778–787 Nucleic Acids Research, 1999, Vol. 27, No. 3

Mapping contacts between gRNA and mRNA in trypanosome RNA editing

Sheldon S. Leung and Donna J. Koslowsky*

Department of Microbiology, Michigan State University, East Lansing, MI 48824, USA

Received September 30, 1998; Revised November 23, 1998; Accepted December 2, 1998

ABSTRACT

All guide RNAs (gRNAs) identified to date have defined 5′ anchor sequences, guiding sequences and a non-encoded 3′ uridylate tail. The 5′ anchor is required for *in vitro* editing and is thought to be responsible for selection and binding to the pre-edited mRNA. Little is known, however, about how the gRNAs are used to direct RNA editing. Utilizing the photo-reactive cross-linking agent, azidophenacyl (APA), attached to the 5′- or 3′-terminus of the gRNA, we have begun to map the structural relationships between the different defined regions of the gRNA with the pre-edited mRNA. Analyses of crosslinked conjugates produced with a 5′-terminal APA group confirm that the anchor of the gRNA is correctly positioning the interacting molecules. 3′ Crosslinks (X-linker placed at the 3′-end of a U_{10} tail) have also been mapped for three different gRNA/mRNA pairs. In all cases, analyses indicate that the U-tail can interact with a range of nucleotides located upstream of the first edited site. It appears that the U-tail prefers purine-rich sites, close to the first few editing sites. These results suggest that the U-tail may act in concert with the anchor to melt out secondary structure in the mRNA in the immediate editing domain, possibly increasing the accessibility of the editing complex to the proper editing sites.

INTRODUCTION

RNA editing in *Trypanosoma brucei* is a post-transcriptional process that inserts and deletes uridylate residues (U) from mitochondrial pre-mRNA molecules (1,2). This phenomenon produces mature mRNAs by creating open reading frames, correcting encoded frameshifts and creating signals for translational initiation and termination. The precise placement of U residues is guided by a small RNA molecule, the guide (g)RNA, which is complementary to portions of the mature mRNA (3). Analyses of gRNAs reveal that they appear to be made up of three functional elements. Contained within the 5′-end of gRNAs is a short sequence known as the gRNA anchor. The anchor can base pair with the mRNA just 3′ of the region to be edited, positioning the gRNA for the editing process (4,5). The second element (the guiding region) provides the information for the specific insertion and deletion of U residues. Finally, at the 3′-end of the gRNA is a non-encoded U-tail (6). The function of the U-tail remains unclear, however, several roles have been suggested by various models of editing (6–8).

In vitro kinetic analyses of products and possible intermediates of RNA editing supports the enzyme cascade model of editing first proposed by Blum *et al*. (3,6). The detection of both 5′ and 3′ cleavage products and the observation that inserted U residues are derived from free UTP has ruled out two previous mechanistic models involving chimeric gRNA/mRNA intermediates (5,9,10). The chimeric intermediate models suggested that the gRNA oligo(U) tail served as the U donor or acceptor during the editing process (7,8,11). The elimination of chimeras as editing intermediates throws the role of the U-tail into question. In the original cleavage–ligation model, it was suggested that the U-tail functions by binding to purine-rich regions upstream of the editing sites, thereby strengthening the interaction of the gRNA and pre-mRNA (6). In *in vitro* editing studies, removal of the gRNA U-tail does not diminish gRNA-directed mRNA cleavage (5). Formation of the edited product, however, was severely diminished, suggesting that it may play a role in holding on to the 5′ mRNA cleavage product during the editing reaction.

To provide insight into the role(s) the gRNA U-tail may play in the editing process, we mapped the interaction of the U-tail of three different gRNAs with their pre-mRNAs. This was accomplished using the photo-reactive crosslinking agent, azidophenacyl (APA), specifically attached to the 3′-end of the gRNA. In addition, we mapped the interaction of the gRNA 5′ anchor with its cognate pre-mRNA by placing the photoagent at the 5′-end of the gRNA. We report here our results, confirming the role of the anchor in correctly positioning the gRNA and providing evidence that suggests that the U-tail does bind purine-rich sequences upstream of editing sites. Interestingly, the U-tail interacted with purine-rich sequences near (5–28 bases) the first editing site, even when the more stable predicted interaction would involve more upstream regions. This crosslinking information was used to generate computer-predicted secondary structure models for the gRNA/pre-mRNA interactions. For all three gRNA/mRNA pairs, the predicted secondary structures are similar. In all cases, the anchor duplex region is correctly paired and secondary structure in the mRNA editing region is likely eliminated. In addition, the gRNA guiding region forms a potential stem–loop positioned across from the first few editing sites. These results suggest that the U-tail may act not only to increase the stability of the RNA

*To whom correspondence should be addressed. Tel: +1 517 432 3362; Fax: +1 517 353 8957; Email: koslowsk@pilot.msu.edu

interactions, but may also work to 'iron out' any secondary structure in the mRNA in the immediate editing domain, possibly increasing the accessibility of the editing complex to the proper editing sites.

MATERIALS AND METHODS

Nucleic acids

Plasmid DNA. 5′CYbUT and 3′A6UT have been previously described (12,13). 5′ND7UMT was prepared by PCR amplification of maxicircle DNA using ND75′NEdc and MODHR3 oligodeoxynucleotides and cloning into pBluescriptII-SK$^-$ (Stratagene). Plasmids for gND7-506 and gA6-14 were gifts from Dr Ulrich Göringer (14). The template for gCYb-558 (15) was created using overlapping oligodeoxynucleotides (T7, gCYb-558-1 and gCYb-1end).

PCR products. mRNA templates for *in vitro* transcription were amplified using the T7 and BIG SK oligodeoxynucleotides. gRNA templates were PCR amplified using the T7 oligodeoxynucleotide and 3′ primers complementary to the 3′-ends of the gRNAs. Amplification of gCYb-558 involved either gCYb-1end or gCYB-558endU5. PCR reactions were performed as per the manufacturer's instructions (Promega).

Oligodeoxynucleotides:

T7	5′-AATTTAATACGACTCACTATAG-3′	22 nt
BIG SK	5′-GGCCGCTCTAGAACTAGTGG-3′	20 nt
ND75′NEdc	5′-CGGGTACCATGACTACATGATAAGTAC-3′	27 nt
TbHR3	5′-CTTTTATATTCACATACTTTTCTGTACC-3′	27 nt
MODHR3	5′-CCGGATCCATGGACGAACTACAAACACGATGCAAAT-3′	36 nt
gND7-506end	5′-AAAAAAAAAATTCACTATATACAC-3′	24 nt
gCYb-558-1	5′-CCTAGAAATTCACATTGTCTTTTAATCCCTATAGTGAGTCG-3′	41 nt
gCYb-1end	5′-AAAAAAAAAATTCCCTTTATCACCTAGAAATTCAC-3′	35 nt
gCYb-558endU5	5′-AAAAATTCCCTTTATCACCTAGAAATTCAC-3′	30 nt
gA6-14end	5′-AAAAAAAAAATAATTATCATATC-3′	23 nt
C-gA6-14	5′-CAGGAATTCCGATAACGAATCAGATTTTGAC-3′	31 nt
A6H-1	5′-CCTAACCTTTCCTGC-3′	15 nt
T7leadercomp	5′-GGTACCCAATTCGCC-3′	15 nt

In vitro transcription

T7 RNA polymerase (200 U) *in vitro* transcription reactions (40 mM Tris–HCl, pH 8.0, 19 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine, 0.01% w/v Triton X-100, 16 U RNAsin, 1 U yeast pyrophosphatase, 4 mM each ribonucleotide) were carried out for 6 h at 37°C. Radioactively labeled transcripts were produced using 50 µCi of [α-^{32}P]ATP, 800 Ci/mmol (NEN). For 5′-end APA modification, transcription was carried out in the presence of 7 mM guanosine 5′-phosphorothioate (GMPS) prepared following Burgin and Pace (16). Transcripts were gel purified on an 8% polyacrylamide (w/v)–7 M urea gel.

