This is to certify that the

dissertation entitled

Characterization of gRNA and mRNA interactions
in trypanosome mitochondrial RNA editing

presented by

Sheldon Se-Chung Leung

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Genetics

Major professor

Date 2/23/01

CHARACTERIZATION OF gRNA AND mRNA INTERACTIONS IN TRYPANOSOME MITOCHONDRIAL RNA EDITING

By Sheldon Se-Chung Leung

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Microbiology
Genetics Program

2001

ABSTRACT

CHARACTERIZATION OF gRNA AND mRNA INTERACTIONS IN TRYPANOSOME MITOCHONDRIAL RNA EDITING

By

Sheldon Se-Chung Leung

Mitochondrial RNA editing in trypanosomes involves the precise addition and removal of uridylates from pre-mRNAs, producing translatable mRNAs. Small RNAs (55-70 nts) known as guide RNAs (gRNAs) direct this process. gRNAs bind to their cognate mRNAs via a 5' anchor sequence while the bases downstream of the anchor guide the editing process. At the 3' end of a gRNA, is a uridine tail whose function(s) has not been clearly defined.

The process of editing also requires the interaction of gRNA/mRNA pairs with a protein complex. This complex interacts with hundreds of gRNA/mRNA pairs, yet these RNAs share no primary sequence motifs that could act as RNA-binding domains. This has led us to hypothesize that the structural features of interacting gRNAs and mRNAs play a role in the specific interaction of gRNA/mRNA pairs with the editing complex.

To examine the structure of gRNA/mRNA pairs, we utilized photo-reactive crosslinking reagents placed at the 5' and 3' ends of gRNAs in order to begin to map the structural relationships between gRNAs and their cognate pre-mRNAs. The results of the crosslinking study confirmed that the 5' anchor sequence of the gRNA basepaired with the mRNA, forming a gRNA/mRNA anchor duplex. In addition, and more importantly,

the data provided the first direct evidence that the U-tail interacted with upstream purine rich sequences. Using these data, similar gRNA/mRNA secondary structure predictions were obtained for the three pairs examined. Each pair formed a gRNA/mRNA anchor duplex, a U-tail/mRNA duplex and a gRNA stem-loop, supporting our overall hypothesis.

Interestingly, the previous study indicated that the gRNA U-tail interacted with pre-mRNA sequences that were to be subsequently edited. This led to an investigation of how the U-tail/mRNA interaction would be affected by editing using crosslinking to map the position of the 3' end of the U-tail along partially edited transcripts. Remarkably, despite the insertion of 6 U's and a doubling of the length of the gRNA/mRNA anchor duplex, the 3' end of the U-tail continued to interact with the same sequence. Secondary structure predictions suggested that the U-tail was involved in the maintenance of the gRNA stem-loop, causing the 3' end of the U-tail to basepair with the same sequence, suggesting that the gRNA stem-loop is an important structural. In addition, preservation of the stem-loop by the U-tail is potentially a novel role for the U-tail.

Solution probing of a 5' crosslinked gRNA/mRNA pair provided additional evidence for our predicted structure of interacting gRNAs and mRNAs. 5' crosslinked molecules are biologically relevant as they support gRNA-directed endonuclease, U-specific exonuclease and terminal uridylyltransferase activities indicating that these molecules interact correctly with the editing complex. The probing data directly shows that the U_{15}-tail protects several mRNA nucleotides predicted to be involved in the U-tail/mRNA duplex. Together with our previous crosslinking studies, these data provide further support for our predicted secondary structure of interacting gRNA/mRNA pairs.

ACKNOWLEDGEMENTS

Many thanks to Dr. Donna Koslowsky, whose lab this work was performed in. Her foresight and enthusiasm to explore our niche in the world of kinetoplastid RNA editing has provided me with a valuable learning experience and a great introduction to the RNA world.

All the members of the lab have helped me along the way, especially Sandi Clement who has been an amazing colleague in the lab. Her advice and encouragement was indispensable. I also shared some fun times in and out of the lab with Catherine Book, Rebecca Zima and Gayani Yahampath.

I must acknowledge my guidance committee Dr. Pamela Green, Dr. Laurie Kaguni and Dr. Ron Patterson for their advice and support. I would like to thank Dr. Green and her lab for letting me participate in joint lab meetings. Not only did these meetings provide me with a different perspective on my work but I also enjoyed learning about *Arabidopsis* RNA processing.

I also owe thanks to Dr. Dipnath Baidyaroy and Jung Park for countless scientific discussions and adventures. Thanks to Jaimie Houghton for dinners, dinners and more dinners. Saori Obayashi, Dawn Greensides, and Kirsten Fertuck have been wonderful friends.

I would also like to recognize the support of my family and my wife Sue.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

INTRODUCTION

RNA editing was first discovered in 1986 in the kinetoplastid protozoa, *Trypanosoma brucei* and *Crithidia fasciculata*. While sequencing the mRNA of the cytochrome oxidase subunit II (COII) gene, Benne et al. (1986) found the insertion of 4 extra U's not encoded in the mitochondrial genome. This was totally unexpected, as the prevailing dogma was that DNA templates contained all of the protein coding information carried by mRNAs. Many examples of RNA editing (with distinctly different mechanisms) have since been identified in a wide range of organisms outside of the kinetoplastid protozoa, but is most noticeably absent from bacteria.

Trypanosoma brucei brucei

This African parasite is a serious economic pest that infects domesticated livestock, causing the fatal disease, Nagana. The vectors for this parasite are several *Glossina* species commonly known as tsetse flies. Cycling through these two hosts results in a complex lifecycle that requires a tremendous amount of developmental gene regulation at many levels. Our lab is specifically interested in the developmental regulation of mitochondrial gene expression. Differentiation from the bloodstream form to the insect form involves a dramatic shift in metabolism. In mammals, the slender bloodstream form's primary source of energy is derived from glycolysis which is sequestered in the unique organelle, the glycosome. Mitochondrial activity is minimal as cytochromes and several enzymes involved in the citric acid cycle are not present. Transformation into the stumpy bloodstream form results in a modest increase in mitochondrial activity with the

initiation of synthesis of enzymes involved in mitochondrial oxidative phosphorylation. Once ingested by the fly, the stumpy form is induced to differentiate into the procyclic (insect) form. The procyclic mitochondrion is fully functional and becomes the primary source of ATP (Bienen et al., 1991; Clarkson et al., 1989).

Mitochondrial DNA

Kinetoplastid mitochondrial DNA consists of two classes of circles known as maxicircles and minicircles. These DNA molecules are catenated to form a unique and large network known as a kinetoplast. Analogous to mitochondrial genomes of other organisms, the maxicircle encodes several ribosomal genes and other genes required for mitochondrial activity. There are about 50 copies of the 22 kB maxicircle per mitochondrion. Minicircles are much more abundant with 5,000-10,000 copies per network. Each of the 1 kB circles encode three guide RNAs (gRNAs) which are essential components of editing. While the functions of both circles have been identified, the role of the network has yet to be defined.

General Aspects of Editing

RNA editing in the mitochondria of kinetoplastids involves the precise post-transcriptional insertion and deletion of uridylates (U's) from mRNA transcripts. Editing is an essential process that produces translational start and stop codons as well as open reading frames. The isolation and N-terminus amino acid sequencing of apocytochrome B (CYb) demonstrated that mitochondrial edited transcripts were translated *in vivo* and that editing in apocytochrome B mRNA created the AUG start codon (Horvath et al.,

2000). In *T. brucei* a significant portion of mitochondrial transcripts are edited, with twelve of seventeen mRNAs modified. The amount of editing within each mRNA varies; ranging from 4 U's inserted (COII) to 552 U's added and 88 U's deleted (NADH dehyrdogenase subunit 7, ND7). Regulation of editing is also both stage and transcript specific resulting in three categories of mRNAs: (1) only edited in bloodstream forms, (2) only edited in procyclic forms and (3) constituitively edited (reviewed in (Alfonzo et al., 1999; Estevez and Simpson, 1999; Hajduk and Sabitini, 1998; Stuart et al., 1997)). ND7 is an exception as it contains two editing domains. The 5' domain is edited in both stages while the 3' domain is only edited in the bloodstream form, resulting in two forms of the transcript and most likely, two forms of the protein.

Guide RNAs

Small RNAs (50-70 nts) known as guide RNAs (gRNAs) direct the precise insertion and deletion of U's (Blum et al., 1990). They were identified by their complementarity to short stretches of edited mRNAs. All identified gRNAs contain three functional elements. At the 5' end of the molecule is the anchor, a short sequence (4-15 nts) that is complementary to the mRNA sequence. It serves to basepair the gRNA with its cognate mRNA forming an anchor duplex. Therefore gRNAs act in *trans*, with the only exception being COII whose gRNA is co-transcribed at the 3' end of the message. Immediately 3' to the anchor is the information or guiding sequence. This sequence directs editing as U's are inserted and/or deleted until the mRNA becomes fully complementary to this guiding sequence. It is important to note that during editing, U's can be inserted across from G's in the guiding sequence. Therefore, gRNAs cannot be

considered conventional templates, as characterized RNA polymerases do not incorporate U's across from G's. Finally, at the 3' end of the gRNA is a post-transcriptionally added U-tail. The average length of the U-tail is estimated to be 15 nts *in vivo* (Blum and Simpson, 1990). The function of the U-tail has not been fully resolved although several roles for the U-tail have been proposed. It has been hypothesized that the U-tail could act as an anchor in addition to the 5' anchor of the gRNA, by basepairing to purine rich sequences in the mRNA, thus stabilizing the gRNA and mRNA interaction (Blum and Simpson, 1990). Alternatively, it has been proposed that the U-tail is a reservoir from which U's flow to and from the mRNA during the insertion and deletion of U's (Blum et al., 1991; Cech, 1991). However, current evidence does not support the latter hypothesis, which will be discussed in detail below.

gRNAs participate in the editing process despite their different sequences. This prompted the Göringer lab to examine the secondary structure of gRNAs to determine if they formed a common structural motif that could be recognized by proteins involved in editing. Enzymatic and chemical probing of gRNAs revealed that these molecules did indeed share a common secondary structure (Schmid et al., 1995). The gRNA's anchor sequence formed a weak stem-loop that would be expected to easily melt in order for a gRNA to bind an mRNA. The guiding sequence formed a more substantial stem-loop, while the U-tail was shown to be single-stranded. The guiding sequence stem-loop is bound by the protein gBP21 (Hermann et al., 1997). This protein is hypothesized to stabilize the stem-loop, supporting a role for structure in gRNAs. Although gBP21 is not required for editing, it has been shown to tightly associate with the editing machinery (Lambert et al., 1999).

Direction of Editing

The overall direction of editing is 3' to 5' along the mRNA. gRNAs are short RNAs that can only direct the editing of a small stretch of RNA. Extensively edited mRNAs therefore require multiple gRNAs to guide editing. A survey of gRNAs revealed that only a fraction of gRNAs contain anchors that are complementary to never edited sequences, while the rest contain anchors complementary to edited sequences. This suggested that the small fraction of gRNAs would interact first, immediately downstream of the first editing site. By directing editing of the mRNA they would produce the complementary sequence for the next gRNA's anchor (Maslov and Simpson, 1992). This is supported by the characterization of partially edited mRNAs, which were found to be edited at the 3' end while unedited at the 5' end. In addition, because the editing process causes gRNAs and mRNAs to be basepaired , one would predict that this would block the next gRNA from basepairing (Blum et al., 1990). The removal of a gRNA to allow the next gRNA to anneal could be carried out by mHEL61p, a mitochondrial DEAD-Box helicase (Missel et al., 1995). The knockout of this protein resulted in a significant reduction in the accumulation of fully edited mRNAs (Missel et al., 1997). However, mitochondrial lysate lacking the helicase was able to complete a single editing event. These results support a role in which the helicase is required to remove gRNAs to allow the next gRNA to anneal, enabling editing by more than one gRNA.

Editing within the domain of a single gRNA

It was initially thought that the choice of editing sites was simply a matter of a mismatched basepair immediately upstream of the anchor duplex. However, a significant fraction of partially edited mRNAs isolated from mitochondria contained edited sequence that did not match edited or unedited RNAs. This sequence was termed a junction (Koslowsky et al., 1991). Upstream of the junction was unedited sequence, while the sequence 3' of the junction was correctly edited. It was speculated that these molecules were undergoing editing within the junction when they were isolated. Due to the amount of this type of partially edited RNA it is difficult to dismiss them as "dead-end" products. In light of this, three models have been proposed to explain these RNAs as intermediates of editing. Decker and Sollner-Webb (1990) hypothesized that these molecules could be created if editing within a block directed by a single gRNA was random. Fortuitous insertions and deletions that would extend the anchor duplex would be protected via the gRNA, preserving correctly edited sites. This would require indiscriminate endonucleolytic cleavage of the mRNA. However, current *in vitro* editing assays support an editing mechanism in which endonucleolytic cleavage is gRNA directed and not random. This contradicts the above model if these partially edited mRNAs are intermediates. Another possibility is that editing does occur in a strict 3' to 5' direction but that these editing events have been directed by an inappropriate or misaligned gRNA. RNAs with junctions that were complementary to inappropriate gRNAs have been isolated (Sturm and Simpson, 1990). Finally, editing could be dictated by the stability of the interaction of the gRNA and mRNA. By editing in a stepwise fashion at sites that progressively increase the thermodynamic stability of the gRNA/mRNA interaction,

editing could proceed until the guiding sequence and mRNA are fully basepaired (the most stable interaction) (Koslowsky et al., 1991). During these cycles of realignment, the partially edited mRNA will not necessarily match the final edited sequence. Although there is no evidence to exclude either of the latter two models, mis-editing by inappropriate gRNAs would produce a dead-end product. This would require an additional investment of resources to correct these non-functional products. In the thermodynamic model, RNAs containing junctions are direct intermediates of editing. Further resolution of this issue will require the development of *in vitro* assays capable of multiple rounds of editing.

Models of Editing

Three models have been proposed to describe the mechanism of editing: Cleavage-Ligation (CL), Transesterification (TE), and Cleavage-Ligation with Chimera (CL-C). Two important differences distinguish the CL model from the other two models: (1) both the TE and CL-C models hypothesize that the U-tail is the source of U's for the editing process and (2) this predicts the production of an editing intermediate, termed a chimera. Chimeras are hybrid molecules consisting of an mRNA truncated at the 5' end, and joined to a gRNA via its U-tail. Currently, there is mounting evidence favoring the CL model, as will discussed in a later section.

Cleavage-Ligation. The initial step in this model is gRNA directed endoribonuclease cleavage (Figure 1) (Blum et al., 1990). Cleavage is always immediately 5' to the anchor duplex at the first mismatch. This is because U's are added or removed from the 3' end of

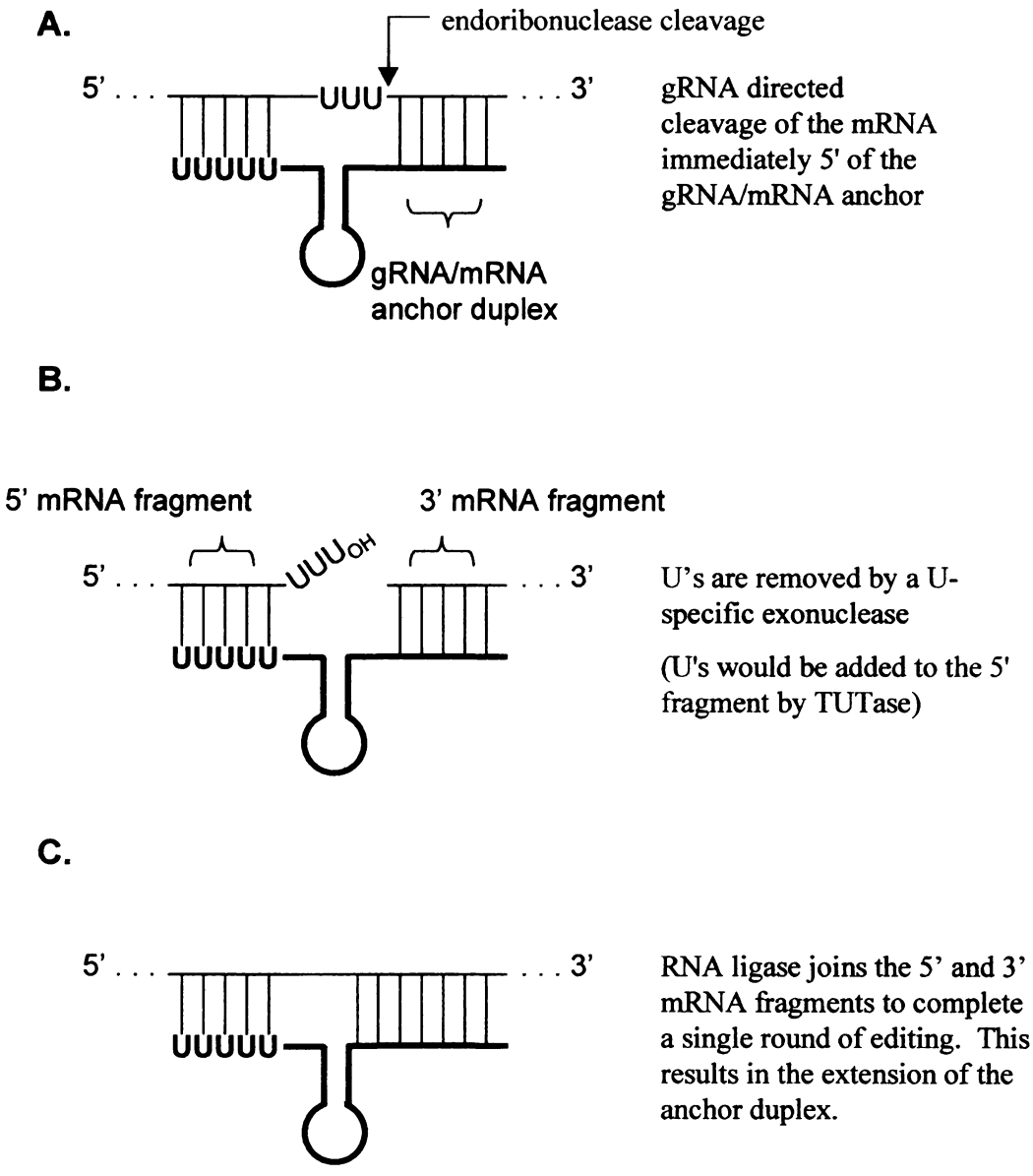


Figure 1. The steps of the Cleavage-Ligation model of editing. A. Endoribonuclease cleavage, B. Deletion of U's and C. Ligation. The mRNA is shown on top, while the gRNA is in bold on the bottom. The gRNA U-tail is represented by a series of bold U's.

the 5' cleavage product, implying that the U's would cycle to and from a pool of free UTP, not the gRNA's U-tail. Recent evidence suggests that deletion and insertion activities are carried out by a U-specific exonuclease and a terminal uridylyl transferase, respectively (Cruz-Reyes and Sollner-Webb, 1996). The final step of rejoining the cleavage fragments is carried out by an RNA ligase.

Transesterification. The recovery of chimeras from mitochondrial RNA, led to the hypothesis that editing could occur via two transesterification reactions (Blum et al., 1991; Cech, 1991). The first transesterification reaction would involve nucleophilic attack by the 3' OH of the U-tail, at the phosphodiester bond in the editing site, joining the gRNA to the mRNA and releasing a 5' mRNA product. With the U-tail linked to the gRNA, it was tempting to suggest that U's would move to and from the U-tail during deletion and insertion events. This model requires that editing sites not only be defined by a mismatch between the gRNA and mRNA immediately upstream of the anchor duplex, but also whether the insertion or deletion of U's was required (Figure 2B). Deletion events would involve nucleophilic attack at the phosphodiester bond 5' of the U's to be removed. To insert U's, the attack would occur at the phosphodiester bond just 5' to the anchor duplex. In both cases, the number of U's deleted or inserted would be controlled by the second transesterification reaction. The 3' OH of the 5' cleavage product would attack the U-tail at the appropriate phosphodiester bond to remove or add the correct number of U's. This would also rejoin the 5' and 3' mRNA fragments, and detach the gRNA. The appeal of this model was that no additional energy source was required and a similar mechanism is employed in RNA splicing.

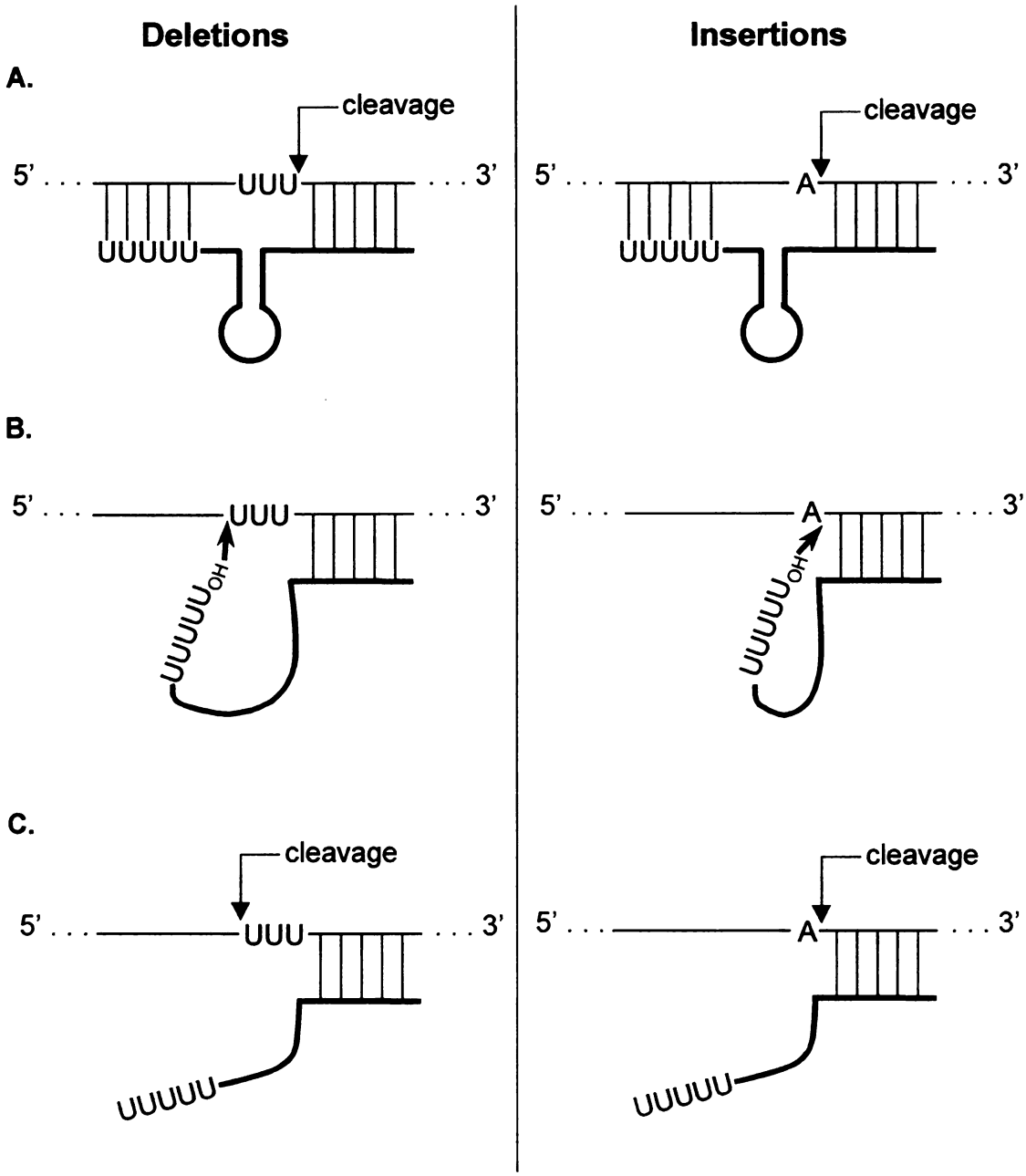


Figure 2. Selection of editing site. A. Cleavage Ligation, B. Transesterification, C. Cleavage Ligation with chimera. Deletion activity would remove the 3 U's in the mRNA. Insertional activity would add U's between the A and the gRNA/mRNA anchor duplex. The mRNA is shown on top while the bold line on the bottom represents the gRNA

Cleavage-Ligation with Chimera. This model also incorporated the chimera as an editing intermediate (Cleavage-Ligation with Chimera, CL-C) (Rusché et al., 1995) using essentially the identical steps of the TE model, except that endoribonuclease and ligase activities would replace the transesterification steps. The initial step of cleavage is also influenced by the event (deletion or insertion) at the editing site (Figure 2C). Cleavage would occur upstream of the U's to removed and just 5' of the anchor duplex during insertion. This cleavage would produce 5' and 3' mRNA fragments, but obviously no chimera. Chimera formation would occur via ligase activity, joining the gRNA's U-tail to the 3' cleavage product's 5' end. A second endonucleolytic cleavage would act on the U-tail to control the number of U's inserted or deleted. Finally a ligase would join the 5' and 3' mRNA fragments together. The enzymatic activities required in this model have been identified (Bakalara et al., 1989; Pollard et al., 1992).

