different cis-acting elements obtained from the in vitro selection are described here that are able to compensate for a suboptimal editing site. Two proteins of a highly enriched editing fraction from a high apparent molecular weight that was subsequently UV-cross-linked to two different in vitro selected RNA editing substrates. Cytidines immediately 5′ of the editing site inhibit both editing and the cross-links to these proteins, but both can be compensated by an in vitro selected hairpin loop. The result identifies two proteins that contribute to the specificity of the editing reaction. The in vitro selection also suggested that the cytochrome b mRNA A + U element is functioning as a cis-acting gRNA and this was subsequently confirmed.

EXPERIMENTAL PROCEDURES

Editing Extract—A 2-liter L. tarentolae culture (UC strain) was grown to a density of 1.5 x 10^8 cells/ml at 27 °C in BHI medium (Difco) supplemented with 0.01 mg/ml hemin. The culture was lysed under isonicotinic conditions using a Stansted Disrupter (Energy Service Co.) at 1200 p.s.i. as previously described (17). The mitochondrial fraction was enriched by flotation through a Renografin-60 (diatrizoate meglumine; Bracco Diagnostics) density gradient (24). After one wash in SHE (0.25 M sorbitol, 20 mM Heps, pH 7.5, 2 mM EDTA), the mitochondrial fraction was resuspended in 3.8 ml of lysis buffer, 50 mM Tris (pH 8.7 at 22 °C), 20 mM MgCl₂, 1 mM EDTA, 0.5 mM dithiorthreitol, 10% glycerol, 1 mg/ml Pefabloc, 10 μg/ml leupeptin, and 200 μl of 10% Triton X-100 was added. The fraction was left on ice for 7 min with occasional mixing and then centrifuged at 11,000 x g for 4 min. The supernatant was loaded onto a 4-ml Q-Sepharose column pre-equilibrated in wash buffer (25 mM Heps, pH 7.5, 10 mM MgCl₂, 1 mM EDTA, 20 mM KCl, 0.5 mM dithiorthreitol, 10% glycerol, 1 mg/ml Pefabloc, 10 μg/ml leupeptin), and 200 μl of 10% Triton X-100 was added. The fraction was heated at 65 °C for 20 min and then centrifuged, and the RNA was ethanol-precipitated. The RNA was resuspended in 3 μl of distilled H₂O and 1 μl of urea dye, and the products were resolved on a 60-cm-long/0.4-mm-thick 9% polyacrylamide, 8 m urea gel. The edited signal was quantified after Phosphorimaging scanning (Amersham Biosciences).

Selection-Amplification—The strategy used to select optimal editing sites has been previously described (9). A 100,000 radiolabeled RNA containing a 23-nucleotide random region (Fig. 1A) was treated with editing extract, the products were resolved on a 5% polyacrylamide 8 M urea gel, and the gel was exposed to x-ray film for 12 h at −80 °C. The region of the gel corresponding to 1–3 U insertions was excised and eluted. After ethanol precipitation and resuspension in distilled H₂O, cDNA was synthesized using avian myeloblastosis virus reverse transcriptase (RT) at 42 °C and primer RP-1 (ACACAAGAACAATCTAAACC), which is complementary to nucleotides 51–70 (Fig. 1A). The cDNA product was PCR-amplified using primers RP-1 and RP-1 (TAATACGACTCAC-TATAGGTCTTTATTTGGAATTAATTTA), which contains a T7 promoter sequence at its 5′ end and has sequence corresponding to the amino acids 1–27 of the recombinant (Fig. 1A). Low stringency annealing and extension conditions were used so that cDNA resulting from the reverse transcription of RNAs containing insertions within the upstream fixed sequence could be amplified (9).