Attachment of photoaffinity agents

Following Burgin and Pace (16), azidophenacyl bromide (Sigma) was incubated with gND7-506 and gCYb-558 to label the 5′-end of the transcripts with azidophenacyl. 3′-Photoagent-labeled gRNAs were produced using the protocol of Oh and Pace (17).

Crosslinking of gRNAs and pre-mRNAs

Reactions contained 90 pmol of gRNA in the presence of 45 pmol of pre-mRNA. Mitochondrial extract was fractionated via glycerol gradients (5,18). Each 0.5 ml fraction was then tested for activity using the deletion assay (5,19). Hybridizations were carried out under RNA editing conditions (5). Reactions were heated to 60°C for 2 min and cooled to 27°C at a rate of 2°C/min. If the reaction was to contain protein, 7 µl of active fraction was added at this point. Reactions were then incubated a further 20 min at 27°C. Reactions were transferred to 120 µl GeNunc modules and irradiated using a Stratalinker (Stratagene) with 312 nm bulbs for 20 min while on ice. Reactions were kept 5 cm from the bulbs and shielded by a polystyrene Petri dish during irradiation (blocks wavelengths <300 nm). Crosslinked RNAs were resolved on 6% polyacrylamide–7 M urea gels, cut out and eluted overnight at room temperature (0.3 M NaOAc, 0.2% w/v SDS). RNAs were ethanol precipitated and resuspended in 15 µl of H_2O.

Primer extension analysis

An aliquot of 5 µl of crosslinked RNA or 2–5 ng of control RNA was mixed with 5′-^{32}P-labeled BIG SK (50 000 c.p.m.) and heated to 90°C for 2 min in 50 mM KCl, 20 mM Tris–HCl, pH 8.5, 0.5 mM Na_2EDTA and 8 mM $MgCl_2$. Reactions were cooled at 2°C/min to 45–50°C and primer extension (33 mM KCl, 13 mM Tris, pH 8.5, 0.33 mM EDTA, 5 mM $MgCl_2$, 11 mM DTT) carried out for 30 min using AMV reverse transcriptase (Seikagaku). Sequencing reactions were carried out using 0.4 mM of each dNTP and 0.2 mM of each ddNTP. Reactions were resolved on 8% (w/v) denaturing polyacrylamide gels.

RNase H analysis

The reaction conditions of Konforti *et al*. (20) were followed: 50 mM Tris–HCl (pH 8.3), 10 mM DTT, 60 mM NaCl and 0.1 pmol of primer A6H-1. Digestion was performed using 2.5 U of RNase H (Epicentre) for 30 min at 37°C. Reactions were run out on 8% (w/v) denaturing polyacrylamide gels and blotted onto Nytran for 30 min at 0.5 mA/cm^2 using a TE77 SemiPhor electroblotter. Cleavage products were identified using northern hybridization with mRNA- and gRNA-specific probes as indicated in Figure 4.

Secondary structure predictions

Sequences were analyzed using programs in the GCG software package. MFOLD was used to predict RNA secondary structures and Plotfold was used to generate connect files (21). Currently, no program is available that can fold two separate molecules. Therefore, the gRNA and mRNA were joined using a linker of 10 non-base pairing N residues. No changes were observed using linkers of increasing size. Connect files were imported into RNAdraw (22) to graphically display the predictions. Based on the location of crosslinks mapped, the 3′-most uridylate was base paired to the appropriate base in the pre-mRNA using the force option of MFOLD. The temperature parameter was set at 27°C, the optimal growth temperature of insect-stage trypanosomes.

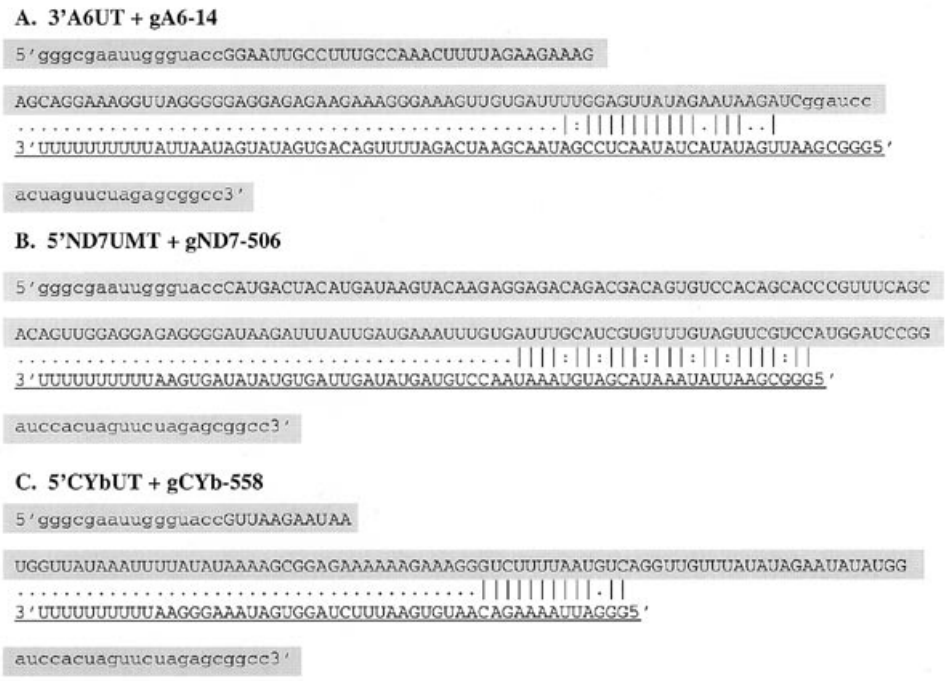


Figure 1. Sequence and alignment of the gRNA/mRNA substrate pairs. The sequence of the unedited mRNA transcript (shaded gray) is shown with the gRNA sequence (underlined) aligned beneath the editing domain. The complementarity between the gRNA and mRNA in the anchor duplex region is shown with : indicating G:U base pairing and | for Watson–Crick base pairing. Vector sequence located at the 5′- and 3′-ends of the mRNA are shown in lower case. (**A**) 3′A6UT + gA6-14; (**B**) 5′ND7UMT + gND7-506; (**C**) 5′CYbUT + gCYb-558.