Biochemical Evidence for a model of RNA editing

The development of *in vitro* editing assays was a significant breakthrough allowing for the characterization of the requirements of editing as well as testing of the different models (Byrne et al., 1996; Seiwert and Stuart, 1994). The observations from the experiments described below favor the Cleavage-Ligation model. *In vitro* editing is gRNA directed and single editing events extend the anchor duplex by the insertion and deletion of U's as predicted. Thus modifications in the sequence of gRNA or mRNA lead to corresponding predictable changes in the number of U's inserted or deleted. Components of the *in vitro* editing reaction included mitochondrial lysate, mRNA, gRNA, Mg^{2+}, ATP and UTP (only for insertional editing). The requirement for UTP in

insertional editing suggests that the U-tail is not the source of U's for editing (Burgess et al., 1999).

Recent reports also suggest that optimal conditions for *in vitro* editing at deletion and insertion sites are different (Cruz-Reyes et al., 1998b). Standard editing conditions support robust U deletional activity while insertional activity is only weakly supported. Optimal U insertional editing conditions included lower ATP concentrations (0.3 mM vs. 3 mM) and higher concentrations of UTP (0.15-0.5 mM vs. 0.05 mM). ATP has been suggested to inhibit cleavage at sites of U insertion thereby reducing the amount of edited RNA produced (Cruz-Reyes et al., 1998a). However, as Igo et al. (2000) pointed out, the assay used could not detect if higher ATP concentrations increased ligase activity, religating unedited cleavage fragments, thus inhibiting editing.

Cleavage at editing sites. Mitochondrial lysate contains endoribonuclease activity that could support both of the cleavage-ligation models. Piller et al. (1997) further characterized this activity and demonstrated that there are three endoribonuclease activities present in mitochondrial lysate. These activities can be separated using glycerol fractionation or chromotography. Cleavage by the endoribonuclease that associates with editing activity is gRNA directed. The absence of gRNA or the use of anti-sense RNA to block the anchor of gRNAs inhibits cleavage at the correct editing site (Adler and Hajduk, 1997). The current *in vitro* editing assay in *T. brucei* involves short mRNA substrates allowing for the direct visualization of editing intermediates and products. 3' end-labeling of the mRNA results in observation of a 3' cleavage fragment during editing (Seiwert et al., 1996). The presence of a 3' cleavage product does not

support the transesterification model. Transesterification produces chimeras involving the 3' mRNA fragment ligated to the gRNA.

3' OH group of gRNAs. The role of the 3' OH group was investigated using gRNAs that lacked the hydroxyl group or had a blocked 3' end by co-transcribing the gRNA upstream of the mRNA. In both cases, these gRNAs were able to direct editing, suggesting there is no requirement for a 3' OH group in editing (Burgess et al., 1999; Kapushoc and Simpson, 1999). This also contradicts the transesterification model. These experiments also demonstrated that while editing could still occur in the absence of a 3' OH, chimera formation could not. Therefore the 3' OH is only required to ligate the gRNA to the 3' mRNA fragment to produce chimeras. This separation of editing and chimera formation strongly indicates that chimeras are not RNA editing intermediates. These observations, thus argue against the two models that use chimeras as editing intermediates (TE and CL-C).

Role of the U-tail. Free UTP is an absolute requirement for insertional editing, suggesting the U-tail is not the source of U's for editing. This raises questions concerning the function(s) of the U-tail. Several observations from *in vitro* editing experiments have provided clues to the role of the U-tail. Removal of the U-tail does not affect gRNA directed cleavage at the editing site (Seiwert et al., 1996), however, it does increase chimera formation and block the production of edited mRNA. Substituting the U-tail with a sequence that basepairs with a higher affinity to the mRNA also supports accurate cleavage, but chimera formation is abolished and editing activity is restored (Seiwert et

al., 1996). In the *Leishmania tarentolae* editing assay, such a substitution actually increased the amount of edited product (Kapushoc and Simpson, 1999). Finally, experiments using mRNAs with gRNAs supplied in *cis* upstream of the mRNA, also do not require a gRNA U-tail to support editing (Kapushoc and Simpson, 1999). These observations clearly indicate that a poly U tract at the end of a gRNA is not required for a single round of editing. However, the U-tail does appear to play a role in the suppression of chimera formation. It is hypothesized that after cleavage at the editing site, the U-tail is required to tether the 5' fragment, preventing it from being lost from the editosome (Seiwert et al., 1996). If this fragment is lost, the 3' end of the gRNA may be inadvertently ligated to the 3' fragment creating a chimera. In the *in vitro* editing assay, the U-tail is not 100% effective at preventing chimera formation as they are readily detected. However, *in vivo*, chimeras are extremely rare, and are found at levels well below a single copy per cell and up to 5000 times lower than edited mRNAs (Riley et al., 1995). This suggests that the *in vitro* editing assay is missing a component that might help to stabilize the U-tail/mRNA interaction. Such an interaction is mimicked by substituting the U-tail with a sequence that is perfectly complementary to the 5' cleavage fragment, or by supplying the gRNA in *cis*, upstream of the mRNA (Kapushoc and Simpson, 1999). However, one cannot exclude the possibility that chimeras are rapidly degraded or even corrected *in vivo*.

These experiments do not directly address the additional hypothesized function of the U-tail as a supplementary anchor. However, the observation that RNAs with *cis* acting gRNAs edit as efficiently as reactions in which *trans* acting gRNAs are supplied at a 15-30 fold molar excess (intra vs. intermolecular interaction) suggests that the rate of

mRNA and gRNA association may be a rate limiting step in editing (Kapushoc and Simpson, 1999).

Complexes and Proteins Involved in Editing

The editing complex. Other RNA processing activities such as RNA splicing and polyadenylation involve ribonucleoprotein complexes (RNPs) that carry out these functions. Likewise, there is increasing evidence that the enzymes involved in RNA editing associate in a complex, termed an editosome. Sedimentation analysis of *T. brucei* mitochondrial lysate indicates that there are two peaks of editing activity (Corell et al., 1996; Pollard et al., 1992). One complex was found at 19-20S while a larger complex was identified at 35-40S. Both of these complexes contain a gRNA directed endoribonuclease, a TUTase and an RNA ligase. However, only the 35-40S complex is associated with mRNAs and gRNAs. This has led to speculation that the 35-40S complex is in the process of editing mRNAs, while the smaller complex represents a fully competent editosome not yet associated with RNAs. This is supported by observations that the 19-20S complex is capable of both insertion and deletion activity (Cruz-Reyes et al., 1998a).

L. tarentolae also contains two mitochondrial complexes that may be involved in editing (Peris et al., 1994). The T class (T for TUTase containing complex) of RNPs sediments at 10S while the G class (G for major gRNA containing complexes) is found at approximately 20S. The 10S complex contains gRNAs, TUTase and RNA ligase. The 20S complex only contains gRNAs and a gRNA independent U insertion activity. It is

difficult to make any correlation between the complexes of the two species, as the *L. tarentolae* complexes have not been shown to contain gRNA-dependent editing activity.

Identified proteins. To identify proteins that might be involved in editing, crosslinking experiments have been used to isolate proteins that interact directly with gRNAs (Köller et al., 1994; Read et al., 1994). These studies yielded three proteins that appeared to interact with gRNAs with high affinity (9, 25 and 90 kDa). The 25 kDa protein was later identified as gBP21, the gRNA binding protein. As demonstrated in knockout experiments, gBP21 is not required for editing. However, it is tightly associated with the editosome (Lambert et al., 1999) as monoclonal antibodies to the protein immunoprecipitate editing activity (Allen et al., 1998). The 90 kDa protein was shown to specifically bind the U-tail, however, the 9 and 90 kDa proteins have not been further characterized. Similar experiments in *C. fasciculata* yielded three proteins (30, 65, 88 kDa) (Leegwater et al., 1995). All three proteins showed binding affinity to the U-tail. A protein believed to be the homologue of the 88 kDa polypeptide has been cloned in *T. brucei*. This protein, called TBRGG1, contains an RGG repeat, a known RNA binding motif (Vanhamme et al., 1998). TBRGG1 potentially associates with the editosome as it co-fractionates with RNA editing activity.

Another U binding protein, RBP16, was discovered by Hayman and Read (1999). This protein contains a RNP1 and a RGG-like RNA-binding motifs in the N- and the C-termini, respectively. A role for this protein in editing has yet to be established.

Using an array of monoclonal antibodies raised against protein components of the 35-40S editing complexes, a novel protein was identified (Madison-Antenucci et al.,

1998). The RNA editing associated protein (REAP-1) is a 45 kDa protein that has a 21 amino acid motif repeated eight times. The presence of positively charged amino acids within this motif suggests that REAP-1 may be an RNA-binding protein. In addition, monoclonal antibodies to REAP-1 inhibit *in vitro* editing activity, potentially indicating a role in editing.

Editosome assembly. The process by which mitochondrial proteins, mRNAs and gRNAs interact to form a complete editosome remains unclear. mRNAs and gRNAs do not contain common primary sequences that could act as RNA-binding motifs. gRNAs do contain U-tails which are bound by proteins, although the U-tail is unlikely to be the major recognition domain as mitochondrial rRNAs also have poly U tracts at their 3' ends (Adler et al., 1991). Despite the lack of an apparent sequence motif, gRNAs and mRNAs independently support the formation of protein complexes (Göringer et al., 1994; Koslowsky et al., 1996). Gel shift assays reveal that protein assembly on both RNAs produce four RNPs (G1-G4 and M1-M3 and M5 for gRNA and mRNA dependent complexes, respectively). Assembly of these RNPs was specific, as unrelated RNAs were not able to inhibit complex formation (Göringer et al., 1994; Koslowsky et al., 1996). These RNPs may represent editing complexes in various stages of assembly although this has not been established. In addition, the relationship between the mRNA and gRNA RNPs has not been investigated.

Investigation of the conditions under which complex formation is inhibited provided insight into what the protein complex was recognizing on the RNAs. Complex assembly on the gRNAs was affected by ionic strength and was disrupted by heparin

(Göringer et al., 1994; Koslowsky et al., 1996). The ability of heparin to prevent complex formation indicated that the polyanion was able to block electrostatic interactions between the gRNA and proteins. In the case of mRNA complexes, increasing K^+ and Mg^{2+} concentrations inhibited assembly (Koslowsky et al., 1996). Both K^+ and Mg^{2+} have been shown to play a role in RNA structure suggesting that higher levels of cations stabilized structures inappropriate for complex formation. These observations support a model in which the motif recognized by the protein complex involves negative charges on the gRNA and or mRNA phosphate backbone arranged by the structure of the RNAs. Therefore the motif can be masked by blocking electrostatic interactions or by the formation of improper structures. This model also explains the non-sequence specific binding of the mRNA and gRNA by the editosome in agreement with absence of a common sequence element within the RNAs.

An Overview of this thesis

In this introduction I have summarized the present state of kinetoplastid mitochondrial RNA editing research. While significant progress has been made, many questions still remain unanswered. Currently, most labs are focused on identifying the protein components of the editosome. With both *T. brucei* and *L. tarentolae* genome sequencing projects well underway and the battery of monoclonal antibodies generated against the editing complex, the components of the editosome will be characterized in the near future.

A natural progression of this work will be to examine how the editing complex is put together and how it assembles on gRNAs and mRNAs. Therefore the identification

of gRNA/mRNA features that could interact with these proteins would significantly contribute to our understanding of editosome assembly. With this in mind, our lab has approached the question of editosome assembly from the aspect of the RNA. RNAs are necessarily key players in the process of editing. It's the mRNA's coding sequence which is edited and the gRNA that provides the information for this process.

However, gRNA/mRNA pairs do not share any common sequences that could form RNA-binding motifs for proteins involved in editing. This has led to the overall hypothesis that interacting gRNAs and mRNAs form a core structure whose features are recognized by the editosome, enabling the editing complex to interact with the RNAs. In this thesis I examine the hypothesis that gRNA/mRNA pairs form a common structure, a crucial component of our overall hypothesis. In the following chapters I describe experiments that support my hypothesis. My initial experiments involved crosslinking the 5' and 3' ends of gRNAs to their cognate mRNAs. The positions of these crosslinks were mapped and incorporated into computer generated secondary structure predictions that suggest that interacting gRNA/mRNAs do form a common structure, with three structural elements: a gRNA/mRNA duplex, a U-tail/mRNA duplex and a gRNA stem-loop (Chapter 1). These crosslinking studies led to my interest in the U-tail/mRNA interaction as editing proceeds (Chapter 2). Using the technique of crosslinking again, secondary structure predictions were developed with partially edited mRNAs. The results of these experiments indicate that the U-tail cannot continue to act as an anchor or tether as editing proceeds. Instead I propose that the U-tail acts to maintain the gRNA stem-loop predicted in my initial experiments. This suggests that this structure is an important feature of gRNA/mRNA pairs, potentially playing a role in protein-RNA

interactions. Chapter 3 describes the use of standard probing techniques to obtain enzymatic and chemical probing data that support the predicted structure of gRNA/mRNA pairs from Chapter 1. In the last chapter I briefly summarize my results and describe directions this project might take to progress towards a more complete understanding of the assembly of active editing complexes on interacting gRNAs and mRNAs.

References

Adler B. K., Hajduk S. L. 1997. Guide RNA requirement for editing-site-specific endonucleolytic cleavage of preedited mRNA by mitochondrial ribonucleoprotein particles in Trypanosoma brucei. Mol. Cell. Biol. **17**(9):5377-85.

Adler B. K., Harris M. E., Bertrand K. I., Hajduk S. L. 1991. Modification of Trypanosoma brucei mitochondrial rRNA by posttranscriptional 3' polyuridine tail formation. Mol. Cell. Biol. **11**(12):5878-84.

Alfonzo J. D., Blanc V., Estevez A. M., Rubio M. A., Simpson L. 1999. C to U editing of the anticodon of imported mitochondrial tRNA(Trp) allows decoding of the UGA stop codon in Leishmania tarentolae. EMBO J. **18**(24):7056-62.

Allen T. E., Heidmann S., Reed R., Myler P. J., Göringer H. U., Stuart K. D. 1998. Association of guide RNA binding protein gBP21 with active RNA editing complexes in Trypanosoma brucei. Mol. Cell. Biol. **18**(10):6014-22.

Bakalara N., Simpson A. M., Simpson L. 1989. The Leishmania kinetoplast-mitochondrion contains terminal uridylyltransferase and RNA ligase activities. J. Biol. Chem. **264**(31):18679-86.

Benne R., Van den Burg J., Brakenhoff J. P., Sloof P., Van Boom J. H., Tromp M. C. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell **46**(6):819-26.

Bienen E. J., Saric M., Pollakis G., Grady R. W., Clarkson A. B., Jr. 1991. Mitochondrial development in Trypanosoma brucei brucei transitional bloodstream forms. Mol. Biochem. Parasitol. **45**(2):185-92.

Blum B., Bakalara N., Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell **60**(2):189-98.

Blum B., Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell **62**(2):391-7.

Blum B., Sturm N. R., Simpson A. M., Simpson L. 1991. Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. Cell **65**(4):543-50.

Burgess M. L., Heidmann S., Stuart K. 1999. Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit in vitro U insertion. RNA **5**(7):883-92.

Byrne E. M., Connell G. J., Simpson L. 1996. Guide RNA-directed uridine insertion RNA editing in vitro. EMBO J. **15**(23):6758-65.

Cech T. R. 1991. RNA editing: world's smallest introns? Cell **64**(4):667-9.

Clarkson A. B., Jr., Bienen E. J., Pollakis G., Grady R. W. 1989. Respiration of bloodstream forms of the parasite Trypanosoma brucei brucei is dependent on a plant-like alternative oxidase. J. Biol. Chem. **264**(30):17770-6.

Corell R. A., Read L. K., Riley G. R., Nellissery J. K., Allen T. E., Kable M. L., Wachal M. D., Seiwert S. D., Myler P. J., Stuart K. D. 1996. Complexes from Trypanosoma brucei that exhibit deletion editing and other editing-associated properties. Mol. Cell. Biol. **16**(4):1410-8.

Cruz-Reyes J., Rusché L. N., Piller K. J., Sollner-Webb B. 1998a. T. brucei RNA editing: adenosine nucleotides inversely affect U-deletion and U-insertion reactions at mRNA cleavage. Mol Cell **1**(3):401-9.

Cruz-Reyes J., Rusché L. N., Sollner-Webb B. 1998b. Trypanosoma brucei U insertion and U deletion activities co-purify with an enzymatic editing complex but are differentially optimized. Nucleic Acids Res. **26**(16):3634-9.

Cruz-Reyes J., Sollner-Webb B. 1996. Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. Proc. Natl. Acad. Sci. USA **93**(17):8901-6.

Decker C. J., Sollner-Webb B. 1990. RNA editing involves indiscriminate U changes throughout precisely defined editing domains. Cell **61**(6):1001-11.

Estevez A. M., Simpson L. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. Gene **240**(2):247-60.

Göringer H. U., Koslowsky D. J., Morales T. H., Stuart K. 1994. The formation of mitochondrial ribonucleoprotein complexes involving guide RNA molecules in Trypanosoma brucei. Proc. Natl. Acad. Sci. USA **91**(5):1776-80.

Hajduk S. L., Sabitini R. S. 1998. Mitochondrial mRNA Editing in Kinetoplastid Protozoa. In: Grosjean H., Benne R., editors. Modification and Editing of RNA. Washington, D.C.: ASM Press. p 377-393.

Hayman M. L., Read L. K. 1999. Trypanosoma brucei RBP16 is a mitochondrial Y-box family protein with guide RNA binding activity. J. Biol. Chem. **274**(17):12067-74.

Hermann T., Schmid B., Heumann H., Göringer H. U. 1997. A three-dimensional working model for a guide RNA from Trypanosoma brucei. Nucleic Acids Res. **25**(12):2311-8.

Horvath A., Berry E. A., Maslov D. A. 2000. Translation of the edited mRNA for cytochrome b in trypanosome mitochondria. Science **287**(5458):1639-40.

Igo R. P., Jr., Palazzo S. S., Burgess M. L., Panigrahi A. K., Stuart K. 2000. Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing [In Process Citation]. Mol. Cell. Biol. **20**(22):8447-57.

Kapushoc S. T., Simpson L. 1999. In vitro uridine insertion RNA editing mediated by cis-acting guide RNAs. RNA **5**(5):656-69.

Köller J., Norskau G., Paul A. S., Stuart K., Göringer H. U. 1994. Different Trypanosoma brucei guide RNA molecules associate with an identical complement of mitochondrial proteins in vitro. Nucleic Acids Res. **22**(11):1988-95.

Koslowsky D. J., Bhat G. J., Read L. K., Stuart K. 1991. Cycles of progressive realignment of gRNA with mRNA in RNA editing. Cell **67**(3):537-46.

Koslowsky D. J., Kutas S. M., Stuart K. 1996. Distinct differences in the requirements for ribonucleoprotein complex formation on differentially regulated pre-edited mRNAs in Trypanosoma brucei. Mol. Biochem. Parasitol. **80**(1):1-14.

Lambert L., Müller U. F., Souza A. E., Göringer H. U. 1999. The involvement of gRNA-binding protein gBP21 in RNA editing-an in vitro and in vivo analysis. Nucleic Acids Res. **27**(6):1429-36.

Leegwater P., Speijer D., Benne R. 1995. Identification by UV cross-linking of oligo(U)-binding proteins in mitochondria of the insect trypanosomatid Crithidia fasciculata. Eur. J. Biochem. **227**(3):780-6.

Madison-Antenucci S., Sabatini R. S., Pollard V. W., Hajduk S. L. 1998. Kinetoplastid RNA-editing-associated protein 1 (REAP-1): a novel editing complex protein with repetitive domains. EMBO J. **17**(21):6368-76.

Maslov D. A., Simpson L. 1992. The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing. Cell **70**(3):459-67.

Missel A., Norskau G., Shu H. H., Göringer H. U. 1995. A putative RNA helicase of the DEAD box family from Trypanosoma brucei. Mol. Biochem. Parasitol. **75**(1):123-6.

Missel A., Souza A. E., Norskau G., Göringer H. U. 1997. Disruption of a gene encoding a novel mitochondrial DEAD-box protein in Trypanosoma brucei affects edited mRNAs. Mol. Cell. Biol. **17**(9):4895-903.

Peris M., Frech G. C., Simpson A. M., Bringaud F., Byrne E., Bakker A., Simpson L. 1994. Characterization of two classes of ribonucleoprotein complexes possibly involved in RNA editing from Leishmania tarentolae mitochondria. EMBO J. **13**(7):1664-72.

Piller K. J., Rusché L. N., Cruz-Reyes J., Sollner-Webb B. 1997. Resolution of the RNA editing gRNA-directed endonuclease from two other endonucleases of Trypanosoma brucei mitochondria. RNA **3**(3):279-90.

Pollard V. W., Harris M. E., Hajduk S. L. 1992. Native mRNA editing complexes from Trypanosoma brucei mitochondria. EMBO J. **11**(12):4429-38.

Read L. K., Göringer H. U., Stuart K. 1994. Assembly of mitochondrial ribonucleoprotein complexes involves specific guide RNA (gRNA)-binding proteins and gRNA domains but does not require preedited mRNA. Mol. Cell. Biol. **14**(4):2629-39.