Localization of U Insertions—RNAs were treated with extract, and the slower migrating bands with insertions were excised, eluted, and precipitated for reverse transcription. Two parallel reverse transcription reactions were performed using primers RP-2 (TCTGAATTCCTAACTAAAAAGCACCC), which is complementary to nucleotides 1–18 (Fig. 1A), and RP-3 (TCTGAATTCCTAACTAAAAAGCACCC), which is complementary to nucleotides 51–70; both primers also have EcoRI sites on their 5′ ends to facilitate subsequent cloning. Full-length cDNA was gel-purified, eluted in the presence of 10 μg of carrier tRNA, and ethanol-precipitated. Terminal transferase (Roche Applied Science) was used to add an oligo(dA) tail to the isolated cDNA. The tiled cDNA was PCR-amplified for cloning using FP-2 (TCTGAAATTTTGTATTTTGTTTGTATTTTTTTTTTTT) and either RP-2 or RP-3 as appropriate.

Chemical Modifications—The dimethyl sulfate (DMS) and 1-cyclohexyl-1′-3′-dithiopropan-2-ol carbodiimide metho-p-toluene sulfonate (CMCT) modification reactions were performed as previously described (26) with the exception that the CMCT concentration was reduced to 2.6 μg/ml, and the CMCT denaturing reaction was performed at 70 °C for 1 min. The modifications were detected by reverse transcription using several different primers: RP-4 (CATAAAATAGACAAACACC), complementary to nucleotides 14–16 (Fig. 2A), and either RP-5 (CAGCTCCAATTTATCCC) complementary to nucleotides 33 to 16 of RNA B-1 or RP-6 (CCCTTTAAAAATTATCCC) complementary to nucleotides 33 to 16 of RNA C-1 (Fig. 2A).

UV Cross-linking—One pmol of radiolabeled linear RNA was heated in 5 μl of 25 mM Tris (pH 8.1 at 27 °C), 0.2 mM EDTA at 65 °C for 3 min, 11 μl of binding buffer (50 mM KCl, 25 mM MgCl₂, 7.5 mM potassium phosphate, 50 mM dithiorthreitol, 2.5 mM ATP, 1 mg/ml Pefabloc, and 10 μg/ml leupeptin) was added, and the RNA was left at room temperature for 10 min. After the addition of 9 μl of the SP-fractionated editing extract (0.2 μg of protein) the RNAs were incubated at 27 °C for 20 min. The RNA-editing extract solution was transferred onto paraffin oil and exposed to 120,000–150,000lux for 1 h. Prior to UV-cross-linking, a 254-nm light source in a model 1800 Stratallinker (Stratagene). The products were treated with 10 μg of RNase A for 10 min and then resolved on a 7% SDS-polycrylamide gel. The cross-linked products were quantified after Phosphorimager scanning.

RESULTS

Diverse Editing Substrates—Selection-amplification (21, 22) was initially exploited in order to identify the sequence and structural requirements for the A + U-dependent editing reaction. A random mixture of the four standard nucleotides was substituted at 23 positions of the A + U element (Fig. 1A). RNAs within the random population that had the sequence and structure able to support in vitro U additions independent of added gRNA were partitioned from the bulk unedited popula-
In vitro selection of editing substrates. A, sequence of the 32-nucleotide A + U element (nucleotides 19–50) from the 5'-untranslated sequence of the cytochrome b mRNA (nucleotides 5371–5402 in LEI RPMAX) (18) and flanking artificial fixed sequence. A 23-nucleotide region of the element was replaced with a random mixture of the four standard nucleotides. Those RNAs within the random population that are substrates for the in vitro U insertion reaction were partitioned from the bulk unedited population on the basis of their slower mobility on a polyacrylamide gel. The slower migrating RNA was excised from the gel and amplified by RT-PCR and T7 RNA polymerase-mediated transcription for a subsequent cycle of enrichment. The fixed sequences flanking the random region were exploited as the RT-PCR primer binding sites, and the primer annealing and extension conditions were chosen such that editing sites within the fixed sequence would be regenerated. B, the sequence of the RNAs obtained after four cycles of selection. Group A RNAs function as cis-acting gRNAs that direct U insertions between nucleotides 16 and 21 (9). Groups B–F are defined by different 8–14-nucleotide sequences that were each selected twice (boxed). Group G RNAs are unrelated in primary sequence both to the seven other groups and to each other. C, activity of selected RNAs B-1 and C-1 relative to the parental A + U RNA and the random population used at the start of the in vitro selection.