RESULTS

gRNA and mRNA substrates utilized

To investigate the contribution of the 5′ anchor and 3′ U-tail to the gRNA/mRNA interaction, crosslinking experiments were carried out using three different gRNA/mRNA pairs (Fig. 1).

gA6–14 and 3′A6UT. The ATPase 6 (A6) pre-mRNA is edited throughout the lifecycle of the trypanosome and is the substrate used for *in vitro* editing assays (5,9,10,19,23). 3′A6UT, covers 99 nt of the 3′-end of A6. Within 3′A6UT there are 34 editing sites, representing 10 deletions and 81 insertions. The mRNA sequence upstream of the anchor duplex is 74% purine, making it an ideal substrate for interaction with the gRNA 3′ U-tail, as predicted by the cleavage–ligation model.

gND7–506 and 5′ND7UMT. Editing of NADH dehydrogenase subunit 7 (ND7) pre-mRNA is unusual in that editing occurs in two distinct domains (5′ and 3′) and that editing within the domains is differentially regulated (24). 5′ND7UMT contains the 5′ domain which is modified by the addition of 71 uridylates and the deletion of 13 uridylates, at 39 sites. Like A6, the editing domain and 5′-UTR are purine biased (61.5%). However, the 5′ND7 sequence is frequently punctuated by short stretches of pyrimidines (Fig. 1B).

gCYb-558 and 5′CYbUT. Cytochrome b (CYb) pre-mRNA is only edited at its 5′-end resulting in the insertion of 34 uridylates at 13 sites. 5′CYbUT contains 88 nt of the 5′-end of CYb (25). The short (19 nt) editing domain is purine-rich (95%), however, the upstream 5′-UTR is much less so (59% purines). Editing of CYb is developmentally regulated, occurring only during the procyclic and stumpy bloodstream stages (25,26). The gRNAs used in this study (gA6-14, gND7-506 and gCYb-558) are the initiating gRNAs that start the editing cascade of their respective domains (15,23,24).

gRNA/mRNA anchor duplex interactions

The interaction of the gRNA anchor sequence with its cognate mRNA sequence was examined using a crosslinker localized at the 5′-end of the gRNAs. The gRNA anchor sequence has been shown to be required for *in vitro* editing (5), however, its interaction with mRNA has not been shown directly. Furthermore, this enabled us to confirm that the anchor duplex of the gRNA/mRNA pairs used in this study formed correctly, despite the presence of vector sequence in the mRNAs.

To label the 5′-end of the anchors of gND7-506 and gCYb-558, the gRNAs were synthesized in the presence of GMPS (16,27). The resulting thiophosphate group at the 5′-terminus of the gRNA was then coupled to an APA group. The gCYb-558 anchor begins at the approximate 5′-end of the synthesized gRNA, making it an appropriate substrate for this modification (Fig. 1C). However, the anchors of both gND7-506 and gA6-14 do not begin precisely at the 5′-end of the synthesized gRNA, precluding the use of GMPS to label the anchor region. In order to examine an additional gRNA/mRNA anchor interaction, the sequence of the ND7 mRNA was modified (5′ND7UMT) so as to extend the gRNA/mRNA anchor duplex to the 5′-end of the synthesized

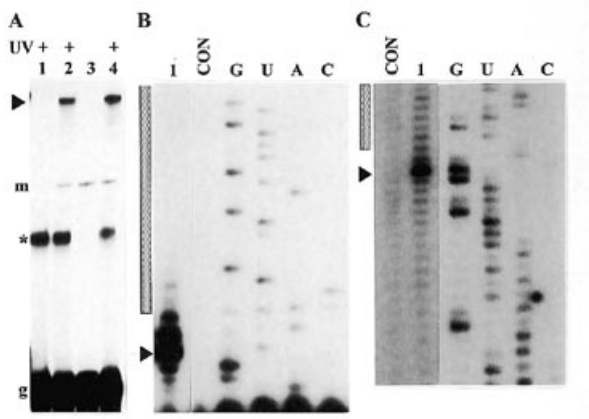


Figure 2. Identification and mapping of 5′ modified gRNA/mRNA intermolecular crosslinked species. (**A**) Radiolabeled 5′APA-gND7-506 (g) crosslinked with trace labeled 5′ND7UMT (m). Lane 1, gND7-506 alone; lane 2, gRNA + mRNA; lane 3, gRNA + mRNA + lysate; lane 4, gRNA + mRNA + lysate. UV + indicates RNAs irradiated at UV_{315}. Large solid arrowhead indicates the major gRNA/mRNA intermolecular crosslinked conjugate. * indicates gRNA intramolecular crosslink. (**B**) Mapping of 5′APA gND7-506/5′ND7UMT crosslinks by primer extension. (**C**) Mapping of 5′APA gCYb-558/5′CYbUT crosslinks. AMV reverse transcriptase and a ^{32}P-end-labeled primer complementary to the 3′ tag sequence of the mRNAs (BIG SK) were used for primer extension as described in Materials and Methods. Lane 1, primer extension of gRNA-crosslinked mRNAs. CON, control RT of non-crosslinked mRNAs. G, U, A and C denote lanes containing RNA sequencing reactions. In (C), RNA sequencing reactions were photographed from a longer exposure of the same gel. Main termination products resulting from the crosslinked gRNAs are indicated by solid arrowheads. The stippled boxes outline the position of the anchor duplexes.

gRNA (Fig. 1B). This sequence modification resulted in an increase of the anchor duplex region from 12 to 26 bp.

gRNAs and mRNAs were allowed to hybridize under editing conditions (5) and irradiated for 20 min on ice. gRNA/mRNA conjugates were identified on a 6% (w/v) denaturing polyacrylamide gel and isolated. Crosslinks were not obtained in the absence of APA modification or if the modified gRNA was paired with an incorrect pre-mRNA (data not shown).

Single crosslink species were obtained for both gND7-506 and gCYb-558. For both 5′ modified gND7-506 and gCYb-558, irradiation at 312 nm resulted in a single major crosslinked species when the gRNAs were paired with the correct pre-edited mRNA (Fig. 2A and data not shown). A second crosslink, not dependent on the presence of mRNA, was also visible. We assume that this species is a gRNA intramolecular crosslink, but it was not characterized. Some minor crosslinks that were not reproducible between experiments were detected occasionally. Only mRNA-dependent crosslinked species that were formed consistently were analyzed. The positions of the generated crosslinks were mapped along the mRNA using a primer specific for vector sequence located at the 3′-end of the mRNA (BIG SK) and reverse transcriptase which stalls approximately one base before (3′ of) the crosslink (16,28). Therefore, the crosslink positions discussed will refer to the base immediately 5′ of the reverse transcription termination product.

5′ modified gND7-506 produced a primary crosslink (strongest termination product) that mapped to one base 3′ of the predicted anchor duplex (Fig. 2B). Two additional termination products were also observed corresponding to the first and second bases of the mRNA anchor. Reverse transcription of crosslinked 5′CYbUT and 5′ modified gCYb-558 mapped a primary crosslink to two bases 3′ of the expected anchor duplex (Fig. 2C). Again, two other strong termination products were observed, flanking the primary crosslink, one and three bases 3′ of the anchor duplex. In the presence of lysate, changes in the position of crosslinking were not observed for either gRNA (data not shown). In both cases, the major crosslink is just 3′ of the predicted anchor duplex and not at the exact 5′-end of the anchor duplex. This may be explained by the fact that the APA group randomly interacts with C-H and N-H bonds in the immediate proximity, not necessarily with the base it is paired across from. Taking this possibility into consideration, the crosslink data indicate that both gRNA/mRNA anchor duplexes correctly form.

gRNA U-tail interactions

To examine the gRNA U-tail interaction with its mRNA, APA groups were placed at the 3′-ends of gA6-14, gND7-506 and gCYb-558 synthesized with U_{10} tails using the protocol of Oh and Pace (17). *In vivo*, gRNAs have U-tails which average 15 U residues in length (6,7). A U_{10} tail length was chosen for this study as we felt that a U_{10} tail would interact in a similar fashion as the *in vivo* U_{15} tail and because the U_{10} construct gave us less problems with T7 polymerase stuttering and tail length heterogeneity. Crosslinks between the gRNAs and their pre-edited mRNAs were obtained as described above. The sites of crosslinking were determined by primer extension using reverse transcriptase (RT). In most cases, comparison of extension products from crosslinked RNA with reaction products from non-crosslinked RNA and sequencing reactions can identify the individual crosslinked nucleotides. In our case, generation of a crosslink physically links the gRNA to the mRNA. Therefore, termination products that may be due to secondary structure interactions between the gRNA and mRNA cannot be mimicked in our control non-crosslinked RNAs. Hence, interpretation of the RT data must be carefully done.