Riley G. R., Myler P. J., Stuart K. 1995. Quantitation of RNA editing substrates, products and potential intermediates: implications for developmental regulation. Nucleic Acids Res. **23**(4):708-12.

Rusché L. N., Piller K. J., Sollner-Webb B. 1995. Guide RNA-mRNA chimeras, which are potential RNA editing intermediates, are formed by endonuclease and RNA ligase in a trypanosome mitochondrial extract. Mol. Cell. Biol. **15**(6):2933-41.

Schmid B., Riley G. R., Stuart K., Göringer H. U. 1995. The secondary structure of guide RNA molecules from Trypanosoma brucei. Nucleic Acids Res. **23**(16):3093-102.

Seiwert S. D., Heidmann S., Stuart K. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. Cell **84**(6):831-41.

Seiwert S. D., Stuart K. 1994. RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro. Science **266**(5182):114-7.

Stuart K., Allen T. E., Heidmann S., Seiwert S. D. 1997. RNA editing in kinetoplastid protozoa. Microbiol Mol Biol Rev **61**(1):105-20.

Sturm N. R., Simpson L. 1990. Partially edited mRNAs for cytochrome b and subunit III of cytochrome oxidase from Leishmania tarentolae mitochondria: RNA editing intermediates. Cell **61**(5):871-8.

Vanhamme L., Perez-Morga D., Marchal C., Speijer D., Lambert L., Geuskens M., Alexandre S., Ismaili N., Göringer U., Benne R. and others. 1998. Trypanosoma brucei TBRGG1, a mitochondrial oligo(U)-binding protein that co-localizes with an in vitro RNA editing activity. J. Biol. Chem. **273**(34):21825-33.

CHAPTER 1

MAPPING CONTACTS BETWEEN gRNAs AND mRNAs IN TRYPANOSOME RNA EDITING

The results of this chapter have been published in the article: Leung, S.S. and Koslowsky, D.J. 1999. Mapping contacts between gRNAs and mRNAs in trypanosome RNA editing. Nucleic Acids Res. 27(3):778-87.

INTRODUCTION

In kinetoplastids, mitochondrial RNA editing involves the precise post-transcriptional insertion and deletion of uridylate residues (U) from mitochondrial pre-mRNA molecules (Alfonzo et al., 1997; Estevez and Simpson, 1999; Hajduk and Sabitini, 1998; Stuart et al., 1997). The information for this phenomenon is contained within a small RNAs (55-70 nts) known as guide RNAs (gRNAs). Analyses of gRNAs reveals that they can broken down into three functional elements. At the 5' end of gRNAs is an anchor sequence which enables the gRNA to basepair to the correct sequence on its cognate mRNA. The information sequence which guides editing is found immediately downstream of the anchor sequence. The 3' end of the gRNA is post-transcriptionally modified by the addition of approximately 15 U's, creating a U-tail. A survey of the primary sequence of gRNAs and mRNAs reveals a lack of common sequence motifs suggesting that the structure of the two interacting RNAs could play a role in the assembly of an active editing complex on the RNAs. Our approach to this hypothesis was limited by the dearth of information concerning the role of the U-tail in the editing process. While the role of the 5' anchor and information sequence elements of the gRNA have been demonstrated using *in vitro* assays, the function of the U-tail is less clear, with several possibilities proposed by different models of editing (Adler and Hajduk, 1997; Byrne et al., 1996; Kable et al., 1996; Seiwert et al., 1996).

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. (Blum et al., 1990; Blum and Simpson, 1990). 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 (Kable et al., 1996; Seiwert et al., 1996). The chimeric intermediate models suggested that the gRNA oligo(U) tail served as the U donor or acceptor during the editing process (Blum et al., 1991; Cech, 1991; Sollner-Webb, 1991). With the elimination of chimeras as editing intermediates, we are left with the question: what is the role of the U-tail? 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 (Blum and Simpson, 1990). In *in vitro* editing studies, removal of the gRNA U-tail does not diminish gRNA-directed mRNA cleavage (Seiwert et al., 1996). 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, the interaction of the U-tail of three different gRNAs with their pre-mRNAs was mapped. 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. The results of these investigations confirm the role of the anchor in correctly positioning the gRNA and provide 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, supporting our hypothesis. In all cases, the anchor duplex region is correctly paired and secondary structure in the immediate editing region is 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 to increase the stability of the gRNA/mRNA interaction. In addition, the basepairing of the gRNA to the mRNA at sequences flanking the initial editing sites, removes 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

Templates for *in vitro* transcription

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 oligonucleotides and cloning into pBluescriptII-SK- (Stratagene). Plasmids for gND7-506 and gA6-14 were gifts from Dr Ulrich Göringer (Schmid et al., 1995). The template for gCYb-558 (Riley et al., 1994) 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 nts
BIG SK	5'-GGCCGCTCTAGAACTAGTGG-3'	20 nts
ND75'NEdc	5'-CGGGTACCATGACTACATGATAAGTAC-3'	27 nts
TbHR3	5'-CTTTTATATTCACATACTTTTCTGTACC-3'	27 nts
MODHR3	5'-CCGGATCCATGGACGAACTACAAACACGATGCAA AT-3'	36 nts
gND7-506end	5'-AAAAAAAAAATTCACTATATACAC-3'	24 nts
gCYb-558-1	5'-CCTAGAAATTCACATTGTCTTTTAATCCCTATAGT GAGTCG-3'	41 nts
gCYb-1end	5'-AAAAAAAAAATTCCCTTTATCACCTAGAAATTCA C- 3'	35 nts
gCYb-558endU5	5'-AAAAATTCCCTTTATCACCTAGAAATTCAC-3'	30 nts
gA6-14end	5'-AAAAAAAAAATAATTATCATATC-3'	23 nts
C-gA6-14	5'-CAGGAATTCCGATAACGAATCAGATTTTGAC-3'	31 nts
A6H-1	5'-CCTAACCTTTCCTGC-3'	15 nts
T7leadercomp	5'-GGTACCCAATTCGCC-3'	15 nts

***In vitro* transcription**

T7 RNA polymerase (200 U) *in vitro* transcription reactions (40 mM Tris-HCl, pH 8.0, 19 mM MgCl2, 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 [[alpha]-32P]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 (Burgin and Pace, 1990). Transcripts were gel purified on an 8% polyacrylamide (w/v)-7 M urea gel.

Attachment of photoaffinity agents

Following Burgin and Pace (Burgin and Pace, 1990), 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 (Oh and Pace, 1994).

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 (Pollard et al., 1992; Seiwert et al., 1996). Each 0.5 ml fraction was then tested for activity using the deletion assay (Seiwert and Stuart, 1994). Hybridizations were carried out under RNA editing conditions (Seiwert et al., 1996). 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 cpm) 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/cm2 using a TE77 SemiPhor electroblotter. Cleavage products were identified using northern hybridization with mRNA- and gRNA-specific probes as indicated in Figure 1-4.

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 (Zuker, 1994). 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 (Matzura and Wennborg, 1996) 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.

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 (Figure 1-1).

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 (Bhat et al., 1990). 3'A6UT, covers 99 nts 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

Figure 1-1. 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.

A. 3′A6UT+gA6-14

```
5′gggcgaauuggguaccGGAAUUGCCUUUGCCAAACUUUUAGAAGAAAG
AGCAGGAAAGGUUAGGGGGAGGAGAGAAGAAAGGGAAAGUUGUGAUUUUGGAGUUAUAGAAUAAGAUCggaucc
                                                  |:|||||||||| |||   |
3′UUUUUUUUUUAUUAAUAGUAUAGUGACAGUUUUAGACUAAGCAAUAGCCUCAAUAUCAUAUAGUUAAGCGGG5′

acuaguucuagagcggcc3′
```

B. 5′ND7UMT+gND7-506

```
5′gggcgaauuggguaccCAUGACUACAUGAUAAGUACAAGAGGAGACAGACGACAGUGUCCACAGCACCCGUUUCAGC
ACAGUUGGAGGAGAGGGGAUAAGAUUUAUUGAUGAAAUUUGUGAUUUGCAUCGUGUUUGUAGUUCGUCCAUGGAUCCGG
                                                |||||:||:|||:|||:||:|||||:||
3′UUUUUUUUUUAAGUGAUAUAUGUGAUUGAUAUGAUGUCCAAUAAAUGUAGCAUAAAUAUUAAGCGGG5′
auccacuaguucuagagcggcc3′
```

C. 5′CYbUT+gCYb-558

```
5′gggcgaauuggguaccGUUAAGAAUAA
UGGUUAUAAAUUUUAUAUAAAAGCGGAGAAAAAAGAAAGGGUCUUUUAAUGUCAGGUUGUUUAUAUAGAAUAUAUGG
                                                  |||||||||| :|
3′UUUUUUUUUUAAGGGAAAUAGUGGAUCUUUAAGUGUAACAGAAAAUUAGGG5′
auccacuaguucuagagcggcc3′
```

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 (Koslowsky et al., 1990). 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 (Figure 1-1B).

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 nts of the 5'-end of CYb (Feagin et al., 1987). 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 (Feagin et al., 1987). 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 ((Bhat et al., 1990; Koslowsky et al., 1990; Riley et al., 1994).

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 (Seiwert et al., 1996), however, its interaction with mRNA has not been shown directly. Furthermore, this

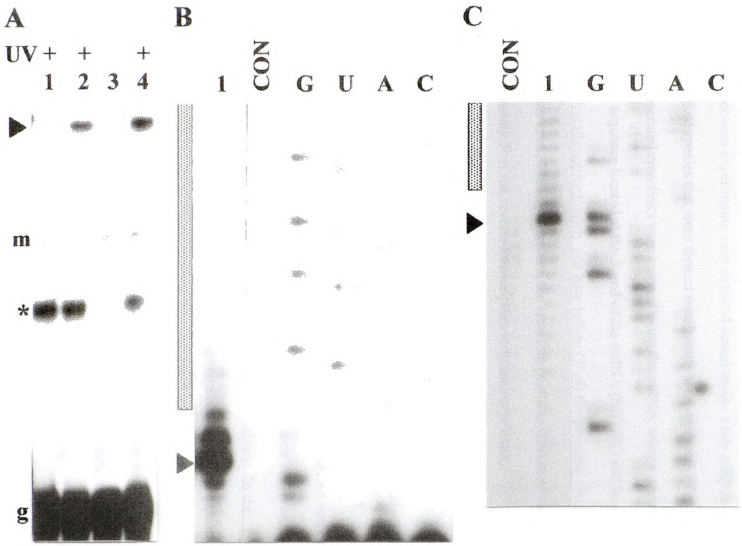
enabled us to confirm that the anchor duplex of the gRNA/mRNA pairs used in this study formed correctly, despite the presence of vector sequence in the mRNAs.

To label the 5'-end of the anchors of gND7-506 and gCYb-558, the gRNAs were synthesized in the presence of GMPS (Burgin and Pace, 1990; Sampson and Uhlenbeck, 1988). 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 (Figure 1-1C). 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 gRNA (Figure 1-1B). 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 (Seiwert et al., 1996) 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 (Figure 1-2A 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

Figure 1-2. 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 UV315. 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 32P-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.


A
UV + + +
1 2 3 4
m
*
g
B
1 CON G U A C
C
CON 1 G U A C

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 (Burgin and Pace, 1990; Denman et al., 1988). Therefore, the crosslink positions discussed will refer to the base immediately 5' of the reverse transcription termination product.

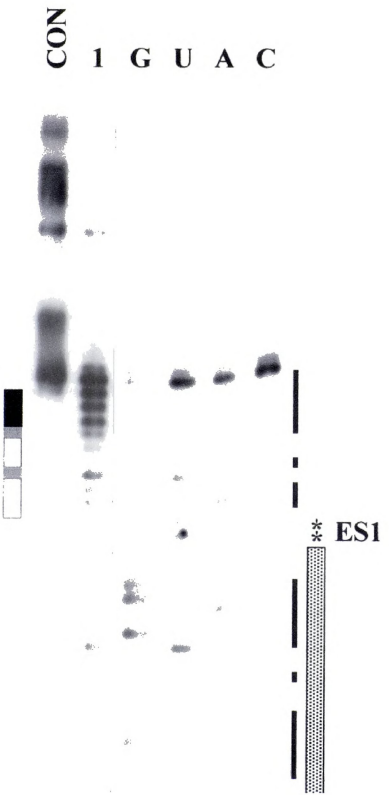
5' modified gND7-506 produced a primary crosslink (strongest termination product) that mapped to one base 3' of the predicted anchor duplex (Figure 1-2B). Two additional termination products were also observed corresponding to the first and second bases of the mRNA anchor. Reverse transcription of crosslinked 5'CYbUT and 5' modified gCYb-558 mapped a primary crosslink to two bases 3' of the expected anchor duplex (Figure 1-2C). 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 U10 tails using the protocol of Oh and Pace (Oh and Pace, 1994). *In vivo*, gRNAs have U-tails which average 15 U residues in length (Blum and Simpson, 1990). A U10 tail length was chosen for this study as we felt that a U10 tail would interact in a similar fashion as the *in vivo* U15 tail and because the U10 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 (Figure 1-3). A minor termination product was observed within the anchor duplex region located 5 nts from the 5'-end of the anchor duplex (mRNA orientation). Minor termination products were also observed corresponding to stops at residues 3-8 upstream

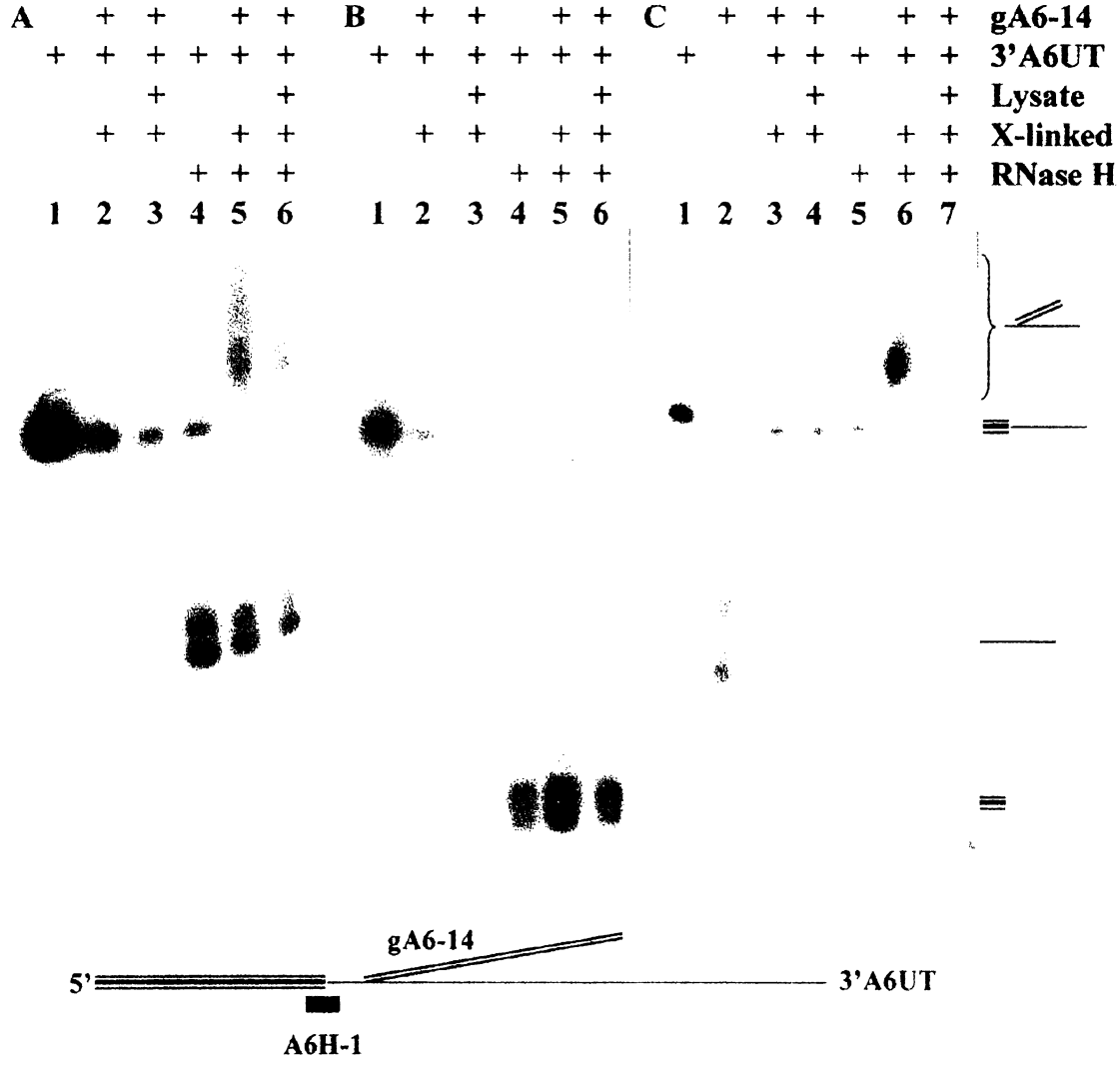
Figure 1-3. Mapping of 3'APA-gA6-14/mRNA intermolecular crosslinks by primer extension. CON, control lane, primer extension of non-crosslinked 3'A6UT. Strong termination products are observed in the purine-rich region located 13-35 nt upstream of the first editing site (ES1). Lane 1, primer extension 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.


CON
1 G U A C
** ES1

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 these data. However, the ladder of termination products found from 2 to 12 nts 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 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 (Figure 1-4). A6H-1 hybridizes ~30 nts 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 nts in length (Figure 1-4A and B, lanes 4), corresponding to the 3' and 5' halves of

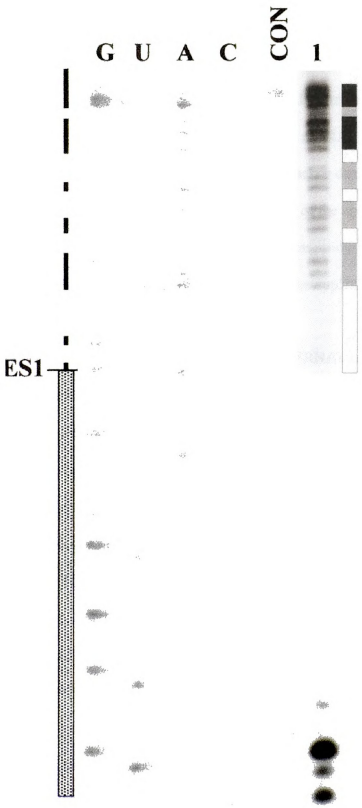
Figure 1-4. RNase H analysis of 3' modified gA6-14 and 3'A6UT conjugates. After RNase H digestion using A6H-1, products were resolved on an 8% denaturing gel and blotted onto Nytran. The three probes used were (A) BIG SK, complementary to the 3'-terminus of the mRNA, (B) T7leadercomp, complementary to 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).


A
B
C
gA6-14
3'A6UT
Lysate
X-linked
RNase H
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6 7
gA6-14
5'
3'A6UT
A6H-1

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 (Figure 1-4A and B, lane 2, and C, lane 3). In Figure 1-4A, 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 (Figure 1-4B) 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 (Figure 1-4C).

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 (Figure 1-5). Termination products not present in the control RT reactions were observed beginning just 5' of the anchor duplex and extending 28 nts further upstream. Two populations of dominant termination products, separated by a single base, were observed, corresponding to crosslinks 26-28 and 21-24

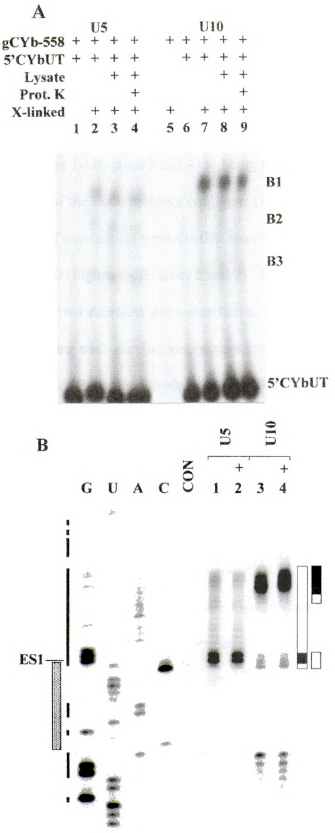
Figure 1-5. Mapping of 3'APA-gND7-506/5'ND7UMT intermolecular crosslinks by primer extension. G, U, A and C are RNA sequencing lanes used to map the position of the crosslinked gRNA. A primer extension reaction using non-crosslinked 5'ND7UMT is shown in the control (CON) lane. The sequence and control 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.


G
U
A
C
CON
1
ES1

nts upstream of the anchor. Distinct termination products were also observed at nts 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 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 nts 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; U10 and U5. In RNA interactions utilizing the 3'-end modified U10 gCYb, a single major and two minor mRNA-dependent crosslinked species were identified (Figure 1-6A, U10). 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 U5 tails. However, in reactions with U5 gCYb, the B2 and B3 crosslinked species were more pronounced (Figure 1-6A, U5). Analysis of the most abundant crosslinked species (B1) generated with U10 gCYb again produced a ladder of termination products beginning just 5' of the anchor duplex and extending ~17 nts upstream (Figure 1-6B, lanes 3 and 4). The strongest stops observed were 14-16 nts upstream of the anchor duplex. Closer to the anchor (3-13 bases) are minor crosslinks followed by three stronger

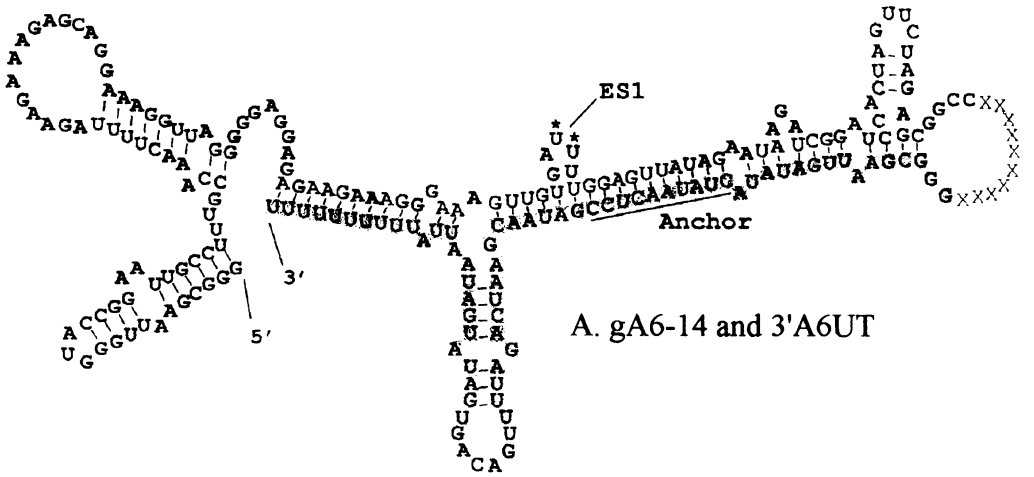
Figure 1-6. Identification and analysis of 3'APA-modified gCYb-558 crosslinked to 5'CYbUT. (A) Identification of intermolecular crosslinks using either gCYb-558U5 or gCYb-558U10. No crosslinks were obtained in the absence of UV treatment (lanes 1 and 6) or in the absence of mRNA (lane 5). Three crosslinked species (B1, B2 and B3) were obtained with gRNAs with both U5 and U10 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-558U5; lanes 3 and 4, termination products obtained when the mRNA is crosslinked to gCYb-558U10; 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.