In addition to selecting RNAs that are permutations of the A + U element, it was possible that cis-acting gRNAs would be obtained from the selection. In fact, 14 of the RNAs selected after four cycles of selection clearly function as cis-acting gRNAs that guide U insertions between nucleotides 16 and 21 of the fixed sequence (labeled Group A in Fig. 1B). The first six RNAs within group A were used to define the critical RNA sequence and structure required for efficient gRNA-directed editing and have been described (9). However, the majority of the selected RNAs did not appear to be related to the group A RNAs, and there also did not appear to be any obvious predicted structural similarity among these RNAs. Several selected sequences 8–14 nucleotides in length are each present within two different RNAs and were used to define the groups labeled B–F (Fig. 1B). Even within each group, however, it was difficult to fold the two RNAs into a structure that placed the consensus sequence within a similar conformation. The majority of the selected RNAs did not appear to be related either to the seven groups or to each other and were defined as group G.

Since it was possible that RNAs were selected without meeting the intended criterion of being an editing substrate, two of the RNAs labeled B-1 and C-1 were chosen for further characterization. Treatment of both RNAs with editing extract resulted in a ladder of slower migrating products not detected with the starting random population (Fig. 1C). The slower migrating products are consistent with RNAs B-1 and C-1 being substrates for the in vitro U insertion reaction. However, unlike the parental A + U sequence, the predominant product formed with the editing of RNAs B-1 and C-1 is consistent with the insertion of only one U.

Secondary Structure of RNAs B-1 and C-1—The location of the in vitro U insertions was determined for both RNAs B-1 and C-1. After treating the RNAs with extract, the edited product was resolved on a denaturing polyacrylamide gel, excised and reverse-transcribed in two parallel reactions using two non-overlapping primers. The two different reverse transcription primers permitted U insertions to be detected throughout the entire sequence within each population of edited RNA. After the reverse transcriptions, a 3' oligo(dA) extension was added to the cDNAs using terminal transferase and the product was PCR-amplified for cloning and sequencing. The major editing site for both RNAs B-1 and C-1 is located between nucleotides 59 and 62 (Fig. 2A). There is ambiguity in the precise localization because nucleotides 60 and 61 are Us, and it has been demonstrated for other RNAs that editing can occur both 5' and 3' of a genomically encoded U (9). In addition to the major editing site(s), both RNAs also had several minor editing sites. The multiple editing sites were of interest, because it is also a characteristic of the A + U element-dependent reaction (18).

In order to obtain insight into the relevant secondary structure of RNAs B-1 and C-1, conformations that placed a potential guiding nucleotide opposite the major site of the A + U element-dependent reaction (18). These were subsequently tested by altering the number of purines at positions that potentially could act as guiding nucleotides in order to determine whether the number of Us inserted during the editing reaction was altered in the predicted manner. Changing the number of purines from zero to two at position 21 had the predicted effect on the number of insertions within both RNAs (Fig. 2B). Sequencing confirmed that the change in the number of insertions was occurring between nucleotides 59 and 62 (not shown), the major editing site. The relatively low background of incorrect insertions, especially when there are no guiding nucleotides at position 21...
is also consistent with the other editing sites making only a minor contribution to the editing signal. Knowing the location of the guiding nucleotides and the major site of insertion provided constraints on the possible secondary structures. Most of the non-group A selected RNAs can be folded into hairpin loop structures similar to RNAs B-1 and C-1.
(not shown); these include the group B–F RNAs and RNAs G-1 through G-30 (Fig. 1B).

In order to refine the predicted secondary structures, both RNAs B-1 and C-1 were probed with the RNA-modifying agents DMS and CMCT (Fig. 2, C and D). Whereas DMS modifies As and Cs at Watson-Crick positions that are not involved in secondary or tertiary interactions, CMCT modifies Gs and Us at unprotected Watson-Crick positions. Sites of modification were detected by RT-primer extension. The intensity from each RT product of the treated RNA was corrected for the corresponding modification-independent termination, and the ratio of the corrected intensity obtained under native conditions to that obtained under denaturing conditions was calculated for each position (Fig. 2A). Values could not be calculated at those positions where there are strong modification-independent RT terminations or limited reactivity under denaturing conditions.