gA6-14/3′A6UT interaction. Irradiation of 3′ modified gA6-14 produced a single mRNA-specific crosslinked species (data not shown). Reverse transcription of gRNA/mRNA conjugates produced a series of termination products along the mRNA (Fig. 3). A minor termination product was observed within the anchor duplex region located 5 nt from the 5′-end of the anchor duplex (mRNA orientation). Minor termination products were also observed corresponding to stops at residues 3–8 upstream of the anchor duplex. Major termination products were observed at residues 9–14. However, this region of the mRNA contains a triple A, triple G sequence which induces a strong premature termination during control reverse transcription of pre-edited A6 mRNA alone. Minor termination products were also observed farther upstream, however, again, these stops mostly correlate with stops observed in the control lanes. Because of the strong stops in the control RT reactions, it is difficult to interpret this data. However, the ladder of termination products found from 2 to 12 nt upstream of the anchor duplex is clearly not present in the control lanes and we interpret these stops as being due to the presence of a gRNA crosslink. Of these, the strongest crosslinks are at positions 9–12 upstream of the anchor. Strong stops are also observed at positions 13 and 14, but because of the stops found in the control extensions, we cannot determine if these are due to

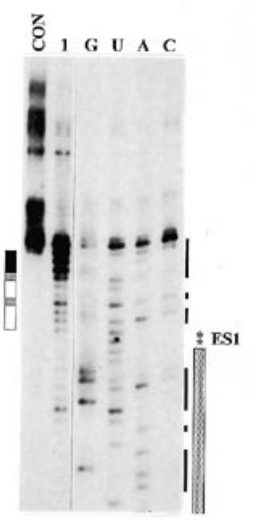


Figure 3. Mapping of 3′APA-gA6-14/mRNA intermolecular crosslinks by primer extension. CON, control lane, primer extension of non-crosslinked 3′A6UT. Strong termination products are observed in the purine-rich region located 13–35 nt upstream of the first editing site (ES1). Lane 1, primer extension of 3′A6UT crosslinked to 3′APA-gA6-14. Unique termination products are observed at nucleotides located 1–12 nt upstream of the first editing site (ES1). RNA sequencing reactions are designated G, U, A and C. The intensities of the termination products are indicated as black (strongest), gray (intermediate) or white (weakest) boxes. Adjacent to the sequence, the black line highlights the purines found within the mRNA sequence. The stippled box indicates the position of the anchor duplex.

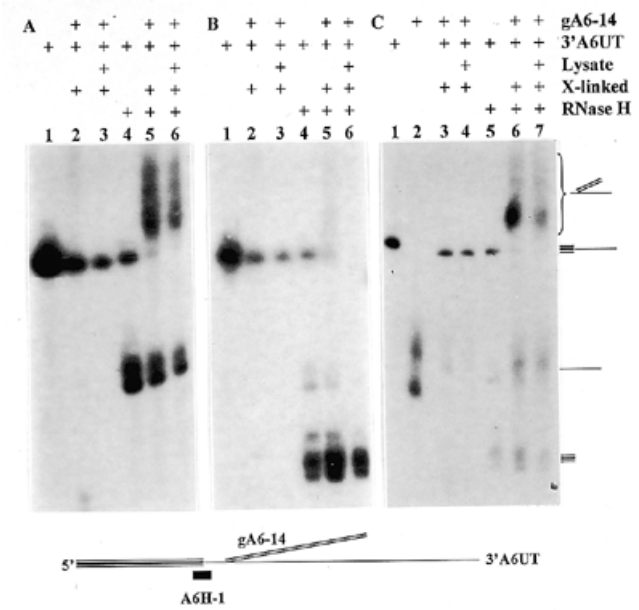


Figure 4. RNase H analysis of 3′ modified gA6-14 and 3′A6UT conjugates. After RNase H digestion using A6H-1, products were resolved on an 8% denaturing gel and blotted onto Nytran. The three probes used were (**A**) BIG SK, complementary to the 3′-terminus of the mRNA, (**B**) T7leadercomp, complementary to the 5′-terminus of the mRNA, and (**C**) C-gA6-14, complementary to the gRNA. (A and B) Lane 1, control undigested mRNA; lanes 2 and 3, control undigested crosslinked RNAs generated in the absence (lane 2) or presence (lane 3) of mt lysate. The full-length crosslinks found near the top of the gel are not shown. Non-crosslinked mRNA can be detected in these lanes despite gel purification of the crosslinked species. Lane 4, control mRNA treated with RNase H and A6H-1; lanes 5 and 6, RNase H + A6H-1-treated crosslinked RNAs generated in the absence (lane 5) or presence (lane 6) of mt lysate. (C) is identical to the other two lanes except that an additional lane (lane 2, untreated gA6-14) was included as a control. On the right hand side are drawings representing the different RNA species. The single, triple and double lines represent the 3′ cleavage fragment, the 5′ cleavage fragment and the gRNA, respectively. In these experiments the mRNA used in generation of the crosslinks was trace labeled for accurate quantitation. Therefore, light bands corresponding to the mRNA fragments can be observed both in (B) (3′ fragment in lanes 4 and 5 and crosslinked fragment in lane 5) and in (C) (3′ and 5′ fragments in lanes 5–7).

the presence of a crosslinked nucleotide. The primer extension stops observed within the anchor duplex region were located just downstream of a 5′-GGAG-3′ sequence in the mRNA anchor. While these stops are not found in the control RT reactions, they may be due to anchor duplex formation between the linked RNAs as this region of the duplex contains three G:C base pairs.

To provide additional confirmation that a gRNA crosslink was responsible for the pattern of termination stops observed, the crosslinked RNA was subjected to oligodeoxynucleotide (A6H-1)-directed RNase H digestion (Fig. 4). A6H-1 hybridizes ~30 nt upstream of the anchor duplex region. Digestion with RNase H in the presence of this oligodeoxynucleotide would cleave the mRNA into two fragments ~76 and 49 nt in length (Fig. 4A and B, lanes 4), corresponding to the 3′ and 5′ halves of the mRNA, respectively. Presence of the crosslinked gRNA to either half would cause it to run with an abnormal electrophoretic mobility. Despite gel purification of the crosslinked species, non-crosslinked mRNA can be detected in the crosslinked RNA lanes (Fig. 4A and B, lane 2, and C, lane 3). In Figure 4A, we can see the 3′ mRNA fragment in the +RNase H lanes in both control (non-crosslinked, lane 4) and crosslinked (lanes 5 and 6) samples. However, in the crosslinked RNA reactions less of the 3′ fragment is observed at the predicted ~76 nt size range compared with the control. Instead a large smear of RNA of slower mobility is observed, indicating that migration of the 3′-half is retarded as would be expected if it were crosslinked to gRNA. Probing of an identical blot with an oligodeoxynucleotide probe specific for the 5′ fragment (Fig. 4B) indicates that the 5′ fragment migrates at the predicted size. The presence of gRNA in the slower migrating species was further confirmed by probing with an oligodeoxynucleotide specific for gA6-14 (Fig. 4C).

gND7-506/5′ND7UMT interaction. Crosslinking of modified gND7-506 in the presence of 5′ND7UMT resulted in two conjugate bands of different electrophoretic mobilities (data not shown). Reverse transcription analyses of both species indicated that the gRNA was crosslinked to the mRNA in approximately the same position (data not shown). Reverse transcription generated a ladder of termination products very similar to that observed for gA6-14 + 3′A6UT (Fig. 5). Termination products not present in the control RT reactions were observed beginning just 5′ of the anchor duplex and extending 28 nt further upstream. Two populations of dominant termination products, separated by a single base, were observed, corresponding to crosslinks 26–28 and 21–24 nt upstream of the anchor. Distinct termination products were also observed at nt 8–11, 13–15 and 17–19. Termination products were again observed upstream of nt 28, however, corresponding stops are also observed in the control RT