A
U5
U10
gCYb-558 + + + + + + + + +
5'CYbUT + + + + + + + + +
Lysate + + + +
Prot. K +
X-linked + + + + + + + +
1 2 3 4 5 6 7 8 9
B1
B2
B3
5'CYbUT
B
U5
U10
G U A C CON 1 2 3 4
ES1

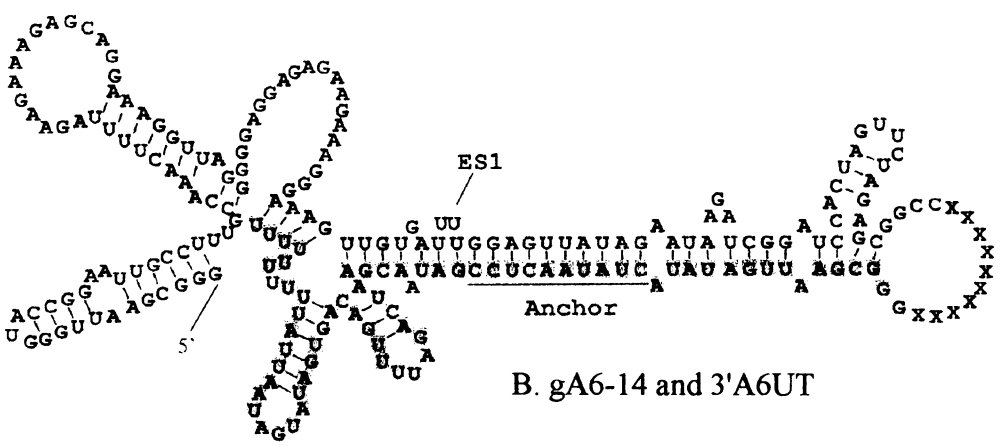
termination products corresponding to the last base of the anchor and the two bases just 5' of the anchor (mRNA orientation). In addition, termination products were also observed at the 3' boundary (mRNA orientation) of the anchor duplex. These termination products were not consistent in their appearance, however, and varied from being quite pronounced (Figure 1-6B, lanes 3 and 4) to being almost non-existent.

Analyses of the same mobility conjugate (B1) generated with the U5 gCYb showed a series of termination products that spanned the same nucleotides as those observed with U10 (Figure 1-6B, 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 U5 gCYb crosslinks. While in Figure 1-6B the intensity of the termination products in this anchor region were much more pronounced for the U10 gRNA substrates, in other experiments, no difference between the U5 and U10 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 U5 and U10 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 U5 and U10 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.

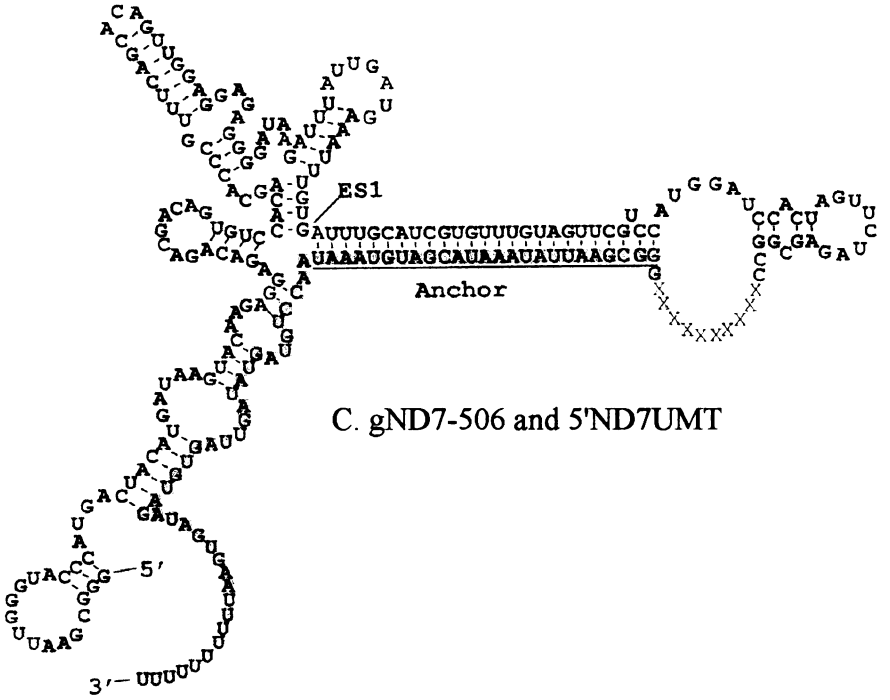
Figure 1-7. Comparison of the secondary structure predictions for the interactions of four gRNA/mRNA substrate pairs. The structures on top: A, C and E represent the initial secondary structure predictions generated with no constraints. The structures on the bottom: B, D, F and G were made with a forced base pair between the gRNA U10 nucleotide (U5 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 U10 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 U10 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-558U10/5'CYbUT interaction. In the initial prediction, the CYb message forms a very stable stem-loop structure, excluding the gRNA (E). After input of U10 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-558U5/5'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.



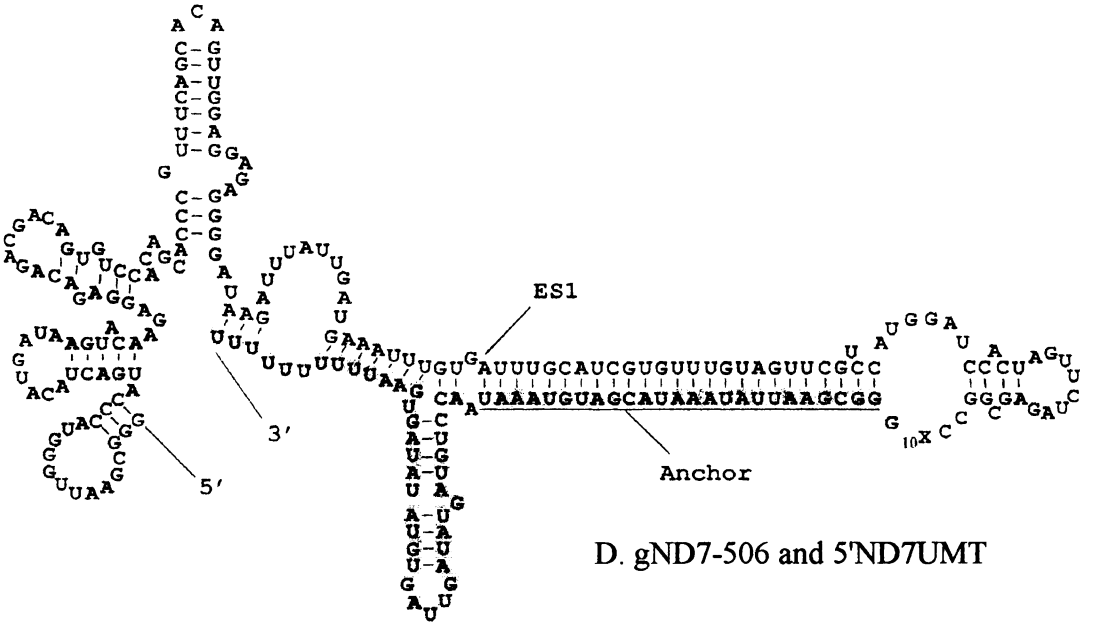
A. gA6-14 and 3'A6UT



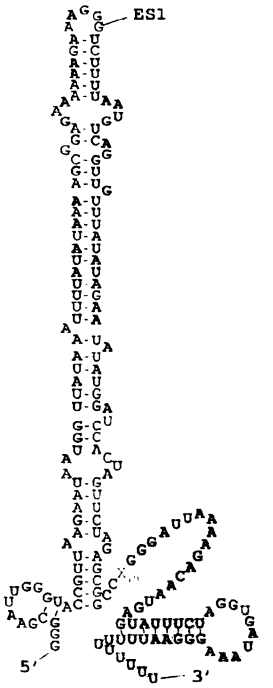
B. gA6-14 and 3'A6UT



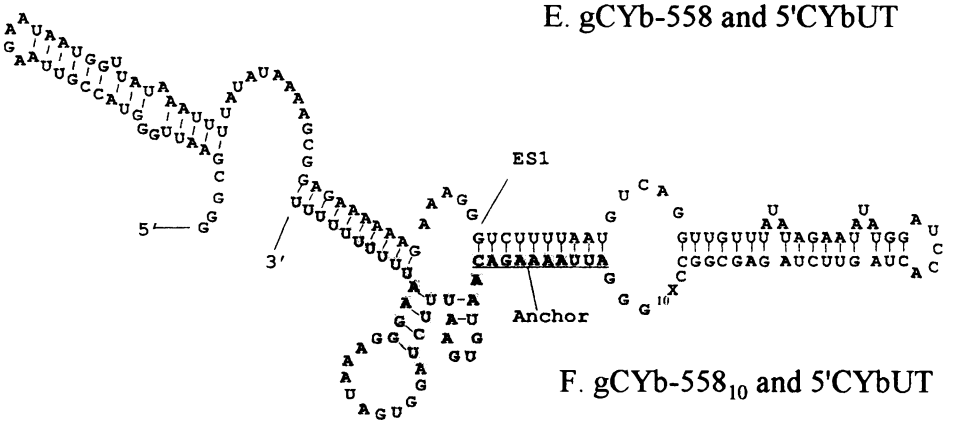
C. gND7-506 and 5'ND7UMT



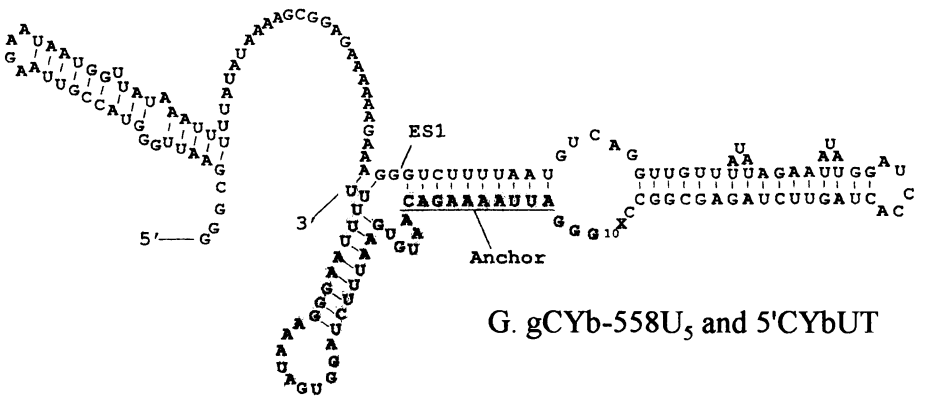
D. gND7-506 and 5'ND7UMT



E. gCYb-558 and 5'CYbUT



F. gCYb-558_{10} and 5'CYbUT



G. gCYb-$558U_5$ and 5'CYbUT

The RT analyses indicate that for all three gRNA/mRNA substrate pairs, the gRNA U10 tail interacts with the mRNA in a region just upstream of the anchor duplex (Figure 1-7). 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 nts 5' of the anchor duplex, respectively (Figure 1-7B and F). The terminal uridylate of gND7-506 crosslinked farther from the anchor duplex with the preferred sites at 21-28 nts upstream (Figure 1-7D). These crosslinking data were incorporated into the computer-predicted secondary structure models by instructing the program to pair the U10 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 (Figure 1-7). In all cases, the anchor duplex region is correctly paired and any secondary structure in the immediate editing domain is eliminated (compare Figure 1-7A and B for 3'A6U, Figure 1-7C and D for 5'ND7UMT and Figure 1-7E, 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

To begin to develop a structural model to support our hypothesis of a core gRNA/mRNA structure, we investigated gRNA/mRNA interactions using photoaffinity crosslinkers at the 5' and 3' ends of three gRNAs. 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 nts 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. Although within the sequence of 3'A6UT, there are virtually no pyrimidines within a 50 nt region beginning at 10 nts upstream of the anchor duplex, the dominant crosslink was mapped to nts 10-12. Interestingly, within the first 10 nts are 5 U's which the U-tail could not basepair with. 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 nts 10-12 indicating that the U-tail 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 nts. This region is extremely purine biased with a single pyrimidine found in the 22 nts directly 5' of the anchor duplex region. Using gCYb-558 with a U10 tail produced dominant crosslinks at nts 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 nts 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 (Figure 1-7D). 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 U10 tail, a single dominant conjugate (B1) was always observed, with two minor conjugates (B2 and B3) consistently appearing. The drop in tail length to U5 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 nts 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 only 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 U10 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 U5 to U15 (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 U5 tail also gave a ladder of termination products which corresponded with those observed with the U10 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 (Figure 1-7A, C and E). However, when the 3' crosslinking data was incorporated into the computer-predicted structures, the structures generated were all very similar (Figure 1-7B, 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 Figure 1-7C with D and Figure 1-7E 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.

Schmid et al. (Schmid et al., 1995) suggested that 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 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 (Blum and Simpson, 1992; Blum et al., 1991; Koslowsky et al., 1992; Read et al., 1992). 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.

The addition of active mitochondrial lysate did not affect the pattern of crosslinking. Similarly, in detailed studies of gND7-506 complexed with the gRNA

binding protein gBP21, Hermann et al. (Hermann et al., 1997) found that the protein binds to the guiding region stem-loop, 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.

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. These data indicate that the gRNA 5' anchor does position the gRNA by basepairing the mRNA just 3' of the editing domain. In addition the 3' crosslinking data provides the first direct evidence that the U-tail interacts with upstream purine rich sequences. This supports a role for the U-tail in both stabilization of the gRNA/mRNA interaction and tethering of the 5' cleavage product during editing. 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. At the same time, the 5' anchor and the U-tail duplex with the mRNA flanking the first few editing sites, possibly working together to remove mRNA secondary structure in the immediate editing domain providing the editing complex increased access to the correct editing sites.

ACKNOWLEDGEMENTS

M. Harris and N. Pace provided 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.

REFERENCES

Adler B. K., Hajduk S. L. 1997. Guide RNA requirement for editing-site-specific endonucleolytic cleavage of preedited mRNA by mitochondrial ribonucleoprotein particles in *Trypanosoma brucei*. Mol. Cell. Biol. **17**(9):5377-85.

Alfonzo J. D., Thiemann O., Simpson L. 1997. The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. Nucleic Acids Res. **25**(19):3751-9.

Bhat G. J., Koslowsky D. J., Feagin J. E., Smiley B. L., Stuart K. 1990. An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell **61**(5):885-94.

Blum B., Bakalara N., Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell **60**(2):189-98.

Blum B., Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell **62**(2):391-7.

Blum B., Simpson L. 1992. Formation of guide RNA/messenger RNA chimeric molecules in vitro, the initial step of RNA editing, is dependent on an anchor sequence. Proc. Natl. Acad. Sci. USA **89**(24):11944-8.

Blum B., Sturm N. R., Simpson A. M., Simpson L. 1991. Chimeric gRNA-mRNA molecules with oligo(U) tails covalently linked at sites of RNA editing suggest that U addition occurs by transesterification. Cell **65**(4):543-50.

Burgin A. B., Pace N. R. 1990. Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. EMBO J. **9**(12):4111-8.

Byrne E. M., Connell G. J., Simpson L. 1996. Guide RNA-directed uridine insertion RNA editing in vitro. EMBO J. **15**(23):6758-65.

Cech T. R. 1991. RNA editing: world's smallest introns? Cell **64**(4):667-9.

Denman R., Colgan J., Nurse K., Ofengand J. 1988. Crosslinking of the anticodon of P site bound tRNA to C-1400 of E.coli 16S RNA does not require the participation of the 50S subunit. Nucleic Acids Res. **16**(1):165-78.

Estevez A. M., Simpson L. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. Gene **240**(2):247-60.

Feagin J. E., Jasmer D. P., Stuart K. 1987. Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell **49**(3):337-45.

Hajduk S. L., Sabitini R. S. 1998. Mitochondrial mRNA Editing in Kinetoplastid Protozoa. In: Grosjean H., Benne R., editors. Modification and Editing of RNA. Washington, D.C.: ASM Press. p 377-393.

Hermann T., Schmid B., Heumann H., Göringer H. U. 1997. A three-dimensional working model for a guide RNA from *Trypanosoma brucei*. Nucleic Acids Res. **25**(12):2311-8.

Kable M. L., Seiwert S. D., Heidmann S., Stuart K. 1996. RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA [see comments] [published erratum appears in Science 1996 Oct 4;274(5284):21]. Science **273**(5279):1189-95.

Koslowsky D. J., Bhat G. J., Perrollaz A. L., Feagin J. E., Stuart K. 1990. The MURF3 gene of *T. brucei* contains multiple domains of extensive editing and is homologous to a subunit of NADH dehydrogenase. Cell **62**(5):901-11.

Koslowsky D. J., Göringer H. U., Morales T. H., Stuart K. 1992. In vitro guide RNA/mRNA chimaera formation in *Trypanosoma brucei* RNA editing [see comments]. Nature **356**(6372):807-9.

Matzura O., Wennborg A. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Comput. Appl. Biosci. **12**(3):247-9.

Oh B. K., Pace N. R. 1994. Interaction of the 3'-end of tRNA with ribonuclease P RNA. Nucleic Acids Res. **22**(20):4087-94.

Pollard V. W., Harris M. E., Hajduk S. L. 1992. Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. EMBO J. **11**(12):4429-38.

Read L. K., Corell R. A., Stuart K. 1992. Chimeric and truncated RNAs in *Trypanosoma brucei* suggest transesterifications at non-consecutive sites during RNA editing. Nucleic Acids Res. **20**(9):2341-7.

Riley G. R., Corell R. A., Stuart K. 1994. Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. J. Biol. Chem. **269**(8):6101-8.

Sampson J. R., Uhlenbeck O. C. 1988. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. Proc. Natl. Acad. Sci. USA **85**(4):1033-7.

Schmid B., Riley G. R., Stuart K., Göringer H. U. 1995. The secondary structure of guide RNA molecules from *Trypanosoma brucei*. Nucleic Acids Res. **23**(16):3093-102.

Seiwert S. D., Heidmann S., Stuart K. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. Cell **84**(6):831-41.

Seiwert S. D., Stuart K. 1994. RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro. Science **266**(5182):114-7.

Sollner-Webb B. 1991. RNA editing. Curr. Opin. Cell Biol. **3**(6):1056-61.

Stuart K., Allen T. E., Heidmann S., Seiwert S. D. 1997. RNA editing in kinetoplastid protozoa. Microbiol. Mol. Biol. Rev. **61**(1):105-20.

Zuker M. 1994. Prediction of RNA secondary structure by energy minimization. Methods Mol. Biol. **25**:267-94.

CHAPTER 2

CHARACTERIZATION OF gRNA U-TAIL INTERACTIONS WITH PARTIALLY EDITED mRNA SUBSTRATES

The bulk of this chapter has been published in the article: RNA editing in *Trypanosoma brucei*: Characterization of gRNA U-tail interactions with partially edited mRNA substrates. Leung, S.S. and Koslowsky, D.J. 2001. Nucleic Acids Res. 29(3):

INTRODUCTION

The process of insertion and deletion of uridylates from mitochondrial pre-mRNAs is directed by gRNAs. These small RNAs basepair with their cognate mRNAs via a 5' anchor sequence. Downstream of this sequence is the information that guides the process of editing. Completing the gRNA is a post-transcriptionally added U-tail of approximately 15 U's.

The activities required for editing are associated with the editosome that must interact with the gRNA/mRNA complex. This association does not appear to be sequence specific suggesting that structural elements of gRNA/mRNA pairs could play a role in RNA-protein interactions. This led to the initial investigation (Chapter 1) of the gRNA/mRNA structure by mapping the 5' and 3' ends of gRNAs along their cognate mRNAs (Leung and Koslowsky, 1999). In addition to providing information about the structure of gRNA/mRNA pairs, studying the interaction of the 3' end of the gRNA also provided insight into the role of the gRNA U-tail.

Several roles for the U-tail have been previously proposed by various models of editing. In the cleavage/ligation model for editing, it is hypothesized that the U-tail helps to stabilize the interaction of the gRNA and mRNA by binding to purine rich regions upstream of editing sites (Blum et al., 1990; Blum and Simpson, 1990). *In vitro* studies by Seiwert et al. (1996) demonstrated that while removal of the U-tail did not reduce gRNA directed cleavage of the mRNA the formation of the edited product was strongly suppressed. This led to the proposal that in addition to stability, the U-tail was involved in tethering the 5' cleavage product during the editing reaction. The initial crosslinking study of three different gRNA/mRNA pairs, described in Chapter 1, provided direct

evidence that the U-tail could basepair with upstream purine rich sequences (Leung and Koslowsky, 1999). The crosslinking data indicated that the U-tail interacted just 5-28 nts upstream of the anchor duplex. Although the crosslinking data identified a favored crosslinking site, the data did suggest that the U-tail was able to interact with a range of upstream sequences. Predicted gRNA/mRNA secondary structures incorporating these data were very similar. In addition to a U-tail/mRNA duplex, all structures contained a predicted anchor duplex while secondary structure in the mRNA editing domain was eliminated. In the gRNAs' guiding region a stem/loop was present across from the first few editing sites. These result suggested that the U-tail may act not only to increase the stability of the RNA 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.

Crosslinking of the 3' end of the U-tail with sequences near the first editing site indicated that the U-tail was interacting with mRNA regions that were to be subsequently edited. This raised the interesting question: what happens to the mRNA/U-tail interaction as editing proceeds in the 3' to 5' direction? Furthermore, what changes in the gRNA/mRNA structure would be predicted to occur? To examine these questions, gCYb-558 and its interaction with the following three apocytochrome b (CYb) mRNA substrates were examined: 5'CYbUT, which is unedited and two partially edited substrates (PES), 5'CYbPES1T and 5'CYbPES3T, which have ES1 through ES3 fully edited, respectively. The placement of an azidophenacyl (APA) group at the 3' end of gCYb-558 enabled the use of photoaffinity crosslinking techniques to study how the addition of U's by the editing process affected the positioning of the U-tail. Surprisingly,

reverse transcriptase (RT) analyses of the major crosslinked species for the three different CYb substrates consistently revealed strong termination products at the same five bases. This region is the same region previously identified as being involved in U-tail binding and is located only 4 - 8 nt upstream of the growing anchor in the most edited substrate. This is striking, as editing through ES3 requires the addition of 6 U's and essentially doubles the length of the gRNA/mRNA duplex yet the 3' end of the gRNA interacts with the same sequence. Using this crosslink data, secondary structure models suggest that a previously predicted gRNA stem-loop is maintained as editing proceeds through ES3. This is made possible by incorporating part of the U-tail into the stem-loop. The maintenance of the gRNA stem-loop emphasizes its potential importance and suggests that the U-tail may have the additional role of maintaining important secondary structure motifs required for interaction with the editing complex.