Most of the nucleotides within the predicted hairpin loop of RNA B-1 (nucleotides 33–50) are reactive under native conditions, and the nucleotides within the predicted helix formed by the pairing of (G-51, C-52, G-53, U-54, U-55, U-56) with (U-32, G-31, C-30, G-29, G-28, A-27) are mostly protected under the native reaction conditions (Fig. 2, A and C). Although the helix could potentially be extended further through the pairing of (A-57, A-58) with (U-26, U-25) and the pairing of (U-48, A-49, U-50) with (A-35, U-34, G-33), the modifications are more consistent with these nucleotides being unpaired. A-21 is more reactive than both A-22 and A-23, which is consistent with it being the unpaired guiding nucleotide. The modification reactions are also consistent with the helix predicted to form through the pairing of (G-62, U-63, U-64 C-65, U-66) with (U-20, A-19, G-18, G-17, G-16, G-15); however, modification-independent RT terminations limit the value of chemical probing within this region.

The chemical probing is also mostly consistent with the proposed folding of RNA C-1 (Fig. 2, A and D). The ambiguity in the RNA C-1 secondary structure is primarily within the region predicted to form the internal G-G bulge (U-49, C-50, G-51, G-52, G-53, A-34). U-49 is not protected, and A-34, C-50, and C-52 are only partially protected, which is not completely consistent with the proposed structure but could result from fraying of base pairs near the end of the helix. Approximately 20% of the RNAs obtained from the fourth cycle of selection can also be folded with analogous G-G bulge loops, including RNAs B-2, D-1, E-1, E-2, and G-1 through G-11. The discrepancy between the in vitro selection of RNAs with the G-G bulge and chemical modification results may reflect an increased stability of the G-G bulged loop region in the presence of editing extract relative to the chemical probing when no protein is present.

Mutagenesis is also consistent with the proposed structures. Substitutions within RNA B-1 that would disrupt the helix potentially formed by the pairing of (G-51, C-52, G-53, U-54, U-55, U-56) with (U-32, G-31, C-30, G-29, G-28, A-27) also inhibited editing (Fig. 3A, M1 mutation; Fig. 3C, lanes 1–4). Editing activity was partially recovered by compensatory mutations that restored the helix (Fig. 3A, M2 mutation; Fig. 3C, lanes 5 and 6). The partial recovery (70%) suggests that there could be a sequence preference within this region or that the extra stability associated with the G-G pair at positions 58 and 25 is inhibitory. Similar results were obtained with the corresponding mutations to RNA C-1 (not shown).

Hairpin Loops That Stimulate gRNA-directed Editing—The consensus sequences that initially defined the group B and C RNAs (boldface lowercase italic type in Fig. 3, A and B) are predicted to be within hairpin loops, well removed in secondary structure from the major editing site. Mutation of the consensus sequences (Fig. 3A, M3 of RNA B-1; Fig. 3B, M6 of RNA C-1) significantly inhibited editing (Fig. 3C, lanes 7 and 8; Fig. 3D, lanes 3 and 4). Other mutations within the predicted loop regions (Fig. 3, A and D) are also inhibitory (M4 of Fig. 3C, lanes 9 and 10; M7 and M8 of Fig. 3D, lanes 5–8). Although these mutations would be expected to disrupt the structure flanking the minor editing sites, the magnitude of the inhibition cannot be accounted for by the loss of editing at the minor sites alone. Mutation of the nonselected hairpin loop of RNA B-1 (M5 of Fig. 3A) had no significant effect on editing (Fig. 3C, lanes 11 and 12). Mutation of G-32 within the predicted internal G-G bulge loop of RNA C-1 (Fig. 3B) inhibited 50% of the editing (Fig. 3D, lanes 9 and 10), emphasizing the importance of this structure. In contrast, disruption of the nonselected C-59–U-24 internal bulge loop of RNA C-1 by a U24G mutation had no significant effect on the reaction (Fig. 3D, lanes 11 and 12). The selected hairpin and bulge loops are the first identified cis-acting elements that stimulate gRNA-directed editing.