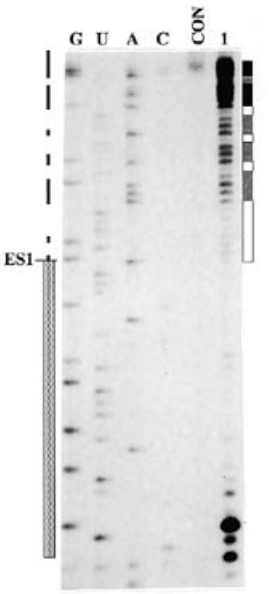


Figure 5. Mapping of 3′APA-gND7-506/5′ND7UMT intermolecular crosslinks by primer extension. G, U, A and C are RNA sequencing lanes used to map the position of the crosslinked gRNA. A primer extension reaction using non-crosslinked 5′ND7UMT is shown in the control (CON) lane. The sequence and control lanes were photographed from a longer exposure of the same gel. Lane 1 contains the primer extension products of 5′ND7UMT crosslinked to 3′APA-modified gND7-506. Strength of the termination products, positions of purine nucleotides in the mRNA and the anchor sequence are indicated as in Figure 3. ES1 marks the position of the first editing site.

reactions. In addition, a strong stop was observed at the start of the anchor duplex (3′, mRNA orientation). These stops differed from those observed when analyzing the 5′ (anchor duplex) crosslinked reactions, in that the stops correlate with the first 3 nt of the anchor duplex with the strongest stop at the third nucleotide. These termination products may be due to the enzyme having trouble reading through the 26 nt anchor duplex formed between the gRNA and the mRNA.

gCYb-558/5′CYbUT interaction. For the last pair of RNAs analyzed, gCYb-558 and 5′CYbUT, we utilized gRNAs with two different U-tail lengths; U_{10} and U_5. In RNA interactions utilizing the 3′-end modified U_{10} gCYb, a single major and two minor mRNA-dependent crosslinked species were identified (Fig. 6A, U_{10}). These individual crosslinked species are designated numerically beginning with the species migrating most slowly in the gel (B1, B2 and B3). Crosslinked species of similar mobilities were also observed when the reactions utilized 3′-end modified gCYb with U_5 tails. However, in reactions with U_5 gCYb, the B2 and B3 crosslinked species were more pronounced (Fig. 6A, U_5). Analysis of the most abundant crosslinked species (B1) generated with U_{10} gCYb again produced a ladder of termination products beginning just 5′ of the anchor duplex and extending ~17 nt upstream (Fig. 6B, lanes 3 and 4). The strongest stops observed were 14–16 nt upstream of the anchor duplex. Closer to the anchor (3–13 bases) are minor crosslinks followed by three stronger termination products corresponding to the last base of the anchor and the two bases just 5′ of the anchor (mRNA orientation). In addition, termination products were also observed

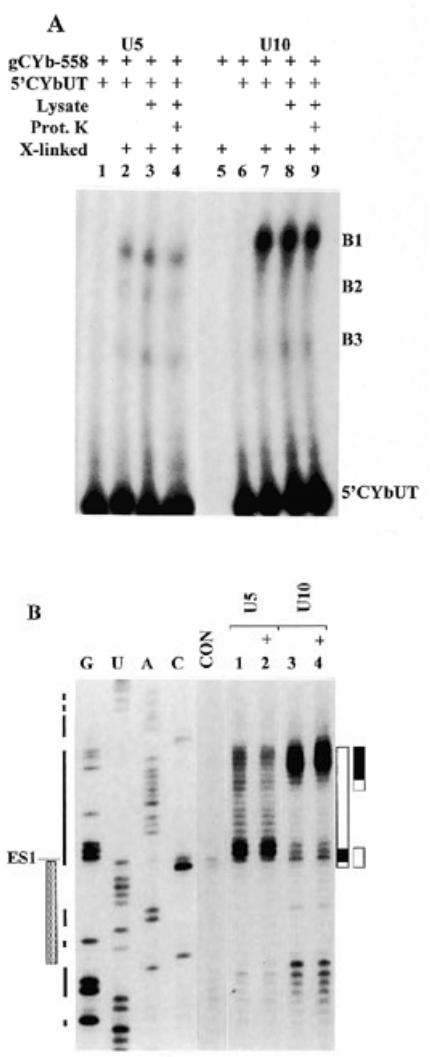


Figure 6. Identification and analysis of 3′APA-modified gCYb-558 crosslinked to 5′CYbUT. (**A**) Identification of intermolecular crosslinks using either gCYb-558U_5 or gCYb-558U_{10}. No crosslinks were obtained in the absence of UV treatment (lanes 1 and 6) or in the absence of mRNA (lane 5). Three crosslinked species (B1, B2 and B3) were obtained with gRNAs with both U_5 and U_{10} tails. The presence or absence of editing active lysate did not affect the ratio of crosslinks obtained (lanes 2, 3, 7 and 8). Proteinase K treatment after crosslinking in the presence of lysate did not affect the mobility of the crosslinked species (lanes 4 and 9). (**B**) Mapping of 3′APA-gCYb-558/mRNA B1 intermolecular crosslinks by primer extension. G, U, A and C are RNA sequencing lanes. CON, control lane containing primer extension products of non-crosslinked 5′CYbUT; lanes 1 and 2, primer extension termination products obtained when the mRNA is crosslinked to gCYb-558U_5; lanes 3 and 4, termination products obtained when the mRNA is crosslinked to gCYb-558U_{10}; lanes 2 and 4, extension products of crosslinks obtained in the presence of editing-active lysate. Position of the anchor duplex, purine nucleotides in the mRNA and intensity of termination products are indicated as in Figure 4. ES1 indicates the position of the first editing site. A decrease in the U-tail length from 10 to 5 uridylates shifts the positions of the dominant crosslinks to nucleotides just 5′ of the first editing site.

at the 3′ boundary (mRNA orientation) of the anchor duplex. These termination products were not consistent in their appearance,

however, and varied from being quite pronounced (Fig. 6B, lanes 3 and 4) to being almost non-existent.

Analyses of the same mobility conjugate (B1) generated with the U_5 gCYb showed a series of termination products that spanned the same nucleotides as those observed with U_{10} (Fig. 6B, lanes 1 and 2). However, the dominant products had shifted so that the two strongest stops correlate to crosslinks with the nucleotides that flank the first editing site. Similar inconsistent primer extension stops were also observed at the 3′ boundary of the anchor duplex region (mRNA orientation) for the U_5 gCYb crosslinks. While in Figure 6B the intensity of the termination products in this anchor region were much more pronounced for the U_{10} gRNA substrates, in other experiments, no difference between the U_5 and U_{10} substrates was observed. Analyses of the faster mobility conjugates (B2 and B3) indicate that in these species, the gCYb U-tails were crosslinked to very different regions (data not shown). For both U_5 and U_{10} gCYb, the B2 band generated a single dominant termination product that correlated with a crosslink located within the 5′-UTR which is not edited in the mature message. The B3 conjugates also mapped to the same position for the U_5 and U_{10} tails with the crosslink located within the 5′ vector sequence of 5′CYbUT.