We have also shown that 3' crosslinked 5'CYbUT and gCYb-558 molecules are biologically relevant as they are recognized and specifically cleaved by the gRNA directed endonuclease at the correct editing site. This demonstrates the usefulness of these crosslinked molecules in the development of a model for gRNA and mRNA interactions in editing.

METHODS AND MATERIALS

Oligodeoxynucleotides

Big SK	5'GGCCGCTCTAGAACTAGTGG3'	20 nt
T7	5'AATTAATACGACTCACTATAG3'	22 nt
CYbCS BamHI	5'CCGGATCCATATATTCTATATAAACAACC3'	29 nt

CYbN	5'GGAGGTACCGTTAAGAATAATGGTTATAAATTT TATATAA3'	40 nt
CYbH-1	5'CAACCTGACATT3'	12 nt
CYbH-2	5'ACCATTATTCT3'	11 nt
CYbPES1	5'CTATATAAACAACCTGACATTAAAAGACAACC TTTCTTTTTTC3'	43 nt
CYBPES3	5'CTATATAAACAACCTGACATTAAAAGACAACA CAAATTTCTTTTTTC3'	47 nt
NgCYb-558(sU)	5'TTATTCCCTTTATCACCTAGAAATTCACATTGTC TTTTAATCCCTATAGTGAGTCGTATTAAATT3'	65 nt
NgCYb-558B	5'AAAAAAAAAAAAAAATTATTCCCTTTATCA3'	30 nt

DNA templates and RNA synthesis

5'CYbUT has been previously described (Koslowsky et al., 1996). Partially edited 5'CYb substrates were created using PCR. To synthesize 5'CYbPES1T and 5'CYbPES3T, the 5'CYbUT DNA template was amplified using T7 and CYbPES1 or CYbPES3 oligodeoxyribonucleotides, respectively. The PCR products obtained from this step were re-amplified using 5'CYbN and CYbCS oligonucleotides. These PCR products were subjected to BamHI and KpnI digestion and cloned into pBluescript SK - (Stratagene). Templates for transcription were obtained from the appropriate plasmids using T7 and Big SK oligonucleotides for PCR. 5'CYbUT and the partially edited RNAs were synthesized by T7 RNA polymerase using a Ribomax kit (Promega) according to manufacturer's directions. mRNAs were gel purified on 6% denaturing polyacrylamide gels. The RNAs were passively eluted in 10 mM Tris pH 7.8., 0.1% SDS, 2 mM EDTA and 0.3 M NaOAc, pH 7.0. gCYb-558 RNA was synthesized using NgCYb-558sU and T7 oligodeoxyribonucletides via the Uhlenbeck single-stranded T7 transcription method

(Milligan et al., 1987). The sequence for the oligodeoxyribonucleotide template for gCYb-558 was redesigned to more closely match the native gRNA sequence without a U-tail (Riley et al., 1994). To improve the homogeneity at the 3' end of the transcribed gRNA, transcription was carried out under low Mg^{2+} conditions (10 mM) (30 mM HEPES pH 8.0, 3 mM Spermidine, 10 mM DTT, and 5 mM KCl). A U_{15} tail was then added to the gRNA by ligation of a U_{15} RNA oligonucleotide (Dharmacon) using a bridging oligodeoxyribonucleotide (NgCYb-558B) and T4 DNA ligase (Moore and Sharp, 1992). Before ligation, the U_{15} RNA oligonucleotide was 5' end-labeled with ^{32}P-ATP. This involved drying down 1.8 nmol of U_{15} RNA oligonucleotide, 250 μCi of ATP γ-^{32}P and 5.25 nmol of cold ATP. The pellet was resuspended in 20 μl of 1X T4 DNA kinase buffer and 75 units of T4 DNA kinase (NEB). Kinase reactions were incubated at 37°C for 1-1.5 hours. To inactivate the kinase, the reaction was incubated at 65°C for 20 minutes. Equimolar amounts of NgCYb-558B (bridging oligonucleotide) and NgCYb-558sU (no U-tail) RNA were added and heated to 70°C for 2 minutes. The molecules were annealed by cooling to 37°C at a rate of 2°C/minute. The reaction conditions for the ligase reaction were as follows: 1X T4 DNA Ligase buffer, 35 U of T4 DNA Ligase (Boehringer Mannheim), 15% PEG 8000 and 120 units of RNasin (Promega). The ligation was incubated overnight at room temperature. The full length product was then gel purified on an 8% denaturing polyacrylamide gel and recovered.

RNA modifications

Photoagent attachment, crosslinking of gRNAs and mRNAs, and primer extension mapping were previously described (Leung and Koslowsky, 1999). The mRNA

to gRNA ratio was 10:1, using a gRNA concentration of 2.5-3.0 μM. Efficiency of crosslinking was quantitated using a Storm phosphorimager (Molecular Dynamics).

RNase H analysis

Each RNase H reaction contained a total of 10 pmol of RNA. If necessary, crosslinks (gCYb-558 was already radioactively labeled as above) and uniformly labeled mRNAs were supplemented with the appropriate cold mRNA to make a total of 10 pmol of RNA per reaction. 5' end-labeled crosslinks and uniformly labeled mRNAs were incubated with 20-30 pmols of each oligodeoxyribonucleotide in 120 mM HEPES pH 8.0, 210 mM NH_4Cl, 7 mM $MgCl_2$, 0.75 mM DTT and 5 U of RNase H (Takara). Reactions were incubated at 55°C for 30 minutes. Products were run out on a 6% denaturing gel and subject to autoradiography for analysis.

Cleavage reactions

The mRNA of crosslinks (0.25-0.5 pmol) were 3' end-labeled with 10 μCi of [5'-^{32}P]-pCp using T4 RNA ligase as previously described by Wahle and Keller (1994). Only the mRNA was end labeled in this process as the 3' end of the gRNA was crosslinked to the mRNA. Glycerol gradients were obtained as previously described (Pollard et al., 1992; Seiwert et al., 1996). RNAs were heated to 60°C and cooled slowly to 27°C at 2°C per minute before the addition of 10 ul of an active glycerol fraction. Reaction conditions were as follows: 20 mM HEPES pH 7.9, 50 mM KCl, 10 mM MgOAc, 0.05 mM DTT, 1 mM EDTA and 5 mM $CaCl_2$. 0.5-2 fmol of crosslink or 3.5-20 fmol of 5'CYbUT was used in cleavage reactions. 10-20 fold excess of gCYb-558 was used where specified.

Control lane conditions were identical except no mitochondrial proteins were added. A T1 ladder was created by incubating 3' end-labeled 5'CYbUT (60 kCPM, 1 fmol) in 12.8 M urea, 40 mM sodium citrate, pH 8.3, and 2 mM EDTA, pH 8.0 with 2 units of T1 ribonuclease (Boehringer Mannheim) for 2 min at 55°C. Reaction products were run out on an 8% denaturing polyacrylamide gel and exposed to film.

Secondary structure predictions

The program RNAstructure ver.3.5 (Mathews et al., 1999) was used to predict secondary structures. Crosslink data was used to force the last U to basepair with the appropriate base in the mRNA. RNAdraw was then used to graphically display the connect files (Matzura and Wennborg, 1996). Lowercase a's were used to create a single molecule for the RNAstructure program as previously described (Leung and Koslowsky, 1999)

RESULTS

CYb substrates and gCYb-558

The editing of apocytochrome b pre-mRNA is limited to a small region near its 5' end and is a developmentally regulated process. Editing inserts 34 U's over 13 sites only during the procyclic (insect) and stumpy bloodstream stages of the trypanosome life cycle (Feagin et al., 1987). Maturation of the first 7 editing sites is guided by gCYb-558 which directs the insertion of 21 U's. gCYb-558 is 59 nt long (including a U_{15} tail) and is able to interact with unedited CYb via an anchor of 13 nt with one mismatch. The unedited mRNA used in this study, 5'CYbUT, contains 88 nt of the 5' end of CYb and has been

previously described (Koslowsky et al., 1996; Leung and Koslowsky, 1999) The partially edited substrates, CYbPES1T and CYbPES3T, are identical in sequence to CYbUT except for the editing events at sites 1 - 3. CYbPES1T is edited at site one by the addition of two uridylates, extending the anchor duplex region by three basepairs. CYbPES3T is fully edited at sites 1 - 3, with a total of 6 uridylate insertions. These editing events extend the anchor duplex by 13 basepairs (Figure 2-1, A-C).

In our previous crosslinking study, we used gRNAs with a U_{10}-tail. The average length of a gRNA U-tail is approximately 15 nts (Blum and Simpson, 1990). However, we found that the T7 RNA polymerase used for transcription of the gRNAs stuttered extensively with a U_{15} template, generating considerable heterogeneity at the 3' ends of the gRNAs. In pursuing the question of how the U-tail interacts as editing proceeds, we made two improvements to our assay. 1.) The sequence of gCYb-558 was redesigned to more closely match that found *in vivo* (Figure 2-1, D, an additional 3 guiding nts) (Riley et al., 1994). 2) gCYb-558 is synthesized with T7 RNA polymerase using a template with no U-tail encoded. A U_{15} RNA oligonucleotide is then ligated to the 3' end of the gRNA using a deoxyoligonucleotide as a bridge and T4 DNA ligase (Moore and Sharp, 1992). This method produces a much more homogeneous population of gRNAs to use in our crosslinking studies.

gRNA U-tail interactions

We have previously analyzed gCYb-558/5'CYbUT crosslinks generated using gRNAs with two different U-tail lengths, U_{10} and U_5. Analysis of the most abundant crosslinked species generated with U_{10} gCYb produced a ladder of termination products

A. 5'CYbUT

```
5'...AUAUAAAAGCGGAGAAAAAAGAAAGGGUCUUUUAAUGUCAGGUUGUUUAUA...3'
                                  |||||||||| :|
                               3'...CAGAAAAUUAGGG...5'   gCYb-558
```

B. 5'CYbPES1T

```
5'...AUAUAAAAGCGGAGAAAAAAGAAAGGUUGUCUUUUAAUGUCAGGUUGUUUAUA...3'
                                    :||||||||||||| :|
                                3'...UAACAGAAAAUUAGGG...5'   gCYb-558
```

C. 5'CYbPES3T

```
5'...AUAUAAAAGCGGAGAAAAAAGAAAUUUGUGUUGUCUUUUAAUGUCAGGUUGUUUAUA...3'
                          |||||||::::|||||||||||||| :|
                     3'...UCUUUAAGUGUAACAGAAAAUUAGGG...5'   gCYb-558
```

D. gCYb-558

```
5'GGGAUUAAAAGACAAUGUGAAUUUCUAGGUGAUAAAGGGAAUAAUUUUUUUUUUUUUUU3'
```

Figure 2-1. Sequence of the 5'CYb mRNA substrates in the editing region (A-C) and full length gCYb-558 (D). Bold U's indicate U's added to create partially edited substrates. The gRNA/mRNA anchor is shown for each 5'CYb substrate with the gRNA aligned below the mRNA. The basepairing between the gRNA anchor and the mRNA is shown by Watson-Crick (|) and non-Watson-Crick basepairs (:). Changes to the sequence of gCYb-558 are underlined (see Methods and Materials).

beginning just 5' of the anchor duplex and extending approximately 17 nt upstream (Leung and Koslowsky, 1999). The strongest stops observed were 14 - 16 nt upstream of the anchor duplex. Analyses of the crosslinks generated with the U_5 gCYb showed a series of termination products that spanned the same nucleotides as those observed with U_{10}. However, the dominant termination products now correlated to crosslinks with the nucleotides that flank the first editing site. Crosslinking this close to ES1 indicated that the U-tail was interacting with mRNA regions that were to be subsequently edited by the interacting gRNA. This led us to investigate what happens to the mRNA/U-tail interaction as editing proceeds through this region.

Crosslinks were produced by annealing gCYb-558 (modified with a 3' APA group on the terminal U) to each of the different CYb substrates and exposing them to 312nm UV light. Crosslinked RNAs were then separated using denaturing PAGE. As previously observed for the gCYb-558 U_{10}/5'CYbUT crosslinks, a dominant band (B1, the species of slowest mobility) and 2 minor, faster bands (B2 and B3) were obtained for each mRNA/gCYb-558 combination used. 3' crosslinking efficiencies were measured for the major B1 bands on a phosphorimager. The efficiencies were very similar for the CYbUT and CYbPES1T substrates (~1%, average of four different crosslinking experiments). However, the efficiency of crosslinking for the CYbPES3T substrate was much lower (0.34%, average of three experiments).

The position of the minor B3 crosslink was mapped using RT and confirmed our previous results that this crosslink occurs in the 5' vector sequence of the mRNA. B2 did not produce any strong terminations (data not shown). Suprisingly, reverse transcriptase analyses of the dominant B1 species for all three different CYb substrates consistently

revealed strong termination products (highlighted by solid circles) at the same five bases (C_{51} G_{52} G_{53} A_{54} G_{55}) (Figure 2-2). These invariable termination products indicate that crosslinking occurs within G_{52} to A_{56}, the RT stops 1 nt 3' of a crosslink (Burgin and Pace, 1990). Termination products just upstream were observed inconsistently (Figure 2-2A). A strong stop was also often observed in the middle of the anchor duplex region of 5'CYbPES3T (Figure 2-2C). However, RNase H digestion (see below) did not indicate the presence of a crosslink. This termination product is more easily explained by the RT having difficulty reading through the long (26 nt) anchor duplex region. The above crosslinks fall within the same sequence that was previously identified using the slightly different gCYb-558 sequence and the shorter U_{10}-tail. What is remarkable is the fact that the U-tail is interacting only 4-8nts upstream of the growing anchor in 5'CYbPES3T.

Additional evidence for a physical crosslink in this region of the CYb substrates was provided by RNase H digestion. By using several oligonucleotides complementary to the mRNA, two oligonucleotides, CYbH-1 and CYbH-2, were found to flank the crosslink (Figure 2-3A). RNase H digestion with both oligonucleotides removed a total of approximately 60 nt from the CYb substrates. Treatment of the crosslink with CYbH-1, CYbH-2, or both oligonucleotides resulted in an RNA species with a faster mobility. However, the digestion products still ran slower than full length mRNA (Figure 2-3A, Lanes 5 - 8). This indicated that gCYb-558 was crosslinked to the mRNA, creating a branched molecule that runs with a slower mobility. This analysis narrowed the location of the crosslink to a region of 43 nt of 5'CYbUT (45 nt and 49 nt for 5'CYbPES1T and 5'CYbPES3T, respectively) spanning the location of the strong upstream RT termination sites. Complete digestion of the crosslinks was not obtained;

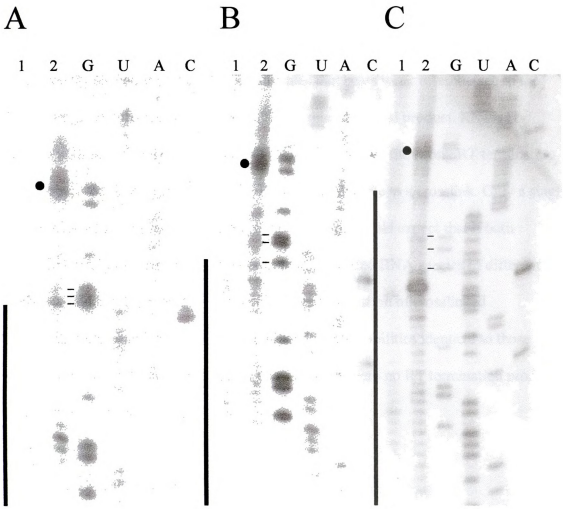


Figure 2-2. Primer extension analyses of 3' APA modified gCYb-558 crosslinked to three CYb mRNA substrates. A. 5'CYbUT, B. 5'CYbPES1T, and C. 5'CYbPES3T. Lane 1, RT of mRNA alone. Lane 2, RT of crosslink. G, U, A, C, represent sequencing lanes. The anchor duplex is highlighted by the thick black line. The major crosslink is identified by a solid circle. The three G's flanking the first 3 editing events are marked with bold horizontal lines.

hence the lighter, full-length crosslink bands (Figure 2-3A, lanes 6-8). In lanes 2 and 4, the short products (<~70nt) are not shown. CYbH-2 appeared to bind a second site weakly resulting in a second product (asterisk in Figure 2-3A). There is a short sequence in the 3' vector sequence of the mRNA substrates that bears weak resemblance to the CYbH-2 target site that could be responsible for this additional product. RNase H digestion with both CYbH-1 and CYbH-2 also indicated that the strong RT termination product observed in the anchor of 5'CYbPES3T was not due to a crosslink. Only a single RNA species was observed after RNase treatment. One would expect that if both termination products were the result of crosslinks, a second RNA species of different mobility would be observed. Furthermore, RNase H digestion of crosslinked 5'CYbPES3T generated the same pattern of bands with mobilities identical to those observed for 5'CYbUT and 5'CYbPES1T crosslinks, where no RT termination product was observed in the anchor.

Cleavage of Crosslinked Substrates

In order to demonstrate that these crosslinks represented biologically relevant molecules, we wanted to determine whether they were substrates for the RNA editing machinery. Direct visualization of editing was not possible due to the branched structure of the crosslinked RNA (Seiwert et al., 1996). In addition, detection of editing using the poison primer assay was hindered by the presence of the crosslinked gRNA and its ability to bind to the mRNA via its anchor sequence (Seiwert and Stuart, 1994). Therefore, we examined whether these molecules could be accurately cleaved by the gRNA-directed

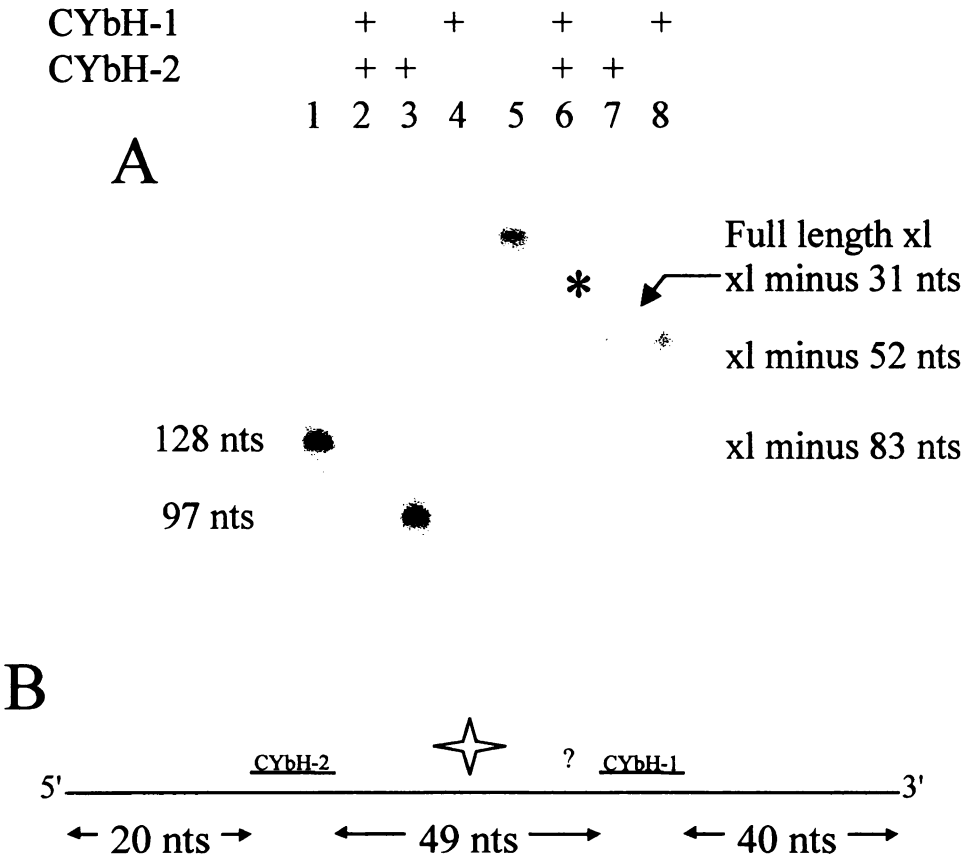


Figure 2-3. RNaseH mapping of crosslinked 5'CYbPES1T and gCYb-558. Lanes 1-4, mRNA only. Lanes 5-8, crosslinked RNA. All lanes were treated with RNase H. Sizes of 5'CYbPES1T mRNA fragments are shown on the left. Descriptions of the 5'CYbPES1T crosslink fragments appear on the right. CYbH-2 appears to weakly bind to a second site, resulting in an additional product (*). (B) Cartoon of the position of CYbH-1 and CYbH-2. The size of uncrosslinked mRNA is shown on the bottom. The position of the crosslink is depicted by the star. ? indicates the position of the RT termination product seen in 5'CYbPES3T.

endonuclease previously identified in mitochondrial fractions (Adler and Hajduk, 1997; Piller et al., 1997).

For these assays, 3' crosslinked molecules were generated using ^{32}P-trace labeled gRNAs and unlabeled mRNAs. The 3' end of purified crosslinked mRNAs were then end-labeled to a high specific activity using T4 RNA ligase and [5'-^{32}P]-pCp and again gel-purified. 3' end-labeled crosslinks (30kcpm, 0.5-2 fmols) were then incubated in an editing reaction (Cruz-Reyes et al., 1998a; Seiwert et al., 1996). UTP and ATP were not included in the reaction mix in order to inhibit ligation and enhance the production of the cleavage product. The 3'-crosslinked molecules were accurately cleaved at ES1 in the presence of active mitochondrial fractions (Figure 2-4). Crosslinked 5'CYbUT yielded a 3' cleavage product that is 59nt in length as expected for cleavage at ES1 (where 2 U's are inserted). This cleavage is consistent with cleavage by the editing endonuclease and not the two other mitochondrial endonucleases previously identified(Piller et al., 1997). A crosslinked species that ran just below the full length crosslink was observed very weakly in the control lane, and enhanced in the presence of mitochondrial proteins. This might indicate the presence of a hypersensitive cleavage site in the crosslinked RNA that is susceptible to cleavage by mitochondrial RNases. The location of this cleavage site was not determined. However, if cleavage was in the mRNA, the cleavage site had to be upstream of the crosslink, due to its mobility. It is also possible that cleavage of the gRNA could also be responsible for this RNA species. An additional product with a mobility of ~91 nt is also observed in the presence of active lysate. It has not been determined whether this product results from cleavage of free mRNA from broken crosslinks or crosslinked RNA. Cleavage of free mRNA would indicate that the cleavage

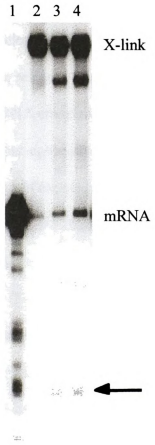


Figure 2-4. Accurate cleavage of 3' crosslinked 5'CYbUT and gCYb-558. Crosslinks (mRNA 3' end-labeled) were assayed for gRNA directed cleavage using standard cleavage conditions. Lane 1, T1 digest of 5'CYbUT. Lanes 2-4, 3' crosslinked RNAs. Lane 2, no Mt fraction. Lane 3, Plus Mt fraction. Lane 4, Mt fraction with 10 fold excess free gCYb-558. 3' crosslinks, free 5'CYbUT from broken crosslinks and cleavage products are indicated by: x-link, mRNA and an arrow, respectively.

site is upstream of the CYb editing domain within the 5'UTR. If the product is a crosslinked molecule, the cleavage site cannot be determined without additional experiments.