The Hairpin Loops Compensate for a Suboptimal Editing Site—Both RNAs B-1 and C-1 have a cytidine upstream of the major editing site, at position 59 (Fig. 3, A and B), which was surprising because cytidines upstream of editing sites are highly disfavored in nature (23) and have previously been shown to inhibit editing of all other tested RNA substrates (6, 8, 9). In order to determine whether the hairpin loops of RNAs B-1 and C-1 were selected to compensate for the upstream cytidine, C-59 was mutated to an A in the context of the M3 and M6 mutations; these are loop mutations that inhibit editing (Fig. 3, C and D; Fig. 4, lanes 3 and 9). The C-59 to A mutation is able to restore editing of an RNA with the M3 mutation to ~50% that of the parental RNA (Fig. 4, lanes 1–6) and completely restores editing of an RNA with the M6 mutation (Fig. 4, lanes 7–12). These results indicate that the hairpin loops were selected, at least in part, to compensate for the suboptimal editing site.

Recognition of an Editing Site—UV cross-linking of an enriched editing extract to the radiolabeled substrate RNAs was performed in order to determine whether the selected loop sequences affect the interaction of RNA binding proteins. RNA C-1 can be cross-linked to three protease-sensitive bands with apparent molecular weights of 80,000, 100,000, and 120,000 (Fig. 5A, lane 1). All three radiolabeled complexes can be competed by treatment with a 2.5–20-fold molar excess of unlabelled RNA C-1 prior to the cross-linking (Fig. 5A, lanes 2–5). However, a molar excess of RNA M6, a mutation of the selected hairpin loop of RNA C-1 that inhibits editing (Fig. 3, B and D), does not compete as efficiently for the complexes. The difference is most apparent for the 100-kDa band (Fig. 5, A and B).

The 80- and 100-kDa bands also become UV-cross-linked to RNA A-1 (Fig. 5D), a different in vitro selected editing substrate (Fig. 1B) (9). Cytidines are highly disfavored upstream of an editing site in RNA A-1 (9), and an A-1 mutant containing upstream cytidines (RNA M16 of Fig. 5C) also does not compete as efficiently as the parental sequence for the 80- and 100-kDa proteins (Fig. 5, D and E). The failure of the 120-kDa protein to be cross-linked to RNA A-1 suggests that it is not essential for general editing site recognition, but the additional interaction could potentially be contributing to the compensation of the suboptimal editing site of RNA C-1. The decreased ability of RNAs M6 and M16, both deficient editing substrates, to compete with the respective parental RNAs for the 80- and 100-kDa cross-links is consistent with these proteins being critical for editing site recognition. The cross-linking results also suggest that the increased binding of RNA C-1 to the 80-, 100-, and 120-kDa proteins is able to compensate for the decreased affinity caused by the upstream cytidine. In contrast, mutation of the selected loop sequence of RNA B-1 does not significantly
Affect cross-linking of the 80- and 100-kDa bands (not shown), which suggests that there are probably several different mechanisms through which the different selected sequences could be compensating for a suboptimal editing site. This is also consistent with the large degree of sequence variation obtained from the in vitro selection (Fig. 1B); multiple solutions to biological problems are frequently obtained from in vitro selections (reviewed in Ref. 28).