Incorporation of crosslinking data into secondary structure models suggests that all three gRNA/mRNA pairs interact to form similar structures

The RT analyses indicate that for all three gRNA/mRNA substrate pairs, the gRNA U_{10} tail interacts with the mRNA in a region just upstream of the anchor duplex (Fig. 7). For gA6-14 and gCYb-558, crosslinking of the terminal uridylate occurred relatively close to the anchor duplex with the preferred sites located from 10–12 and 13–16 nt 5′ of the anchor duplex, respectively (Fig. 7B and F). The terminal uridylate of gND7-506 crosslinked farther from the anchor duplex with the preferred sites at 21–28 nt upstream (Fig. 7D). These crosslinking data were incorporated into the computer-predicted secondary structure models by instructing the program to pair the U_{10} nucleotide with the strongest crosslink site. When this was done, the model secondary structures generated were all very similar and differed substantially from the initial computer predictions (Fig. 7). In all cases, the anchor duplex region is correctly paired and any secondary structure in the immediate editing domain is eliminated (compare Fig. 7A and B for 3′A6U, Fig. 7C and D for 5′ND7UMT and Fig. 7E, F and G for 5′CYbUT). In addition, the gRNA guiding region potentially forms a stem–loop structure positioned across from the first editing site.

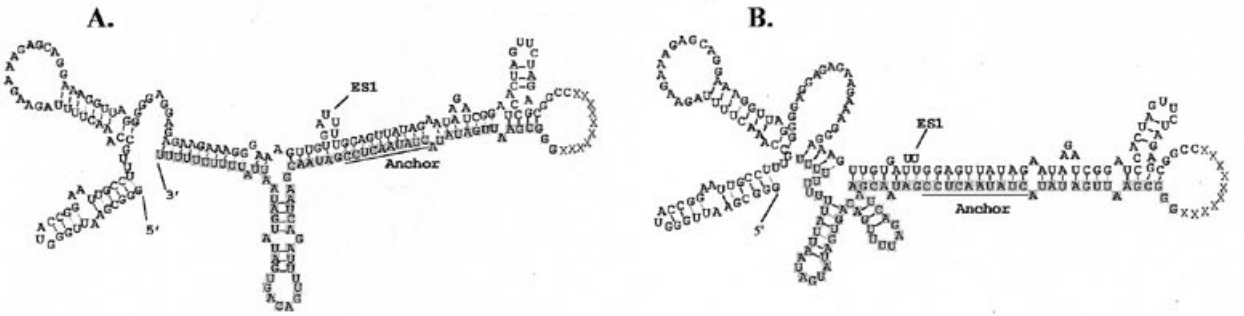
DISCUSSION

Guide RNAs are an essential component of the editing process, supplying the information for the multitude of uridylate insertions and deletions that must occur. Little is known, however, about how the gRNAs are used to direct RNA editing. To approach this question, we have begun to analyze the interactions between gRNAs and mRNAs using comparative photoaffinity crosslinking techniques. Placement of the APA group at the 5′-end of the gRNA allowed us to analyze duplex formation between the gRNA anchor and the mRNA. For both gND7-506 and gCYb-558, the two gRNAs for which anchor duplex interactions were analyzed, the crosslink data suggests that the predicted anchor duplexes correctly form. RT mapping of 5′ crosslink conjugates indicated that the crosslinks were restricted to 2–3 nt surrounding the anchor duplex 3′ (mRNA orientation) border. This may be explained by the fact that the APA group is localized on the 5′-most gRNA nucleotide and can interact with C-H and N-H bonds in its immediate proximity.

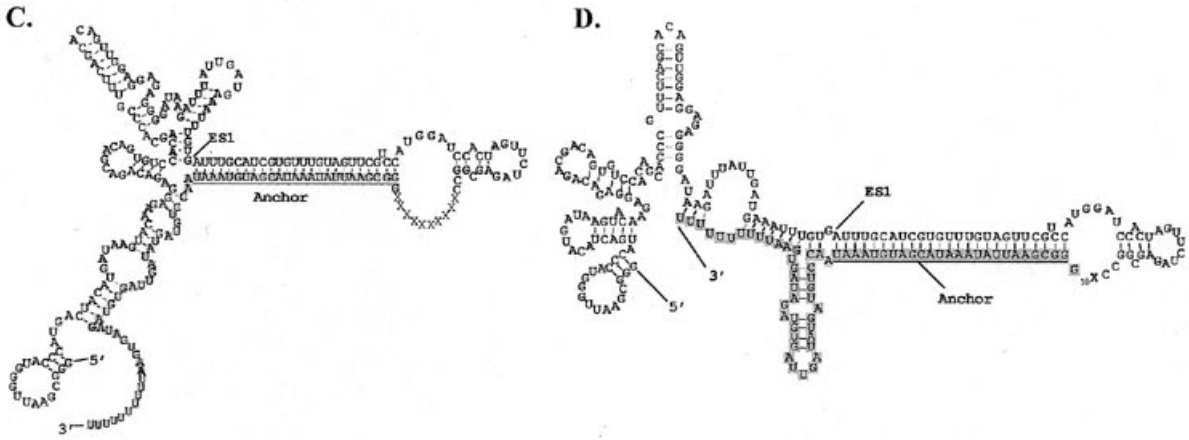
In contrast to the 5′ crosslinks, analyses of gRNA/mRNA conjugates formed when the APA group was placed on the 3′-end of the gRNA U-tail showed a distinctly different crosslinking pattern. RT mapping of the 3′ crosslinked conjugates indicated that the terminal U of the U-tail could interact with a large range of nucleotides located upstream of the first editing site. For all three gRNA/mRNA pairs analyzed, a series of primary crosslinks along with a range of minor crosslinks were detected. In comparing the different gRNA/mRNA interactions, it is interesting that the gRNA U-tails interact in the same relative position, just upstream of the anchor. 3′A6UT is extensively edited throughout most of the message. In the substrate we used, almost no pyrimidine residues are found between 10 and 60 nt upstream of the anchor duplex. However, within the first 10 nt of the anchor, five of the residues are uridylates. Although these five uridylates presumably would not base pair with the gRNA U_{10} tail, strong crosslink sites were mapped at nt 10–12. This suggests that an interaction of the U-tail does not require the entire tail to base pair to the mRNA. For this substrate pair, the strong RT termination signals found in control reactions makes it difficult to determine if stronger crosslinking nucleotides do in fact occur upstream of nt 12. However, we can say that a sizable proportion of the gRNA molecules are crosslinked at nt 10–12 indicating that the U-tail

Figure 7. (Opposite) Comparison of the secondary structure predictions for the interactions of four gRNA/mRNA substrate pairs. Structures A, C and E represent the initial secondary structure predictions generated with no constraints. Structures B, D, F and G were made with a forced base pair between the gRNA U_{10} nucleotide (U_5 for G) and its most dominant crosslink site. The gRNA sequence is shaded gray. The two molecules were linked using a 10 'N' (non-base pairing) linker (represented as X). The anchor duplex regions (underlined, Anchor) and the first editing site (ES1) are indicated. (**A** and **B**) Predicted structures of the gA6-14/3′A6UT interaction. 3′A6UT has a strong purine-rich region upstream of the first editing site (ES1). The most stable interaction in the initial prediction involves the U_{10} tail interacting with a purine-rich region located from 16 to 25 nt upstream of ES1 (A). The predicted structure after input of the crosslinking data is very similar to the initial prediction with the last four uridylates of the U_{10} tail interacting with purines located 7–10 nt upstream of ES1 (B). (**C** and **D**) Predicted structures of the gND7-506/5′ND7UMT interaction. In the initial prediction, the guiding region of the gRNA is predicted to interact with the mRNA well upstream from the region whose editing it directs and the U-tail is not base paired (C). In the predicted structure modified by input of U-tail crosslinking data, the secondary structure in the immediate editing domain is eliminated and the guiding region of the gRNA forms a stem–loop positioned across from the first few editing sites (D). (**E** and **F**) Predicted structures of gCYb-558U_{10}/5′CYbUT interaction. In the initial prediction, the CYb message forms a very stable stem–loop structure, excluding the gRNA (E). After input of U_{10} tail crosslinking data, the U-tail is predicted to interact with a purine-rich region located 6–15 nt upstream of ES1. (**G**) Computer-predicted gCYb-558U_5/5′CYbUT interaction modified by input of U-tail crosslinking data. The gRNA stem–loop structure has shifted, incorporating three of the uridylates in the U-tail into this stem–loop.