Isolation and subsequent handling of the crosslinked substrates always results in crosslink breakage and subsequent release of free mRNA (Figure 2-4 lanes 2, 3 and 4). Therefore we considered whether the released free mRNA could generate the cleavage products. Breakage of the mRNA/gRNA crosslinks would release the two RNAs in equimolar amounts. In the *in vitro* cleavage assays described to date, generation of cleavage product requires the addition of excess gRNA in order to drive the reaction (Adler and Hajduk, 1997; Cruz-Reyes et al., 1998b; Piller et al., 1997). No gRNA directed cleavage of free mRNA is detected when utilizing a 1:1 mRNA:gRNA ratio (data not shown). Efficient cleavage is only observed when the gRNA is supplied in excess. In contrast, cleavage of the crosslinked substrates was relatively efficient in the absence of any added free gRNA (Figure 2-4, lane 3) and the addition of free gRNA to the reaction did not increase the efficiency of cleavage (Figure 2-4, lane 4). These data indicate that the 3' crosslinked substrates support accurate gRNA directed cleavage, suggesting that these crosslinked substrates have been captured in a biologically active state.

Predicted Secondary Structures

To understand how the U-tail could interact with the same sequence in all three cases, we incorporated the cross-link data into computer predicted secondary structures using the computer program, RNAstructure version 3.5 (Figure 2-5) (Mathews et al.,

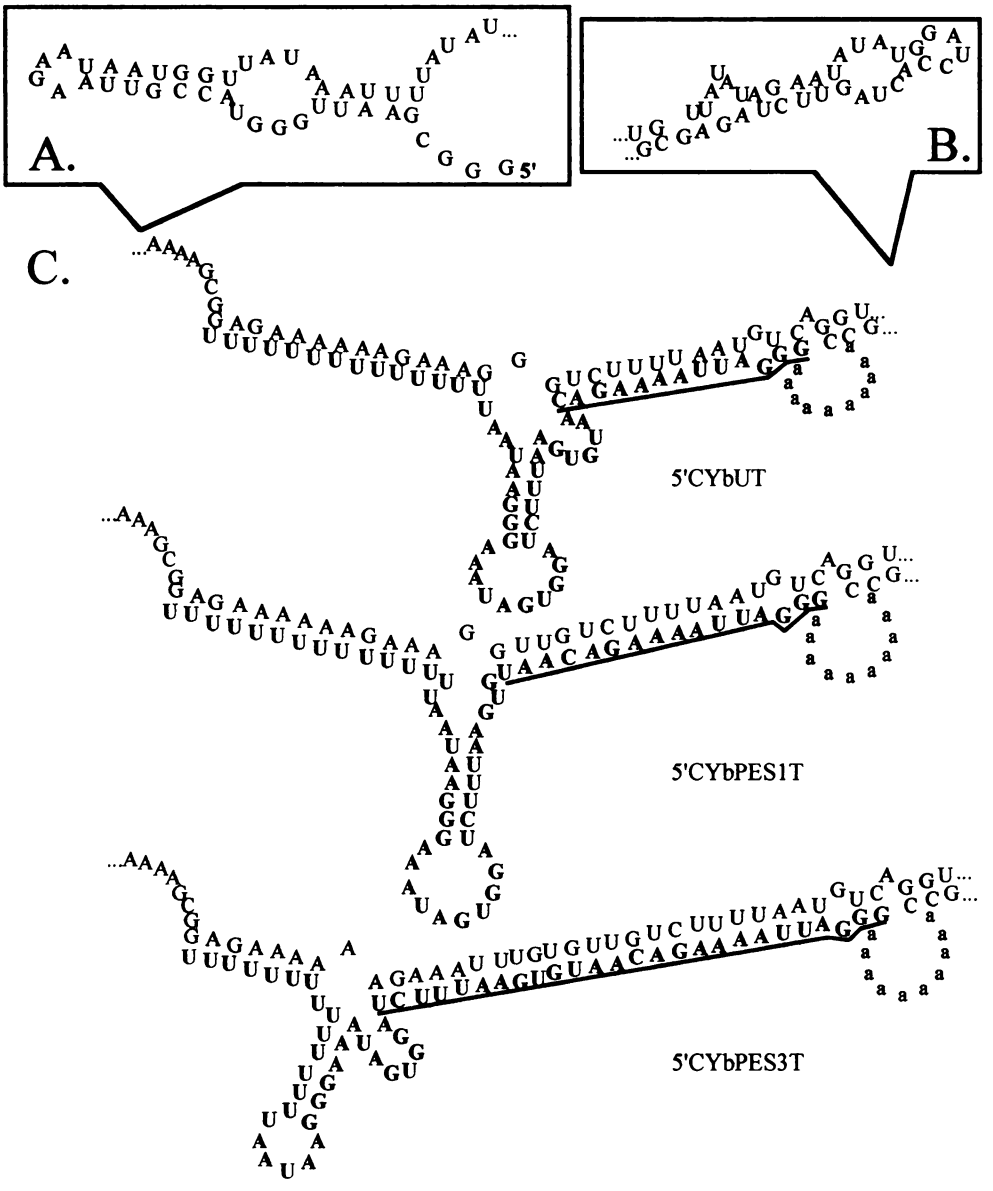


Figure 2-5. Secondary structure predictions incorporating 3' crosslink data. In all three gRNA/mRNA pairs, the 5' and 3' ends of the mRNA produced identical folds, as summarized in panels A and B. The predicted folds for the different gRNA/mRNA pairs highlighting the interaction with the gRNA are shown in C. The gRNA sequence is on the bottom and in bold. The gRNA/mRNA anchors are underlined. The gRNA and mRNA were linked together via a linker of 10 non-pairing bases (a's).

1999). This was done by instructing the program to pair the 3' terminal uridylate of gCYb-558 with G_{53} of the mRNA, one of the dominant crosslinked nucleotides. The structures predicted were very similar to one another (Figure 2-5). The anchor duplex regions are correctly paired and the previously described stem-loop structure formed within the guiding region of the gRNAs (Blum and Simpson, 1990; Leung and Koslowsky, 1999; Schmid et al., 1995) is maintained in all three folds. The predicted structure for gCYb-558 and 5'CYbPES3T (Figure 2-5, C) is particularly interesting, as part of the U-tail is involved in maintaining the stem-loop structure. This shortens the length of the predicted U-tail/mRNA interaction from 13-14 basepairs (in CYbPES1T and CYbUT), to only 7 basepairs for 5'CYbPES3T. One would predict that this would weaken the interaction of the U-tail with 5'CYbPES3T and may explain the decrease in U-tail crosslinking efficiency observed with the CYb PES3T substrate.

DISCUSSION

In an effort to understand how gRNAs and mRNAs interact with the editosome we have begun to develop a structural model of gRNA/mRNA pairs. Both mRNAs and gRNAs necessarily contain different sequences coding for various mitochondrial proteins, leading to the hypothesis that the information required for the assembly and interaction of the editosome on gRNA/mRNA pairs resides in the structure of the RNAs. Previously in Chapter 1, crosslinking techniques were used to characterize the interaction of the gRNA's U-tail with unedited mRNAs sequences in three different gRNA/mRNA pairs. In these studies, we found that the U-tail did interact with upstream purine rich sequences, preferring purines located near the first few editing sites. Interaction of the U-

tail in this region prevented the formation of any mRNA secondary structures in the immediate editing domain possibly providing the editing machinery access to these sites. In addition, the structure predictions for the three gRNA/mRNA pairs contained three structural elements: (1) a gRNA/mRNA anchor duplex, (2) a U-tail/mRNA duplex and a gRNA stem-loop. These computer predictions indicated that the gRNA/mRNA pairs, despite having very different primary sequences, could form similar secondary structures suggesting the formation of a common core architecture which may be important in the assembly of a functional editing complex.

In this current study, we have investigated the interaction of gCYb-558's U-tail with partially edited 5'CYb substrates in order to determine how the change in sequence associated with the editing process might affect the U-tail/mRNA interaction. Our previous studies indicated that the U-tail was interacting with mRNA regions that were to be subsequently edited. We wanted to determine how the increase in the anchor duplex region due to editing, might affect this interaction. One possibility was that the interaction with the mRNA is flexible, with the U-tail "sliding" up the mRNA as editing progressed. Alternatively, the U-tail may move 5' along the mRNA in a stepwise fashion. Once editing proceeds beyond a specific threshold, the U-tail would interact at a new position, farther upstream. This threshold could be defined by the maintenance of a predicted gRNA stem-loop across from the current editing site. Editing could exceed this threshold by creating a large enough anchor duplex, employing nucleotides previously involved in the gRNA stem-loop, resulting in the destabilization of the stem-loop and the interaction of the U-tail.

In this study, we demonstrate that despite editing progressing up to and including ES3 (an increase of 13 basepairs within the anchor duplex region), the U-tail continues to interact with the same purine rich sequence (G_{52} to A_{56}) observed with the unedited mRNA. This crosslink data indicates that the U-tail does not "slide" up the mRNA as editing proceeds. The computer predicted structures generated using this crosslink data, indicate that as the anchor duplex extends, the stem-loop structure in the gRNA can be maintained by alternate basepairs which include the U-tail. This suggests that the role of the U-tail may change as editing proceeds. During the initial stages, its role may be to provide stability during the initial gRNA/mRNA interaction and to help tether the 5' cleavage product. However, as editing proceeds and the size of the gRNA/mRNA anchor increases, the U-tail's contribution to stability is less important. Furthermore, with the 3' most end of the U-tail continuing to interact with the same sequence, the number of U's interacting with the 5' cleavage product decreases, thereby reducing the U-tail's ability to hold on to the 5' cleavage product. This function maybe taken over by protein:RNA interactions as suggested by Burgess et al. (1999), and Kapushoc and Simpson (1999) . Instead of these initial roles, it may be that the U-tail functions to maintain important secondary structure motifs (such as the gRNA stem-loop) by feeding into the structure as editing progresses. This stem-loop could potentially function as a protein binding site (Blum and Simpson, 1990; Hermann et al., 1997; Leung and Koslowsky, 1999; Schmid et al., 1995). Alternatively, as base stacking is a major factor in the stabilization of RNA structures, it may be that the presence of multiple helices that can stack may be important for editing complex stability.

These functions may help to explain why the gRNA has a U-tail. Uridines are able to interact with the upstream purine rich sequences found in pre-mRNA, allowing the U-tail to help stabilize the gRNA/mRNA interaction and bind to the 5' cleavage product. Progression of editing diminishes this requirement and instead the U-tail is needed to maintain the gRNA stem-loop by basepairing with the remainder of the gRNA's guiding region. We hypothesize because uridines are able to bind both A and G, a series of uridines may be the best universal sequence able to carry out the above functions.

ACKNOWLEDGEMENTS

The synthesis and cloning of the partially edited CYb mRNA substrates was done with Catherine Book's assistance.

REFERENCES

Adler B. K., Hajduk S. L. 1997. Guide RNA requirement for editing-site-specific endonucleolytic cleavage of preedited mRNA by mitochondrial ribonucleoprotein particles in *Trypanosoma brucei*. Mol. Cell. Biol. **17**(9):5377-85.

Blum B., Bakalara N., Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell **60**(2):189-98.

Blum B., Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell **62**(2):391-7.

Burgess M. L., Heidmann S., Stuart K. 1999. Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit in vitro U insertion. RNA **5**(7):883-92.

Cruz-Reyes J., Rusché L. N., Piller K. J., Sollner-Webb B. 1998a. *T. brucei* RNA editing: adenosine nucleotides inversely affect U-deletion and U-insertion reactions at mRNA cleavage. Mol. Cell. **1**(3):401-9.

Cruz-Reyes J., Rusché L. N., Sollner-Webb B. 1998b. *Trypanosoma brucei* U insertion and U deletion activities co-purify with an enzymatic editing complex but are differentially optimized. Nucleic Acids Res. **26**(16):3634-9.

Feagin J. E., Jasmer D. P., Stuart K. 1987. Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell **49**(3):337-45.

Hermann T., Schmid B., Heumann H., Göringer H. U. 1997. A three-dimensional working model for a guide RNA from *Trypanosoma brucei*. Nucleic Acids Res. **25**(12):2311-8.

Kapushoc S. T., Simpson L. 1999. In vitro uridine insertion RNA editing mediated by cis-acting guide RNAs. RNA **5**(5):656-69.

Koslowsky D. J., Kutas S. M., Stuart K. 1996. Distinct differences in the requirements for ribonucleoprotein complex formation on differentially regulated pre-edited mRNAs in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **80**(1):1-14.

Leung S. S., Koslowsky D. J. 1999. Mapping contacts between gRNA and mRNA in trypanosome RNA editing. Nucleic Acids Res. **27**(3):778-87.

Mathews D. H., Sabina J., Zuker M., Turner D. H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288**(5):911-40.

Matzura O., Wennborg A. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Comput. Appl. Biosci **12**(3):247-9.

Milligan J. F., Groebe D. R., Witherell G. W., Uhlenbeck O. C. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res. **15**(21):8783-98.

Moore M. J., Sharp P. A. 1992. Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. Science **256**(5059):992-7.

Piller K. J., Rusché L. N., Cruz-Reyes J., Sollner-Webb B. 1997. Resolution of the RNA editing gRNA-directed endonuclease from two other endonucleases of *Trypanosoma brucei* mitochondria. RNA **3**(3):279-90.

Pollard V. W., Harris M. E., Hajduk S. L. 1992. Native mRNA editing complexes from *Trypanosoma brucei* mitochondria. EMBO J. **11**(12):4429-38.

Riley G. R., Corell R. A., Stuart K. 1994. Multiple guide RNAs for identical editing of *Trypanosoma brucei* apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. J. Biol. Chem. **269**(8):6101-8.

Schmid B., Riley G. R., Stuart K., Göringer H. U. 1995. The secondary structure of guide RNA molecules from *Trypanosoma brucei*. Nucleic Acids Res. **23**(16):3093-102.

Seiwert S. D., Heidmann S., Stuart K. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. Cell **84**(6):831-41.

Seiwert S. D., Stuart K. 1994. RNA editing: transfer of genetic information from gRNA to precursor mRNA in vitro. Science **266**(5182):114-7.

Wahle E., Keller W. 1994. 3' end-processing of mRNA. In: Higgins S. J., Hames B. D., editors. RNA Processing. A Practical Approach. New York: IRL Press. p 1-33.

CHAPTER 3

STRUCTURE PROBING OF AN mRNA BOUND TO ITS COGNATE gRNA

INTRODUCTION

Mitochondrial RNA editing in *Trypanosoma brucei* is a unique phenomenon in which uridylates are precisely inserted or less frequently, deleted from mRNA (Estevez and Simpson, 1999; Hajduk and Sabitini, 1998; Stuart et al., 1997). Editing is essential, as this mitochondrial post-transcriptional process is required to produce translatable transcripts. Short RNAs (50-70 nts), termed guide RNAs (gRNAs), contain the information for the sequence modifications. To direct these events, gRNAs must basepair with their cognate mRNAs. This interaction is achieved by the basepairing of an anchor sequence, found at the 5' end of the gRNA, with a complementary sequence in the mRNA. This creates an anchor duplex and serves to correctly position the gRNA along the mRNA to direct the nucleotide changes. In addition to the anchor, gRNAs contain a guiding region and a uridine tail (U-tail). The guiding region acts as a pseudotemplate providing the information for the precise insertion and deletion of uridylates. The approximately 15 nt 3' U-tail is post-transcriptionally added and has been hypothesized to act as an additional anchor, increasing the stability of the interacting RNAs and tethering the 5' mRNA cleavage product during editing (Blum and Simpson, 1990; Seiwert et al., 1996).

The current model for RNA editing invokes a cleavage-ligation mechanism (Estevez and Simpson, 1999; Hajduk and Sabitini, 1998; Stuart et al., 1997). The initial step is gRNA-directed cleavage of the mRNA at the first mismatch between mRNA and gRNA, immediately 5' to the anchor duplex. The 3' cleavage product is basepaired to the gRNA's anchor sequence while the 5' cleavage fragment is hypothesized to be tethered by

the U-tail's interaction with upstream purine rich sequences. A precise number of U's is then added or deleted from the 5' cleavage fragment by a terminal uridylyl transferase (TUTase) or an exonuclease, respectively. The 5' and 3' mRNA fragments are then joined together via a ligase. The enzymatic activities described above are found to associate in a large complex termed the editosome (Corell et al., 1996; Pollard et al., 1992; Rusché et al., 1997). This editosome must have the ability to recognize and bind hundreds of different gRNA/mRNA pairs. However, gRNAs and mRNAs do not contain common sequence motifs that could be used for protein-RNA recognition. Instead, we hypothesize, that interacting gRNAs and mRNAs form a common, higher ordered structure that is recognized by the editosome (Leung and Koslowsky, 1999). Similarily, Schmid et al. (Schmid et al., 1995), using structure probing techniques, have shown that gRNAs form very similar secondary structures which may explain why different gRNAs can interact with the same set of mitochondrial proteins (Köller et al., 1994).

We are interested in whether interacting gRNAs and mRNAs form a common core structure that could be involved in assembly of the editing complex. To examine this question the 5' and 3' ends of three gRNAs were mapped along their cognate mRNAs using photoaffinity crosslinking (Leung and Koslowsky, 1999; Leung and Koslowsky, 2001). The crosslinking data confirmed that the gRNA anchor correctly positions the RNA and that the 3' end of the U-tail interacts with upstream purine sequences, 5-28 nts upstream of the initial editing site. Using this crosslinking data, similar secondary structure predictions were obtained for the three different gRNA/mRNA pairs. Each predicted structure contained a gRNA/mRNA anchor duplex, a U-tail/mRNA duplex and

a gRNA stem-loop similar to the one previously observed in the guiding sequence of gRNAs alone (Schmid et al., 1995).

We describe here, additional support for our predicted structures, using chemical and enzymatic probing techniques. To overcome the technical difficulties associated with the solution probing of two interacting RNAs, we utilized a gRNA/mRNA pair that had been photochemically crosslinked at the 3' most end (mRNA orientation) of their anchor duplex. A 5' crosslink at this site would insure that the anchor duplex would be maintained during probing. In addition, because this crosslink is at the beginning of a known region of basepaired RNA, we anticipated it to be a minor constraint, limiting perturbation of the structure of the interacting RNAs.

We report here that these 5' crosslinked molecules support accurate gRNA directed endoribonuclease activity as well as TUTase, and exonuclease activity, strongly suggesting that these molecules are recognized by the editosome. The structure probing data obtained directly shows that the U_{15}-tail protects several mRNA bases predicted to be involved in the U-tail mRNA duplex. In combination with our previous crosslinking studies, this new data provides additional support for the predicted secondary structure of interacting gRNA/mRNA pairs (Leung and Koslowsky, 1999; Leung and Koslowsky, 2001).

METHODS AND MATERIALS

Templates for transcription

5'CYbUT has been described before (Koslowsky et al., 1996). The A6 partially edited mRNA substrates were created from 3'A6UT using PCR with oligonucleotides T7 and

3'A6PES1 or 3'A6PES3 followed by a second round of amplification using T7 and A6 MOD Anchor (Koslowsky et al., 1996).

Transcription

mRNAs were synthesized in the presence of 5mM guanosine to enable 5' end-labeling. The nucleoside can only be incorporated at the 5' end of transcripts. mRNA transcription reactions also contained 40 mM Tris pH 8.0, 15 mM DTT, 2 mM spermidine, 0.01% Triton X-100, 1 mM GTP and the remaining dNTPs at 4 mM. gRNAs were synthesized using the appropriate oligonucleotide (NgA6-14 or NgCYb-558) and the T7 oligonucleotide via the Uhlenbeck single-stranded T7 transcription method (Milligan et al., 1987). Guanosine 5'-thiophosphate (GMPS) containing gRNAs were transcribed in the presence of 7 mM GMPS and 1mM dNTPs. mRNAs and gRNAs were gel purified on 6% and 8% denaturing polyacrylamide gels, repectively. RNAs were passively eluted from gels in 10 mM Tris pH 7.8., 0.1% SDS, 2 mM EDTA and 0.3 M NaOAc, pH 7.0 and recovered by precipitation.

Crosslinking

Modifications to the gRNAs have been described previously (Leung and Koslowsky, 1999). RNAs were annealed using 2 to 5 molar amounts of mRNA to modified gRNA. Crosslinks were purified on 6% denaturing gels and recovered as described above.

End-labeling

Up to 5 pmols of guanosine-incorporated free or crosslinked mRNA were 5' end-labeled using 50 μCi of γ-^{32}P ATP, 10 units of Kinase (New England Biolabs) and 1X Kinase Buffer for 1 hour at 37°C. Unincorporated γ-^{32}P ATP was removed by gel purifying the RNAs on a 6% denaturing gel.

Structure probing

Single strand-specific probes. All reactions contained 30 Kcpm of 5' end-labeled substrate and 5 μg of yeast tRNA. Substrates were denatured at 60°C for 2 minutes and cooled to 27°C (optimal temperature for procyclic trypanosome growth) at a rate of 2°C per minute. After 20 minutes at 27°C, the enzymatic or chemical probe was added. Reactions were incubated for 10 minutes at 27°C and terminated by phenol extraction. Products were run out on 8% denaturing polyacrylamide gels. The reaction conditions for the individual probes are described below. RNase T1 (Roche) digestions were carried out in 10mM Tris pH 7.5, 50mM KCl and 1mM $MgCl_2$. Probing conditions for RNase T2 (Sigma) were as follows: 10mM Tris pH 7, 50mM KCl and 1mM $MgCl_2$. Mung Bean Nuclease (NEB) reactions contained 30 mM NaOAc pH 7, 50mM NaCl, 5% glycerol and 0.1mM $MgCl_2$. 1mM $ZnCl_2$ was added with the nuclease. Reducing the magnesium concentration from 1mM $MgCl_2$ to 0.1mM $MgCl_2$ only increased the intensities of cleavages observed and did not change the overall pattern obtained (data not shown). Only data from probing reactions that resulted in significantly less than 50% cleavage of the input RNAs was used for interpretation as suggested by (Christiansen et al., 1990).

Probing with MPE-Fe(II). 80uM of MPE-Fe(II) was freshly prepared for each reaction from methidiumpropyl-EDTA (Sigma) and ferrous ammonium sulfate (Sigma). 200 Kcpm of 5' end-labeled RNA and 8 μg of yeast tRNA were heated to 60°C for 2 minutes and cooled to 27°C at 2°C/minute in 10mM Tris pH 7.5, 50mM KCl and 1mM $MgCl_2$. Before the addition of probe a time zero aliquot was taken. MPE-Fe(II) was added to 20uM followed by DTT to 6.25mM. Aliquots were taken at 2 minute intervals up to 6 minutes. Reactions were stopped with 10M urea, 20mM EDTA, 100ug tRNA, 0.1% bromophenol blue and 0.1% xylene cyanol. Aliquots were stored at -80°C until run out on 12% denaturing polyacrylamide gels.

Structure prediction

Structures predictions were obtained using RNAstructure 3.5 and graphically displayed using RNAdraw version 1.1 as previously described (Leung and Koslowsky, 1999; Mathews et al., 1999; Matzura and Wennborg, 1996).