The A + U Element Is a Cis-acting gRNA—The in vitro selection indicated that relatively weak secondary structures with cis-acting gRNAs can form, and these can be identified when both the location of the insertions and the guiding nucleotides are known (Fig. 2). Whereas the parental A + H11001 U element in the context of the natural cytochrome b mRNA sequence induces U insertions at multiple sites within both the 5' and 3'-flanking sequence (18), U insertions are restricted to a limited number of 5' sites when the purine-rich sequence 3' of the A + H11001 U element is mutated (Fig. 6A) (6). The A + U sequence, in this context, can be folded such that two guiding nucleotides are placed opposite the major in vitro editing site (Fig. 6A). This structure differs significantly from that suggested by the chemical probing, in the absence of editing extract, of the A + H11001 U element within the context of the natural cytochrome b mRNA sequence (17). However, changing the number of guiding As at position 43 resulted in the predicted change to the number of insertions, indicating that the structure may account for a portion of the insertions (Fig. 6B).
Additional mutagenesis is also consistent with the proposed structure. Insertion of an A immediately 5' of U-46 is predicted to disrupt the helix upstream of the editing site (Fig. 6A, M9) and inhibits the editing reaction (Fig. 6C, lanes 3 and 4). The mutation is compensated by the insertion of a U immediately 5' of A-21 (M10, lanes 5 and 6), which is predicted to restore the helix (Fig. 6A). Disruption of the major editing site by the insertion of an A immediately 5' of A-23 (Fig. 6A, M11) inhibits the reaction (Fig. 6C, lanes 7 and 8). Likewise, disruption of the helix downstream of the editing site is inhibitory but can be compensated by a mutation that is predicted to restore the helix (Fig. 6A, M12 and M13; Fig. 6C, lanes 9–12). The hairpin loop does not appear to contribute to the reaction, since neither excess of the indicated unlabeled RNA, after RNase A treatment, the cross-linked products were resolved on a 7% SDS-polyacrylamide gel. B, the ratio of the intensity of each of the cross-linked bands formed in the presence of several different concentrations of unlabeled RNA M6 competitor to that formed in the presence of the same concentration of unlabeled RNA C-1 competitor was calculated from three independent sets of cross-linking reactions. If RNA M6 competed as efficiently as RNA C-1 for the complexes, the value of this ratio at each point would have been 1. C, secondary structure of RNA A-1 and description of the mutation (M16) that places cytidines immediately upstream of the editing site. The selected portion of the sequence is in lowercase type. D, radiolabeled RNA A-1 was UV-cross-linked to a fractionated editing extract as described for A. E, quantitation of the RNA A-1 cross-linking results from three independent sets of reactions as described for B.
insertions (M14) nor substitutions (M15) within the loop significantly affect editing (Fig. 6C, lanes 13–16). It is also possible that other features of the A + U sequence element in the context of the natural cytochrome b mRNA, which differs in having a cytidine upstream of the first editing site, could be influencing the reaction.

**DISCUSSION**

The optimal sequence and structure of the RNA required for *in vitro* gRNA-directed insertional editing have previously been defined through selection-amplification and mutagenesis (6, 8, 9). A cytidine located upstream of an insertional editing site was demonstrated to inhibit editing, consistent with a phylogenetic analysis of natural editing sites (23). We have shown here that this inhibition could result, at least in part, from a decreased affinity for 80- and 100-kDa proteins present within a highly enriched editing extract (Fig. 5, D and E). The results strongly suggest that these two proteins are critical for editing site recognition. A small percentage of kinetoplastid insertional editing sites in nature have an upstream cytidine. The selection-amplification results described here indicate that cis-acting elements can compensate for a suboptimal editing site *in vitro* (Fig. 4), and it is possible that analogous elements could also be functioning *in vivo*.

The large degree of sequence variation between the group B and C RNAs (Fig. 1B) suggests that there could be more than one mechanism to compensate for an upstream cytidine. The selected hairpin loop of RNA C-1 facilitates binding to a 100-kDa protein and to a lesser extent to 80- and 120-kDa proteins (Fig. 5, A and B). This could either be through a direct interaction or indirectly through the optimization of the structure near the editing site. Since an upstream cytidine inhibits the interaction with the 80- and 100-kDa proteins (Fig. 5, D and E), the results suggest that the selected RNA C-1 sequence functions, at least in part, by compensating for the decreased binding. In contrast to RNA C-1, mutagenesis of the selected hairpin loop of RNA B-1 did not affect cross-linking (not shown), which suggests that this selected sequence compensates for the suboptimal editing site through a different mechanism. It is possible that RNA B-1 is facilitating the reaction through interaction with proteins of the editing complex other than the 80- and 100-kDa proteins. Since not all RNA-binding proteins would have functional groups suitably orientated for direct UV cross-linking, direct UV cross-linking would be expected to only identify a subset of the proteins involved with editing site recognition.