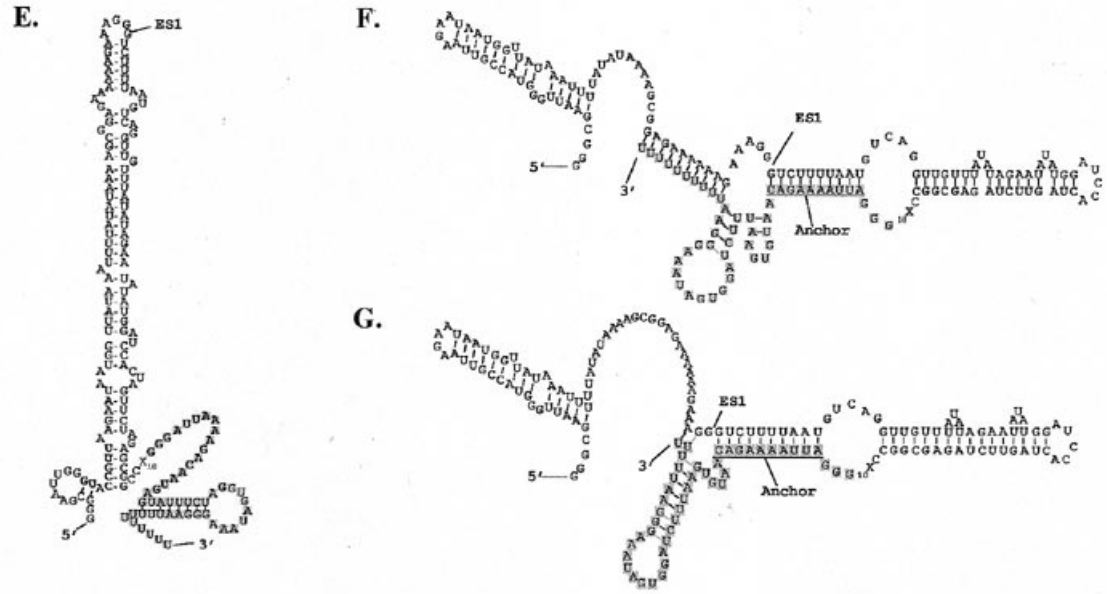
gA6-14 and 3'A6UT



gND7-506 and 5'ND7UMT



gCYb-558 and 5'CYbUT



interaction may involve more constraints than simple purine to U-tail base pairing.

The 5′CYbUT pre-edited substrate differs from 3′A6UT in that the editing domain is quite short, spanning only 19 nt. This region is extremely purine biased with a single pyrimidine found in the 22 nt directly 5′ of the anchor duplex region. Using gCYb-558 with a U_{10} tail produced dominant crosslinks at nt 13–16, just 5′ of the center of this purine-rich region. Distinct crosslinks are again found 3′ (closer to the anchor), but not farther upstream. Shortening the U-tail to just five U residues shifted the dominant crosslinks to within 1–2 nt of the anchor duplex. It is interesting to note that the crosslinks did not move just five bases, but instead shifted 10 bases closer to the anchor. Incorporation of this crosslink data into the computer-predicted secondary structures suggests a shift in the gRNA stem–loop, with three of the uridylates in the U-tail incorporated into the stem–loop structure (Fig. 7D). This again indicates that the entire U-tail does not necessarily base pair with the mRNA (due to the presence of the anchor duplex) and that the position of the U-tail along the mRNA is most likely not driven solely by base pairing interactions with the purine-rich mRNA. It should be noted, however, that shortening the U-tail did alter the populations of gRNA/mRNA conjugates obtained. In the presence of a U_{10} tail, a single dominant conjugate (B1) was always observed, with two minor conjugates (B2 and B3) consistently appearing. The drop in tail length to U_5 shifted the distribution of these populations so that the B2 and B3 conjugates were more abundant. This suggests that the drop in tail length destabilized the most common conformation (detectable by our crosslinking technique).

5′ND7UMT differs from the other two substrates in that it contains only one stretch of 13 purines located from 24 to 37 nt upstream of the anchor duplex. While the strongest crosslinks are found within this purine-rich region, a significant number of crosslinks were also mapped much closer to the anchor duplex in a region which is 50% pyrimidine.

Indeed, for all three RNA pairs, minor crosslinks are observed 3′ of the major crosslinks at almost all nucleotides down to the anchor duplex. This ‘ladder’ pattern of crosslinks is clearly distinct from that observed in the 5′ crosslinking studies. It cannot be explained by reverse transcription read-through nor is it likely that the terminal uridylate bound to a single base while the APA group inserted into a range of neighboring bases. Instead there are two reasonable explanations. (i) *In vitro* transcription of a gRNA with a U_{10} tail results in a population of gRNAs with U-tails of varying lengths. Gel purification improved the homogeneity of the gRNAs, however, the populations used did contain U-tails ranging in size from U_5 to U_{15} (data not shown). This sub-population could be at least partially responsible for the ladder of minor crosslinks. (ii) The interaction of the U-tail with the pre-mRNA may be flexible. The U-tail appears to bind preferentially to a specific region, however, the ability to slide up and down the pre-mRNA sequence could result in the range of minor crosslinks observed. Our data suggests that the heterogeneous populations of gRNAs is unlikely to be the major contributing factor to the minor crosslinks observed. The population of gRNAs showed a Gaussian distribution with respect to the size of the U-tail (data not shown). However, we did not observe a corresponding distribution in the crosslinks. Furthermore, gCYb-558 with a U_5 tail also gave a ladder of termination products which corresponded with those observed with the U_{10} gRNA. The reduction in the number of U residues significantly decreased the amount of stuttering by T7 polymerase. This indicates that the heterogeneity in the gRNA population is not the cause of the ladder of crosslinks observed. Instead, it appears that although the U-tail shows a preference for a particular region, it is capable of binding to a larger range of upstream sequences. This range is constrained, however, in that we did not find crosslinks to the entire range of purine biased sequences available for interaction with the U-tail.

The initial computer structure predictions did not reveal any secondary structures that were common between the interacting RNAs (Fig. 7A, C and E). However, when the 3′ crosslinking data was incorporated into the computer-predicted structures, the structures generated were all very similar (Fig. 7B, D, F and G). In all cases, the anchor duplex region is correctly paired and secondary structure in the mRNA editing domain is eliminated (compare Fig. 7C with D and Fig. 7E with F and G). In addition, the gRNA guiding region forms a stem–loop positioned across from the first few editing sites. These predicted guiding region stem–loop structures are of particular interest as they show similarities to the 3′ stem–loop structures identified in gRNAs by structure probing experiments (14). Schmid *et al.* (14) determined the secondary structures of four different gRNAs from *T.brucei* using a combination of temperature-dependent UV spectroscopy and chemical and enzymatic probing techniques. Alone, gRNA molecules fold into two hairpin elements separated by a single-stranded region of variable length. The 5′-ends of the four gRNAs investigated were all found to be in a single-stranded conformation followed by a small hairpin that contains the anchor sequence. The second hairpin element (stem–loop II) involves the guiding region of the gRNA and is the more stable of the two hairpin elements. For all four gRNAs, the oligo(U) tail was found to be in a single-stranded conformation. The structure predicted for the gA6-14/A6 mRNA interaction shows two hairpins in the guiding region of the gRNA. The 3′-most stem–loop is identical to the 3′ stem–loop observed for gA6-14 alone. The predicted structure for gND7-506/5′ND7UMT also contains a stem–loop in the gRNA that contains many of the same bases as observed in the stem–loop formed by the gRNA alone. Neither of the predicted two stem–loops in gCYb-558 are identical to the identified 3′ stem–loop in gCYb-558. This may be due to the fact that the gCYb-558 sequence we generated differs slightly from that utilized by Schmid *et al.* (14) in their structure probing experiments.