Cleavage reactions

Glycerol gradients of mitochondrial lysates were prepared as previously described (Pollard et al., 1992; Seiwert et al., 1996). RNAs were heated to 60°C and slowly cooled to 27°C at a rate of 2°C per minute. 10 μl of an active glycerol fraction was then added. Reactions were incubated at 27°C for 20 minutes. Conditions of the reactions were as follows: 20 mM HEPES pH 7.9, 50 mM KCl, 10 mM MgOAc, 0.05 mM DTT, 1 mM EDTA and 5 mM $CaCl_2$. Each reaction contained 30 Kcpm of 5' end-labeled RNA (25 fmols). The amount of gRNA used in each reaction is specified in the figure legends.

RESULTS

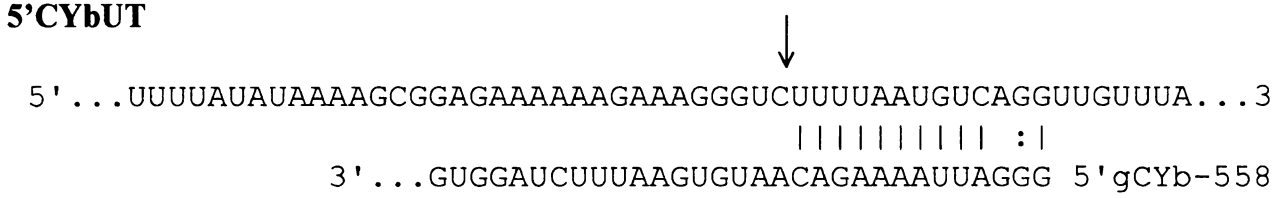
RNA substrates.

5'CYbUT and gCYb-558. Apocytochrome b (CYb) edited message is only produced in the procyclic (insect) and stumpy bloodstream stages of the trypansome life cycle (Feagin et al., 1988; Feagin et al., 1987). Editing involves 13 insertion events, adding 34 U's to the edited domain found near the 5' end of the mRNA. gCYb-558 directs editing at the first seven editing sites, resulting in the addition of 21 U's (Figure 3-1) (Riley et al., 1994). It anneals to the mRNA via an anchor of 13 nt with one mismatch. With a U-tail of 15nts, gCYb-558 is 59 nts in length. The mRNA used in this study was 5'CYbUT, an RNA encompassing 88 nt of the 5' end of unedited CYb (Figure 3-1) (Koslowsky et al., 1996).

3'A6 substrates and gA6-14. Adenosine triphosphate synthase subunit 6 (A6) mRNA is constituitively edited throughout the mRNA. It is extensively edited, with the insertion of 447 U's and the deletion of 28 U's. Editing at the first twelve sites is guided by gA6-14 (Bhat et al., 1990). In order to generate 5' crosslinks using A6 substrates, the original A6 RNA was modified (M) so as to form a gRNA/mRNA anchor duplex which included the 5' most nucleotide of gA6-14. These modified substrates efficiently produced 5' crosslinked molecules. 3'A6UMT contains 91 nts of the 3' end of A6 as well as 16 nts and 20 nts of vector sequence at the 5' and 3' ends, respectively. Two additional A6 substrates were used in this study. Partially edited A6 mRNAs were created with PCR mutagenesis, using 3'A6UMT as a template. 3'A6PES1MT and 3'A6PES3MT are exactly the same as 3'A6UMT except that ES1 and the first 3 editing sites are fully edited, respectively (Figure 3-1).

A.

5'CYbUT



ES1 - Insertion of two U's
5' cleavage product - 67 nts

Figure 3-1. mRNAs and gRNAs used, sequences of interest are shown. (A) 5'CYbUT and gCYb-558. (B) A6 substates and NgA6-14. Bases in bold represent changes in sequence from wild type. gRNAs are shown beneath the mRNA and U's to be deleted are indicated by asterisks. Watson-Crick(|) and G:U (:) basepairs are shown while arrows indicate the expected site of gRNA directed endoribonuclease cleavage.

Figure 3-1 cont'd

B

3'A6UMT

```
                                          **↓
  5'...AGGGGGAGGAGAGAAGAAAGGGAAAGUUGUGAUUUUGGAGUUAUAGUCCAACUUA...3'
                                            |:|||||||||||:||
                    3'...AGUUUUAGACUAAGCAAUAGCCUCAAUAUCGGG 5' NgA6-14
ES1 - deletion of two U's
5' cleavage product - 95 nts
```

3'A6PES1MT

```
                                           ↓
    5'...AGGGGGAGGAGAGAAGAAAGGGAAAGUUGUGAUUGGAGUUAUAGUCCAACUUA...3'
                                           ||:|||||||||||:||
                    3'...AGUUUUAGACUAAGCAAUAGCCUCAAUAUCGGG 5' NgA6-14
ES2 - insertion of two U's
5' cleavage product - 92 nts
```

3'A6PES3MT

```
                                    **↓
5'...AGGGGGAGGAGAGAAGAAAGGGAAAGUUGUUUGUUAUUGGAGUUAUAGUCCAACUUA...3'
                                     |||:|||||:||||||||||:||
                    3'...AGUUUUAGACUAAGCAAUAGCCUCAAUAUCGGG 5' NgA6-14
ES4 - deletion of two U's
5' cleavage product -  89 nts
```

Crosslinked RNAs. gRNAs were transcribed in the presence of guanosine 5' thiophosphate. The resulting thiol group at the 5' end was then attached to an APA group (Burgin and Pace, 1990). To facilitate 5' end-labeling, mRNAs were transcribed in the presence of guanosine. The lack of phosphates in the nucleoside limit guanosine incoporation to the 5' end while making it an efficient substrate for kinase end-labeling. To generate 5' crosslinks, gRNAs and mRNAs were annealed and exposed to 312nm light as previously described (Leung and Koslowsky, 1999). 5' crosslinking produces one mRNA dependent crosslink that has been previously mapped to the 3' ends (mRNA orientation) of the gRNA/mRNA anchor duplexes. Therefore, these crosslinks maintain the anchor duplex, an element that readily forms and is required for the process of editing. Crosslinks were gel purified and the mRNAs in the crosslinks were 5' end-labeled using T4 kinase.

Crosslinked substrates support editosome assembly

To examine whether 5' crosslinked molecules were recognized by the editosome, we first looked to see if they would support the initial enzymatic activity in editing, gRNA-directed endoribonuclease cleavage of the mRNA. To improve visualization of cleavage, ATP and UTP were not added to reactions, suppressing ligation activity. Four different 5' crosslinks were examined, gCYb-558 with 5'CYbUT as well as NgA6-14 with 3'A6UMT, 3'A6PES1MT and 3'A6PES3MT. When incubated in mitochondrial lysate, all 4 crosslinks supported accurate gRNA directed endonuclease cleavage (Figure 3-2).

Crosslinked 5'CYBUT produced a single cleavage product of 67 nts as expected for cleavage at ES1 where two U's are inserted. This differed from 3'A6UMT which produced two products of 95 and 94 nts. The 95 nt product corresponded to correct cleavage at ES1 (a deletion site). 3'A6PES1MT was cleaved at ES2 (a uridylate insertion site) resulting in a single 5' product of 92 nts while cleavage of 3'A6PES3MT at ES4 (a deletion site) produced another doublet (89 and 88 nts). In all cases, the crosslinked RNAs were cleaved more efficiently than the free mRNA controls (Figure 3-2). Considerably less cleavage product was detected utilizing 3'A6UMT, in comparison with the other substrates, suggesting that conditions were not optimal for cleavage at ES1 in A6UMT. Cruz-Reyes et al. (Cruz-Reyes et al., 1998b) have proposed that insertion and deletion sites are differently optimized, with deletion sites requiring ATP (3mM). However, the strong cleavage at ES4 which is also a deletion site, in the absence of exogenous ATP, suggests that this may not be the case. The production of heterogeneous 5' cleavage fragments have been previously observed at sites of deletion and is attributed to U-specific exonuclease activity (Cruz-Reyes and Sollner-Webb, 1996). This suggested the single band observed for the insertional substrates was due to a lack of exogenous UTP, preventing TUTase activity. When UTP was added an additional 5' fragment 1 nt larger was observed for crosslinked 5'CYbUT (Figure 3-3A). This implied that a single U had been added to the 3' end of the 5' fragment via TUTase. The gRNA for 5'CYbUT directs the insertion of two U's at the site in question, however, the addition of only one U was observed. We expect that the addition of two U's, preferentially targets this 5' product for religation as previously reported, making it difficult to detect (Igo et al., 2000; Kable et al., 1996). Similarly, for the deletional substrates the detection of only the

Figure 3-2. gRNA directed endoribonuclease cleavage of 5' crosslinked substrates, (A) 5'CYbUT+gCYb-558. Lane 1, T1 ladder. Lanes 2 and 3 are crosslinked substrate in the presence and absence of lysate, respectively. Lanes 4 and 5 represent cleavage of free 5'CYbUT in the presence or absence of gCYb-558. Free 5'CYbUT not incubated with mitochondrial lysate is shown in lane 6. Lanes 7 and 8 are identical to lanes 4 and 5, except the mRNA has been periodate treated. The triangle indicates most likely circularized mRNA while the dotted arrow shows cleavage by a previously described non-specific endonuclease activity(Piller, 1997). (B) 3'A6UMT+NgA6-14, (C) 3'A6PES1MT+NgA6-14 and (D) 3'A6PES3MT+NgA6-14. Lane 1, crosslink incubated with lysate. Lane 2, crosslink incubated with lysate and 10X molar concentration of free gRNA. Lane 3, no lysate control. Lane 4, free mRNA and gRNA (1:1 ratio) incubated with lysate. Lane 5, the same as lane 4 except a ratio of 1:10 was used. Lane 6, mRNA:gRNA ratio of 1:10 with no lysate. The crosslink (X), mRNA (M), expected cleavage site (arrow) and cleavage products (asterisks) are indicated on the left. The A6 cleavage assays depicted here were carried out by Dr. Koslowsky.

A.

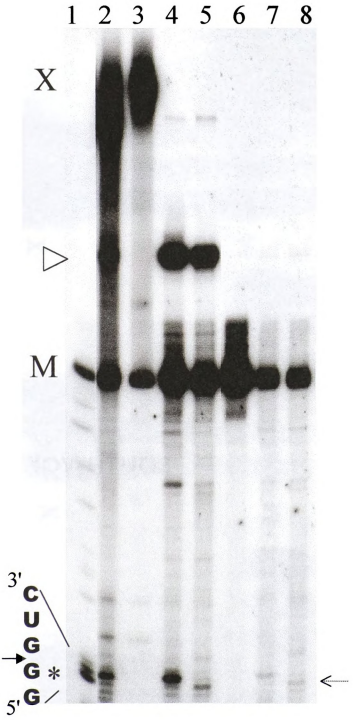
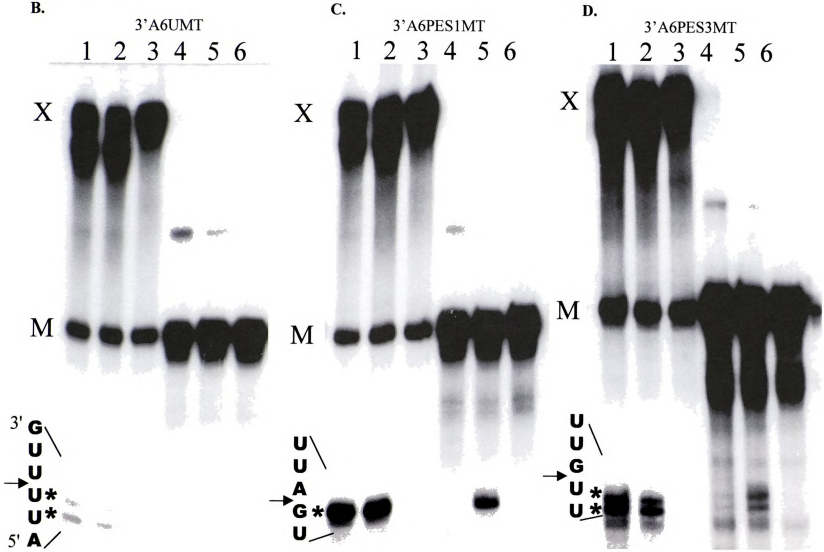


Figure 3-2 cont'd



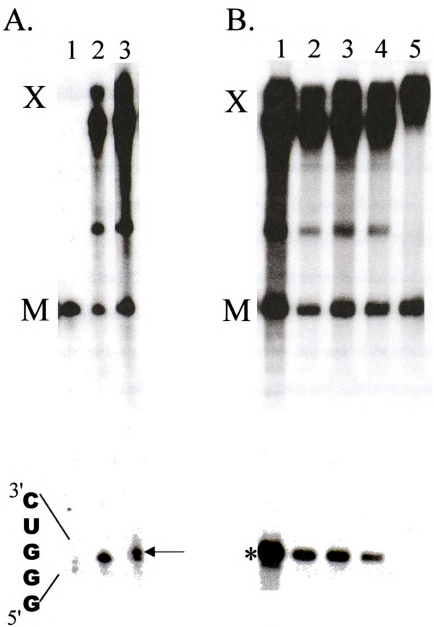


Figure 3-3. The effect of exogenous UTP and gRNA on gRNA directed cleavage of crosslinked 5'CYbUT. (A) Lane 1, T1 ladder of 5'CYbUT. Lane 2, crosslink incubated with lysate. Lane 3, crosslink incubated with lysate and 0.5 mM UTP. The UTP dependent band is highlighted by the arrow. (B) Addition of 6X (lane 2), 12X (lane 3), and 40X (lane 4) excess gCYb-558 to crosslinked 5'CYbUT. Lane 1 contains no exogenous gRNA while lane 5 represents a no lysate control. The crosslink (X), the cleavage product (asterisk) and mRNA (M) are shown on the left.

full length 5' fragment and a 5' fragment one nt smaller may indicate that removal of both targeted uridylates preferentially targets this fragment for religation. Despite not adding exogenous ATP to the reactions, ligase activity was still present as shown by the band above the free mRNA (Figure 3-2A, triangle). This was most likely due to the presence of ligase pre-charged with ATP. Periodate treatment blocks formation of this slower mobility product, suggesting that it is circularized mRNA (Figure 3-2A). The band observed at ES2 in the absence of gCYb-558 (Figure 3-2 A, lane 5, 8) corresponds with a non-gRNA directed endonuclease activity characterized by Piller et al. (1997).

Isolation and handling of crosslinks resulted in the breakage of a fraction of the crosslinks, releasing free mRNA (Burgin and Pace, 1990; Wower et al., 1989). We were concerned that the cleavage could be generated by cleavage of the released mRNA in our reactions. When a crosslink breaks it releases gRNA and mRNA in equimolar amounts. Current *in vitro* editing assays require the use of excess gRNAs to drive the reaction (Byrne et al., 1996; Kable et al., 1996; Seiwert et al., 1996). gRNA directed endonuclease cleavage of gRNAs and mRNAs at a 1:1 ratio is very inefficient (Figure 3-2B). In contrast, crosslinked substrates support efficient cleavage in the absence of exogenous gRNAs. Furthermore, the addition of excess gRNA did not increase cleavage of crosslinked 5'CYbUT (Figure 3-3B).

These data demonstrate that the 5' crosslinked substrates support gRNA directed endonuclease activity in addition to U-specific exonuclease and TUTase activity. This suggests that the 5' crosslinks interact correctly with the editosome.

Utilization of crosslinked RNAs for solution structure probing

Mapping the position of the 5' and 3' ends of gRNAs along their cognate mRNAs allowed us to predict secondary structures using these two constraints (Leung and Koslowsky, 1999; Leung and Koslowsky, 2001). These predicted structures contained three common elements: (1) a gRNA/mRNA anchor duplex, (2) a U-tail/mRNA duplex and (3) a gRNA stem-loop (Figure 3-4). To provide additional evidence for the predicted secondary structures, 5'CYbUT was crosslinked to gRNA with (gCYb-558) and without a U-tail (gCYb-558sU, sans U). These crosslinks were then 5' end-labeled (mRNA) and structure probed in parallel, allowing us to investigate the interactions of the U-tail with it's cognate message. In addition, we also examined the structure of mRNA alone as it is predicted to form a stable stem-loop structure that must be disrupted in order for the gRNA to interact with the mRNA (Figure 3-5). The mRNA is thought to fold back on itself, forming a large stem-loop structure, with several small internal loops (Piller et al., 1995).

Ribonuclease T1 and T2 along with Mung Bean nuclease (MBN) were used to identify single stranded sequences while methidiumpropyl-EDTA-iron(II) (MPE-Fe(II)) was used to probe double-stranded regions. In experiments using crosslinks, mapping cleavages downstream of the crosslinks was not possible as the 5' fragments were linked to gRNA, slowing their mobility unpredictably. Therefore mRNA sequence downstream of the anchor duplex was not examined. The structure probing data obtained correlated well with the predicted structures of free 5'CYbUT and crosslinked 5'CYbUT+gCYb-558.

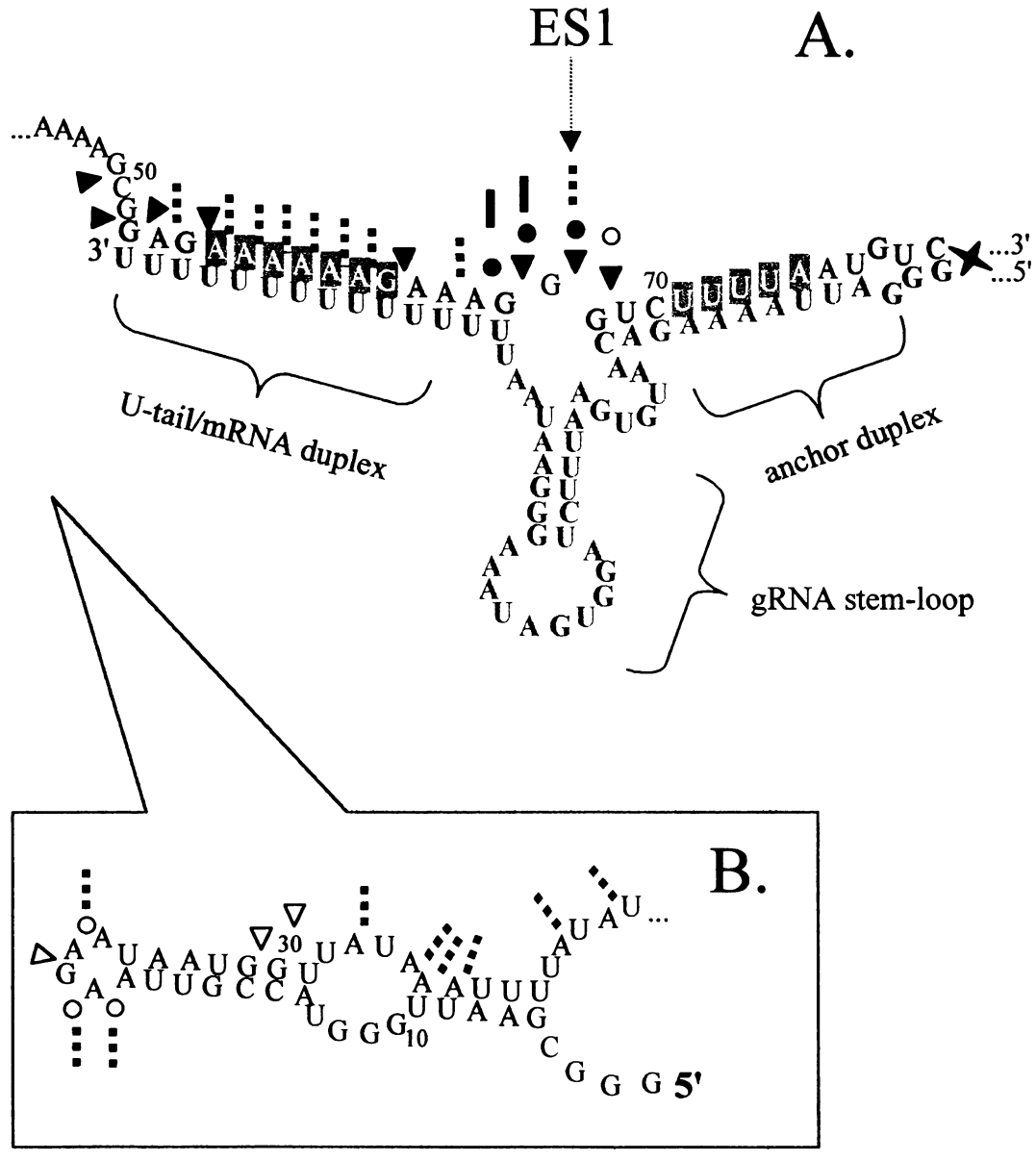


Figure 3-4. Predicted secondary structure of interacting gCYb-558 and 5'CYbUT. (A) Predicted structure of interacting sequences of gRNA and mRNA. (B) Represents the 5' sequence of the mRNA not predicted to interact with the gRNA. Cleavages by T1 (triangle), T2 (circle) , MBN (line) are shown on the structure. Regions of intense MPE-Fe(II) cleavage is highlighted by shaded sequences. Open circles and triangles, and dotted lines indicate weaker cleavages. The crosslink is represented by a star.

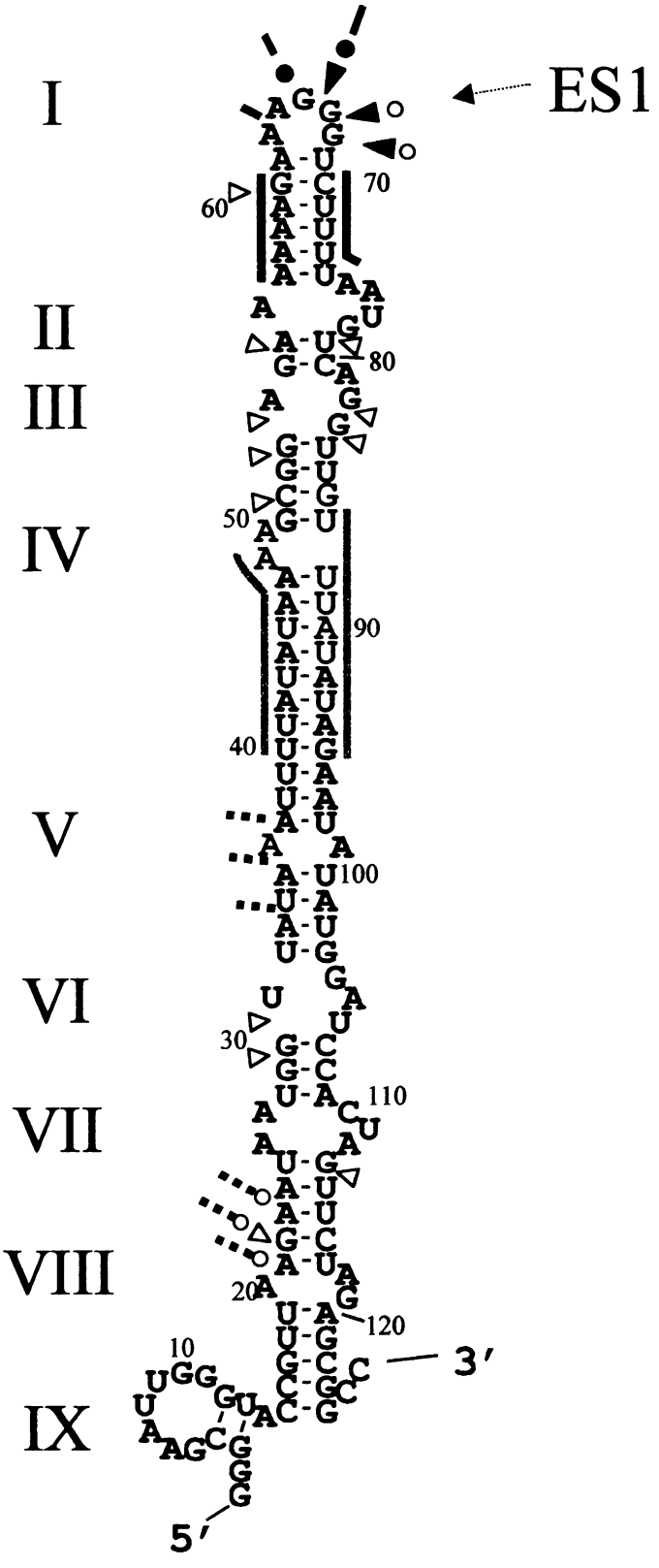


Figure 3-5. Predicted secondary structure of 5'CYbUT alone. Cleavages are mapped on the structure (see Figure 3-4 for legend). Two regions of high MPE-Fe(II) sensitivity are highlighted by the bars in black and dark gray. Roman numerals indicate internal loops and bulges. Editing site one is also indicated (ES1).