Although others have exploited UV cross-linking to identify putative editing-related proteins (29–31), our results differ in several important aspects. First, earlier cross-linking studies exploited relatively crude mitochondrial extracts or lysates. As a result, several unrelated RNA-binding proteins as well as metabolic enzymes with nucleotide binding sites such as glutamate dehydrogenase (31) were detected. In contrast, we used a highly fractionated mitochondrial extract (20), and the background of cross-linking to unrelated RNA binding proteins is significantly reduced. Second, earlier cross-linking exploited a radiolabeled gRNA alone, not a mRNA substrate with a cis-acting gRNA. Since an editing site is formed by the annealing of a gRNA and an mRNA, it is less likely that the protein interactions directly relevant to editing site recognition would have been detected. Third, our cross-linking exploited RNAs that are significantly better *in vitro* editing substrates than the natural RNAs, and binding to the editing complex for some of these RNAs may also be stronger as a result of the selection process. Fourth, the cross-links did not form as well with RNAs containing some of the same mutations that inhibit editing, which supports the likelihood that the cross-linked proteins are related to editing, and this degree of specificity was not previously demonstrated.

The selection-amplification was originally designed to obtain permutations of the A + U element that direct editing independent of added gRNA in order to provide insight into the element’s function. We did not select any RNAs that had significant primary sequence similarity to the cytochrome b mRNA A – U element, but the selection indicated that relatively weak secondary structures could form with cis-acting gRNAs. Although these structures were not always evident, localizing the sites of insertion and identifying the guiding nucleotide(s) provided a powerful structural constraint. Chemical probing or mutagenesis then refined the structure. Using this strategy, the parental A + U element was demonstrated to function as a cis-acting gRNA. The A + U element is ideally suited to act as a guiding sequence, since the blocks of U nucleotides can pair with purines, which are abundant within the cytochrome b editing sites immediately 3’ of the A + U element, and the mismatched A nucleotides potentially could act as guiding nucleotides. Flexibility in the secondary structure could possibly account for editing at multiple sites, although it is also possible that there are additional subtle features of the A + U element that also make it a target for the editing machinery. The relatively low stability of the structure may also account, in part, for the A + U sequence not being able to function in *trans* like other cis-acting guiding sequences.

The biological relevance of the A + U element is still unclear. The 5’-untranslated region of cytochrome b mRNA containing the A + U element was previously cloned from purified mitochondrial RNA and sequenced (19). Of the 49 sequenced clones, 26 contained U deletions immediately flanking the A + U element, sites that overlap the *in vitro* A + U-dependent U insertions. This was a remarkable finding, because the majority of the gRNAs of the *L. tarentolae* UC strain used for these studies have been identified (32, 33), and none can account for the deletions. Although the possibility cannot be unambiguously eliminated that the deletions result from as yet unidentified low abundant gRNAs or possibly by mismatched gRNAs, the A + U element can be folded into low energy conformations in which it could act as a guide for the deletions. The cause of *in vivo* U insertions predominating in *in vivo* whereas deletions predominate in *in vitro* could possibly involve subtle differences in the folding of the RNA in the two environments. The *in vivo* deletions flanking the cytochrome b mRNA A + U sequence are within the 5’-untranslated sequence and as a result could potentially impact either mRNA processing or translation. Alternatively, the deletions themselves may not be biologically significant but rather reflect the presence of the editing machinery interacting with the A + U element. Cytochrome b mRNA is one of the few *L. tarentolae* mRNAs that has a cytidine immediately upstream of an editing site, and it is possible that the A + U element could in some manner be compensating *in vivo* for the suboptimal sequence. This would be analogous to the hairpin loops of RNAs B and C compensating for the upstream cytidine *in vitro* (Fig. 4), but this remains to be tested.