As suggested by Schmid *et al.* (14), the gRNA 5′-most weak stem–loop involving the anchor region would have to melt out in order for it to form the gRNA/mRNA anchor duplex that initiates the editing events. If the second stem–loop (stem–loop II) is maintained during the initial interaction, the U-tail interaction with the mRNA might be constrained, so that it would tend to interact with relatively close upstream regions. This may be what is limiting the ability of the U-tail to interact with mRNA sequence farther upstream. The positioning of the U-tail near the anchor duplex may also explain the generation of chimeric gRNA/mRNA molecules. Most of the chimeras generated *in vitro* and characterized *in vivo* have gRNAs with very short or no U-tails covalently linked to the first few editing sites (7,4,12,29). If the predicted gRNA stem–loop is maintained during the initial editing events, gRNAs missing a U-tail would have their 3′-ends positioned very near the active editing sites possibly allowing the ligation of the gRNA 3′-end to the 3′ cleavage product.

We were surprised that the presence of editing-active lysate did not affect the pattern of crosslinks obtained. However, the initial secondary structure studies of gRNAs alone indicated that the presence of mitochondrial proteins did not change the overall structure (30). In detailed studies of gND7-506 complexed with the gRNA binding protein gBP21, Hermann *et al*. (31) found that the protein binds to the guiding region stem–loop (stem–loop II), with the gRNA structure remaining largely unchanged. This association does appear to increase the stability of the gRNA structure. It may be that the interactions we observed are RNA driven with the proteins possibly reinforcing the preferred interaction. Alternatively, it may be that the molar concentrations of proteins present in our editing lysates were not high enough to affect the structures observed. Currently, we are looking to see if specific editing proteins (such as gBP21) can increase the efficiency of our crosslinking interactions or alter the crosslinking patterns observed.

In this initial study of gRNA/mRNA interactions, photoaffinity crosslinking agents localized to the 5′- and 3′-ends of three different gRNAs were used to map the positions of the gRNA 5′ anchor and 3′ U-tail along their cognate mRNAs. Taken together, the data suggests that the gRNA 5′ anchor does position the gRNA by duplexing with the mRNA just 3′ of the editing domain. In addition, our data suggest that the 3′ U-tail also interacts with the mRNA possibly base pairing with purine-rich upstream sequences. While this interaction appears to be flexible in that we always saw a range of crosslinks, our data does suggest that the U-tail interaction is somewhat constrained to a region relatively close to the first few editing sites. Computer modeling of the RNA interactions indicates that the stem–loop II structure, present in free gRNAs, may be maintained in the initial gRNA/mRNA interaction. With the gRNA stem–loop II maintained, the 5′ anchor and the U-tail mRNA interactions flank the first few editing sites, possibly working together to eliminate mRNA secondary structure in the immediate editing domain. This type of interaction may increase the accessibility of the editing complex to the proper editing sites.

ACKNOWLEDGEMENTS

We thank M. Harris and N. Pace for technical and other advice concerning crosslinking protocols, K. Stuart for help in obtaining editing-active lysate and U. Göringer for supplying us with gRNA plasmids and critical comments on the manuscript. This work was support by NIH grant number AI34155 to D.K.

REFERENCES

1 Sloof,P. and Benne,R. (1997) *Trends Microbiol*., **5**, 189–195.
2 Stuart,K., Allen,T.E., Heidmann,S. and Seiwert,S.D. (1997) *Microbiol. Mol. Biol. Rev.*, **61**, 105–120.
3 Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
4 Blum,B. and Simpson,L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11944–11948.
5 Seiwert,S.D., Heidmann,S. and Stuart,K. (1996) *Cell*, **84**, 831–841.
6 Blum,B. and Simpson,L. (1990) *Cell*, **62**, 391–397.
7 Blum,B., Sturm,N.R., Simpson,A.M. and Simpson,L. (1991) *Cell*, **65**, 543–550.
8 Cech,T.R. (1991) *Cell*, **64**, 667–669.
9 Kable,M.L., Seiwert,S.D., Heidmann,S. and Stuart,K. (1996) *Science*, **273**, 1189–1195.
10 Cruz-Reyes,J. and Sollner-Webb,B. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 8901–8906.
11 Sollner-Webb,B. (1991) *Curr. Opin. Cell Biol*., **3**, 1056–1061.
12 Koslowsky,D.J., Göringer,H.U., Morales,T.H. and Stuart,K. (1992) *Nature*, **356**, 807–809.
13 Koslowsky,D.J., Kutas,S.M. and Stuart,K. (1996) *Mol. Biochem. Parasitol*., **80**, 1–14.
14 Schmid,B., Riley,G.R., Stuart,K. and Göringer,H.U. (1995) *Nucleic Acids Res*., **23**, 3093–3102.
15 Riley,G,R., Corell,R.A. and Stuart,K. (1994) *J. Biol. Chem*., **269**, 6101–6108.
16 Burgin,A.B. and Pace,N.R. (1990) *EMBO J*., **9**, 4111–4118.
17 Oh,B.-K. and Pace,N.R. (1994) *Nucleic Acids Res*., **22**, 4087–4094.
18 Pollard,V.W., Harris,M.E. and Hadjuk,S.L. (1992) *EMBO J*., **11**, 4429–4438.
19 Seiwert,S.D. and Stuart,K. (1994) *Science*, **266**, 114–117.
20 Konforti,B.B., Koziolkiewicz,M.J. and Konarska,M.M. (1993) *Cell*, **75**, 863–873.
21 Zuker,M. (1989) *Science*, **244**, 48–52.
22 Matzura,O. and Wennborg,A. (1996) *CABIOS*, **12**, 247–249.
23 Bhat,G.J., Koslowsky,D.J., Feagin,J.E., Smiley,B.L. and Stuart,K. (1990) *Cell*, **61**, 885–894.
24 Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L., Feagin,J.E. and Stuart,K. (1990) *Cell*, **62**, 901–911.
25 Feagin,J.E., Jasmer,D.P. and Stuart,K. (1987) *Cell*, **49**, 337–345.
26 Feagin,J.E. and Stuart,K. (1988) *Mol. Cell. Biol*., **8**, 1259–1265.
27 Sampson,J.R. and Uhlenbeck,O.C. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 1033–1037.
28 Denman,R., Colgan,J., Nurse,K. and Ofengand,J. (1988) *Nucleic Acids Res*., **16**, 165–178.
29 Read,L.K., Corell,R.A. and Stuart,K. (1992) *Nucleic Acids Res*., **20**, 2341–2347.
30 Köller,J., Müller,U.F., Schmid,B., Missel,A., Kruft,V., Stuart,K. and Göringer,H.U. (1997) *J. Biol. Chem*., **272**, 3749–3757.
31 Hermann,T., Schmid,B., Heumann,H. and Göringer,H.U. (1997) *Nucleic Acids Res*., **25**, 2311–2318.