**REFERENCES**

1. Blum, B., Bakalara, N., and Simpson, L. (1990) *Cell* 60, 189–198
2. Steiert, S. D., Heidmann, S., and Stuart, K. (1996) *Cell* 84, 831–841
3. Kabbe, M. L., Steiert, S. D., Heidmann, S., and Stuart, K. (1996) *Science* 273, 1189–1195
4. Cruz-Reyes, J., and Sollner-Webb, B. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 8901–8906
5. Byrnes, E. M., Connell, G. J., and Simpson, L. (1996) *EMBO J.* 15, 6758–6765
6. Kabbe, A. L., Oppedeur, L. M., McKenzie, B. A., and Connell, G. J. (2001) *Nucleic Acids Res.* 29, 2575–2580
7. Cruz-Reyes, J., Zhelkina, A., Rusche, L., and Sollner-Webb, B. (2001) *Mol. Cell. Biol.* 21, 884–892
Recognition of a Kinetoplastid Editing Site

51175

8. Igo, R. P., Jr., Lawson, S. D., and Stuart, K. (2002) *Mol. Cell. Biol.* 22, 1567–1576
9. Pai, R. D., Oppegard, L. M., and Connell, G. J. (2003) *RNA* 9, 469–483
10. Schnaufer, A., Panigrahi, A. K., Panicucci, B., Igo, R. P., Jr., Wirtz, E., Salavati, R., and Stuart, K. (2001) *Science* 294, 2159–2162
11. Rusche, L. N., Huang, C. E., Piller, K. J., Hemann, M., Wirtz, E., and Sollner-Webb, B. (2001) *Mol. Cell. Biol.* 21, 979–989
12. McManus, M. T., Shimamura, M., Grams, J., and Hajduk, S. L. (2001) *RNA* 7, 167–175
13. Huang, C. E., O’Hearn, S. F., and Sollner-Webb, B. (2002) *Mol. Cell. Biol.* 22, 3194–3203
14. Panigrahi, A. K., Schnaufer, A., Carmean, N., Igo, R. P., Gygi, S. P., Ernst, N. L., Palazzo, S. S., Weston, D. S., Abersold, R., Salavati, R., and Stuart, K. D. (2001) *Mol. Cell. Biol.* 21, 6833–6840
15. Drezdz, M., Palazzo, S. S., Salavati, R., O’Hear, J., Clayton, C., and Stuart, K. (2002) *EMBO J.* 21, 1791–1799
16. Ernst, N. L., Panicucci, B., Igo, R. P., Jr., Panigrahi, A. K., Salavati, R., and Stuart, K. (2003) *Mol. Cell* 11, 1528–1536
17. Connell, G., J. Byrne, E., and Simpson, L. (1997) *J. Biol. Chem.* 272, 4212–4218
18. Brown, L. M., Burbach, B. J., McKenzie, B. A., and Connell, G. J. (1999) *J. Biol. Chem.* 274, 6295–6304
19. Oppegard, L. M., Kabb, A. L., and Connell, G. J. (2000) *J. Biol. Chem.* 275, 33911–33919
20. Oppegard, L. M., and Connell, G. J. (2002) *Int. J. Parasitol.* 32, 859–866
21. Tuerk, C., and Gold, L. (1990) *Science* 249, 505–510
22. Ellington, A. D., and Szostak, J. W. (1990) *Nature* 346, 818–822
23. Burgess, M. L. K., and Stuart, K. (2000) *RNA* 6, 1492–1497
24. Braly, P., Simpson, L., and Kretzer, F. (1974) *J. Protocoll.* 21, 782–790
25. Milligan, J. P., Greebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* 15, 8783–8798
26. Krol, A., and Carbon, P. (1989) *Methods Enzymol.* 180, 212–227
27. Zuber, M. (2005) *Nucleic Acids Res.* 31, 3406–3415
28. Gold, L., Polisky, B., Uhlenbeck O., and Yarus, M. (1995) *Annu. Rev. Biochem.* 64, 763–797
29. Koller, J., Norskau, G., Paul, A. S., Stuart, K., and Goringer, H. U. (1994) *Nucleic Acids Res.* 22, 1988–1995
30. Leegwater, P., Speijer, D., and Benne, R. (1995) *Eur. J. Biochem.* 227, 780–786
31. Bringaud, F., Stripecke, R., Frech, G. C., Freedlund, S., Tuerk, C., Byrne, E. M., and Simpson, L. (1997) *Mol. Cell. Biol.* 17, 3915–3923
32. Maslov, D. A., and Simpson, L. (1999) *Cell* 70, 459–467
33. Thiemann, O. H., Maslov, D. A., and Simpson, L. (1994) *EMBO J.* 13, 5689–5700