Single strand specific nucleases

RNase T1. RNase T1 was used to identify unpaired G's, as T1 specifically cleaves 3' of single stranded G's (Figure 3-6). T1 probing of mRNA alone revealed three G's (G66, G67 and G68) that were very sensitive to cleavage. These G's map to the terminal loop of the predicted structure of 5'CYbUT which contains the first three editing sites (Figure 3-5). The sensitivity of the loop to T1 was in agreement with previous observations (Piller et al., 1995). The remainder of the cleavages in the mRNA were minor, suggesting that the stem-loop structure of 5'CYbUT is stable. Several minor cleavages were observed in internal loops, supporting the presence of loops II and III. Other minor cleavages mapped to short helices adjacent to internal loops (IV, VI, VII and VIII) where one might expect some breathing and hence susceptibility to T1 cleavage.

Probing of gRNA-crosslinked RNA with T1 revealed that the presence of gCYb-558 significantly reduced the susceptibility of the three G's previously found in the terminal loop of 5'CYbUT alone (Figure 3-6B, C). Unexpectedly, its presence did not appear to greatly affect cleavage at any other sites. Furthermore, no significant changes in sensitivity were observed in the presence or absence of the U-tail (gCYb-558 and gCYb-558sU, respectively). The U-tail is predicted to interact with an purine rich sequence which we term the upstream U-tail Stabilization Element (uUtSE). This element covers bases G53 through G67 (Figure 3-4). We expected protection of up to five G's within this sequence in the presence of the U-tail/mRNA duplex. However, no significant differences in the reactivity of these G's was observed using gCYb-558 or gCYb-558sU. G's at the ends of the duplex (G53, G55, G66 and G67) might be cleaved due to breathing, explaining the lack of protection of these bases. On the other hand,

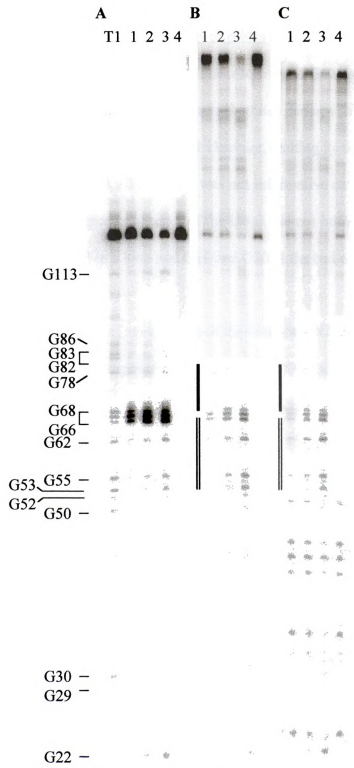


Figure 3-6. RNase T1 digestion of (A) 5'CYbUT alone, (B) 5'CYbUT+gCYb-558 and (C) 5'CYbUT+gCYb-558sU. Lanes 1-3, increasing amounts of T1 (0.025, 0.05 and 0.1 U). Lane 4, undigested RNA. T1, represents a T1 ladder. The solid line and double line denote the anchor duplex and uUtSE, respectively. Cleavage sites are indicated on the left.

equal cleavage of G62 in both cases (plus or minus U-tail) was puzzling and could not be reconciled with the predicted structure. The reactivity of G62 appears to be relatively equal in free 5'CYbUT and the crosslinked substrates perhaps suggesting that the accessibility to G62 is not optimal for T1 cleavage, making its suceptibility to cleavage difficult to assess. Overall, T1 cleavage was not informative concerning the U-tail interaction, as the presence or absence of the U-tail did not influence cleavage. Other T1 cleavages included G22 in a loop and G29 and G30 at the end of a predicted duplex that might breathe.

RNase T2. Continuing our investigation of the U-tail/mRNA interaction we turned to RNase T2 which cleaves single-stranded RNA with no sequence specificity. T2 also highlighted the terminal loop of 5'CYbUT alone, showing a preference for A65-G68 (Figure 3-7). The rest of 5'CYbUT was remarkably resistant to T2 except for A20 and A21 which are present in loop VIII. A23 is also weakly cleaved. Loops VII and VIII flank the short helix A23 is found in, suggesting the helix breathes.

A sharp reduction in cleavage in the region containing the initial editing sites was once again observed in the presence of crosslinked gRNA (Figure 3-7B, C). Despite this inhibition, T2 preferentially cleaved in this region at the same nucleotides as mRNA alone (A65-G68). In contrast to T1, a clear difference in T2 cleavage of the uUtSE was observed, depending on whether the crosslinked gRNA contained a U-tail or not. Crosslinks lacking the U-tail showed increased T2 cleavage of the uUtSE at seven adenylates (A54, A56-A61). gRNAs with a U-tail did provide protection, as cleavage at the same adenylates within the uUtSE was noticeably reduced. This indicated that the U-tail did basepair with the uUtSE, supporting the predicted structure. Cleavages at A20,

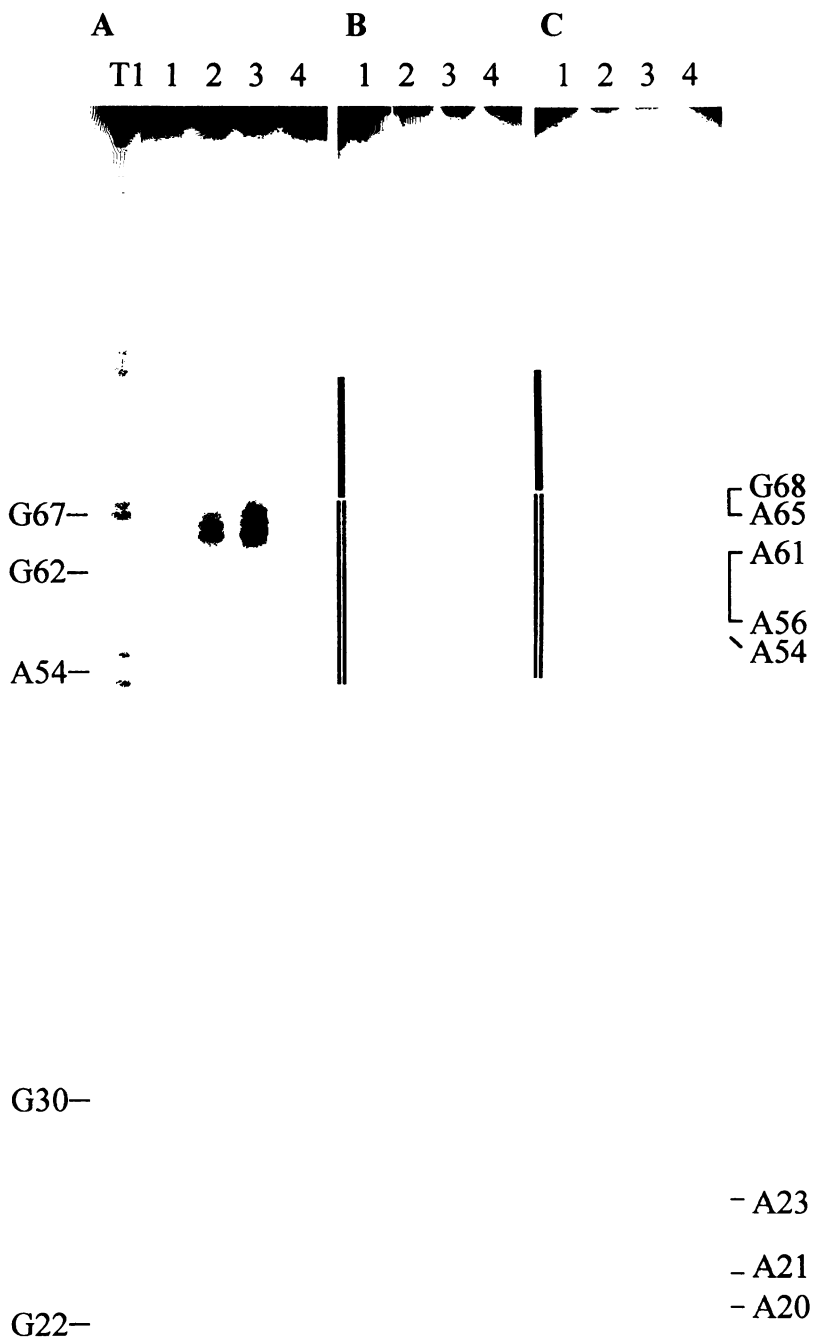


Figure 3-7. RNase T2 probing of (A) 5'CYbUT alone, (B) 5'CYbUT+gCYb-558 and (C) 5'CYbUT+gCYb-558sU. Lanes 1-3, increasing amounts of T2 (0.002, 0.002 and 0.04 U). Untreated RNA is shown in lane 4 and T1 is the ladder. Selected nucleotides of the ladder are indicated on the left. T2 cleavage sites are highlighted on the right. Lines are as in Figure 3-6.

A21 and A23 indicated the presence of a loop in agreement with the T1 data. Additional cleavages with T2 were not consistently observed.

Mung Bean Nuclease. The last single strand specific probe used was mung bean nuclease (MBN). Like T2, this nuclease does not have any strong sequence specificity. Within the mRNA alone, the most sensitive sites were again limited to the terminal loop (A64-G67, Figure 3-8A). We observed that MBN preferentially cleaved at two nucleotides (A65 and G66), as previously observed (Piller et al., 1995). The T1 ladder did not line up with the MBN bands because MBN produces fragments with a 3'OH group as opposed to T1 cleavage which produces a 2'3' cyclic phosphate (Cruz-Reyes et al., 1998a). MBN also cleaves at A20, A21 and A23 within a predicted loop also sensitive to T2. Weak MBN cleavage was observed at A33, A35 and A36. A33 is found within a short 4 bp helix that might breathe, while A35 and A36 highlight a bulge in a duplex, due to a mismatch.

Probing with MBN revealed significant differences between free mRNA and crosslinked mRNA (Figure 3-8B, C). The presence of the gRNA resulted in only limited cleavage of A64-G67, the most sensitive sites in 5'CYbUT alone. However, the most striking difference was the cleavage of the uUtSE sequence. Crosslinks formed with gCYb-558sU (no U-tail) showed significant cleavage at the identical adenylates of the uUtSE sequence, previously observed with T2 probing. The intensity of cleavage approached that of A65 and G66, the most dominant cleavage sites. In contrast, the crosslinks with U-tails, showed a marked reduction in cleavage of the uUtSE. This correlated with protection by the U-tail. These data supported the predicted model, but indicated that the U-tail/mRNA duplex is a flexible interaction that breathes, allowing

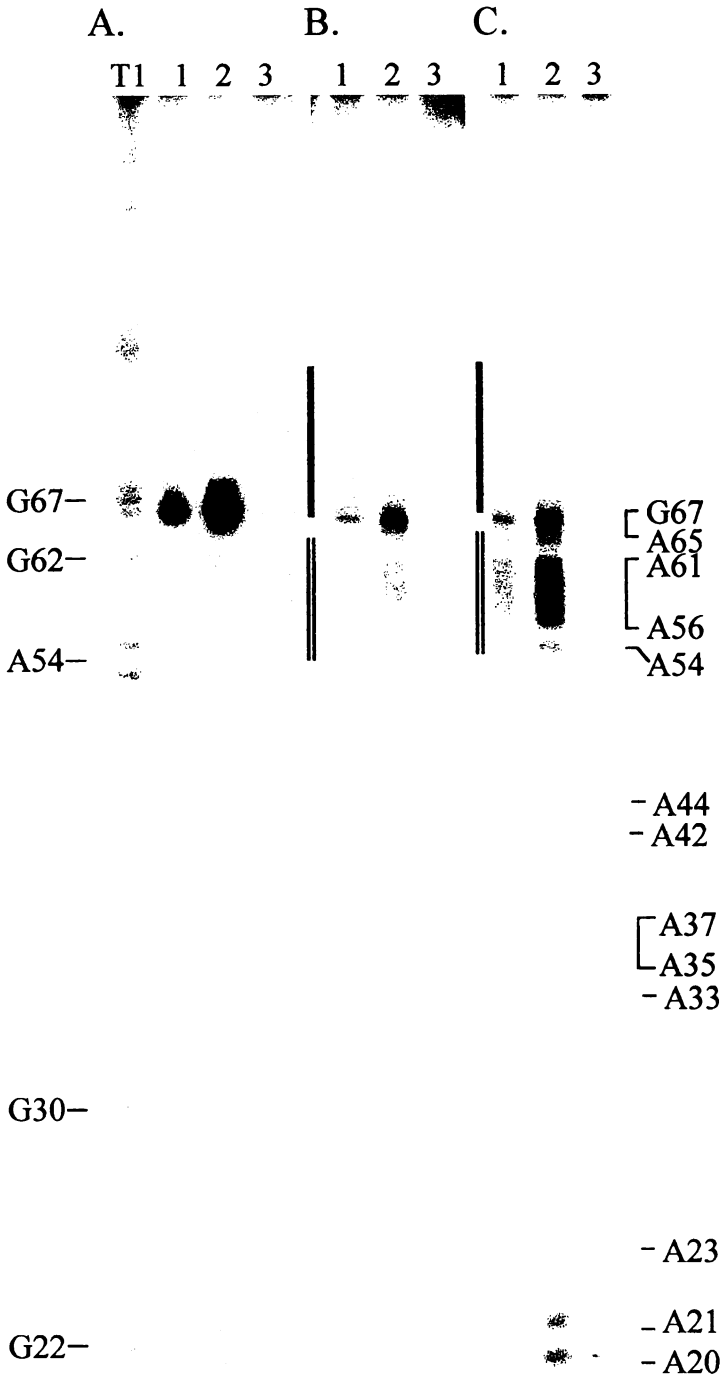


Figure 3-8. Mung Bean Nuclease digestion of (A) 5'CYbUT alone, (B) 5'CYbUT+gCYb-558 and (C) 5'CYbUT+gCYb-558sU. Increasing amounts of MBN (0.25, 0.5 U) were used (Lanes 1-2). RNA not incubated with MBN is shown in Lane 3 while the T1 ladder is labeled T1. Selected nucleotides of the ladder are shown on the left. Sites of MBN cleavage are indicated on the right. Lines are as in Figure 3-6.

MBN to cleave the mRNA some of the time. Cuts at A20, A21 and A23 support the presence of a predicted loop in the mRNA. Weaker cuts at several A's (A33, A35-A37, A42 and A44) were in agreement with two single strand regions in the predicted model.

MPE-Fe(II)

Basepaired regions were examined using methidiumpropyl-EDTA- iron(II) as RNase V1 is no longer commercially available. In the presence of O_2 and DTT, intercalated MPE-Fe(II) produces hydroxyl radicals that diffuse and cleave both strands of a helix (Hertzberg and Dervan, 1984; Kean et al., 1985). Due to this diffusion, MPE-Fe(II) cleavage produces 3-5 cuts on both strands. Intercalation sites are identified by a peak of cuts (3-5 bases) on both strands, with helix geometry causing these cuts to be offset in the 3' direction (Kean et al., 1985; Schultz and Dervan, 1983).

Probing of free 5'CYbUT with MPE-Fe(II) did not produce discrete sets of cuts, as virtually all the nucleotides examined were cleaved (Figure 3-9). However, it was clear that particular regions were much more sensitive to MPE-Fe(II) than others. Using a phosphorimager for quantitation, regions of increased sensitivity were mapped along 5'CYbUT alone, revealing that these regions corresponded to predicted sites of basepaired RNA. Figure 3-10A represents the quantitation of lane 6 in Figure 3-9 (counts represent an arbitrary unit used by the data analysis program, Imagequant 5.0). In this figure, the regions of increased sensitivity are paired (black and dark gray), indicating cleavage on both strands of a duplex and perhaps indicating sites of MPE-Fe(II) intercalation. Gels run for a shorter time indicate that the region of sensitivity at the right end of the graph (A46-A48) continues farther upstream (data not shown). The

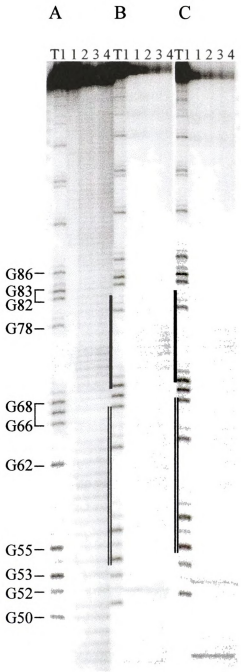
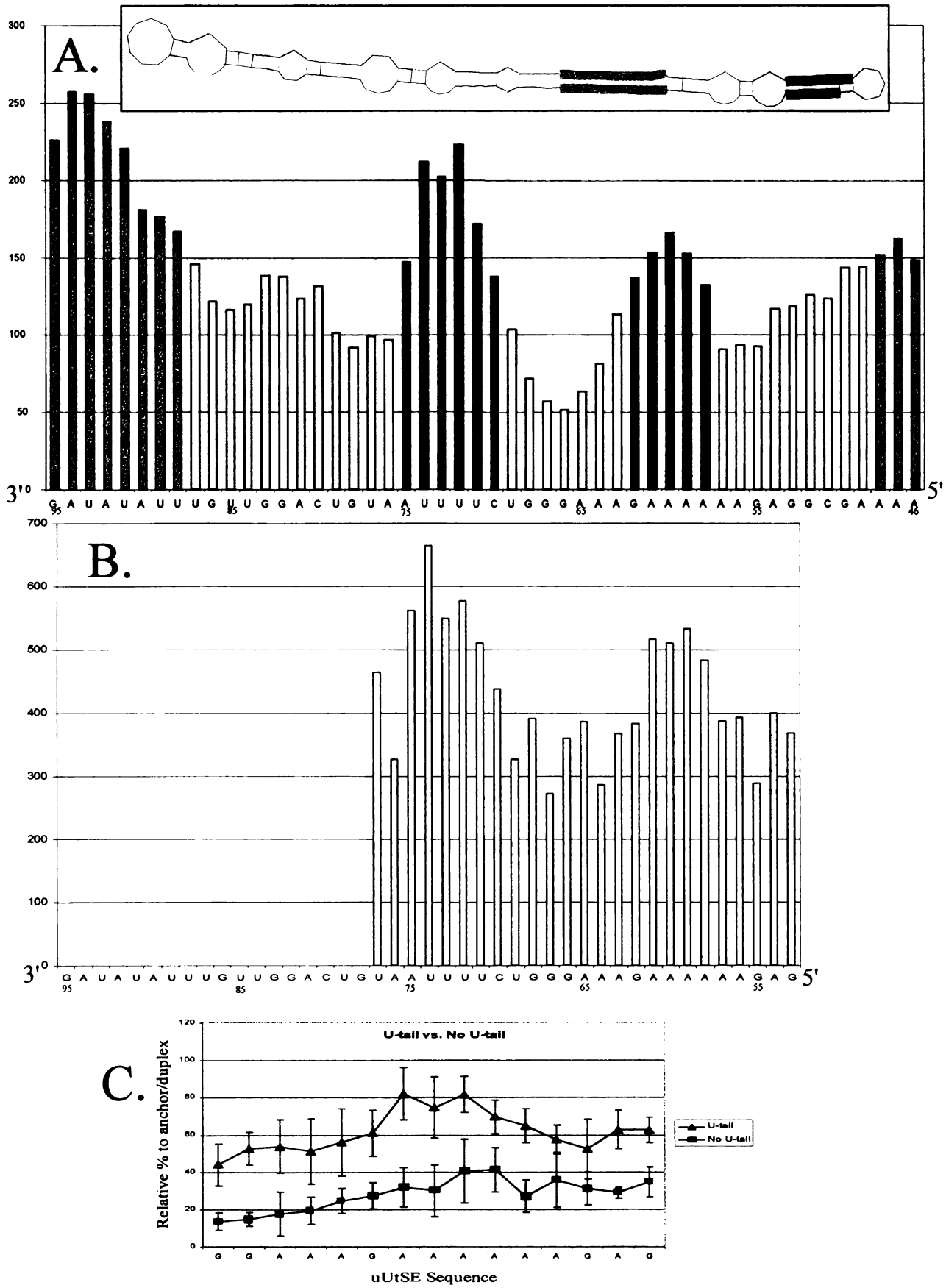


Figure 3-9. MPE-Fe (II) was used to probe double stranded regions of (A) 5'CYbUT alone, (B) 5'CYbUT+gCYb-558 and (C) 5'CYbUT+gCYb-558sU. Aliquots were taken every two minutes (lanes 1-4 correspond to 0, 2, 4 and 6 min incubations). T1 represents the ladder. Lines are as in Figure 6. Selected nucleotides of the ladder are shown on the right. Each reaction (A, B, and C) was loaded and run into the gel before the next reaction, hence the offset in the bands.

Figure 3-10. Quantitation of MPE-Fe(II) cleavage. (A) Cleavage of 5'CYbUT alone was quantitated using a phosphorimager. The regions that show the highest sensitivity to MPE-Fe(II) are highlighted in black and dark gray. These regions correspond to two duplexes outlined by black and dark gray bars in the predicted structure. Data from shorter run gels suggests the region at the right (A46-A48) extends farther upstream, hence the top dark gray line is longer than what is shown in the bar graph. (B) Probing of 5'CYbUT+gCYb-558. The sequence bound by the gRNA anchor and the uUtSE both show increased cleavage by MPE-Fe(II). (C) The presence of the U-tail results in consistently higher MPE-Fe(II) cleavage.


A.
300
250
200
150
100
50
3'
5'
B.
700
600
500
400
300
200
100
3'
5'
C.
U-tail vs. No U-tail
Relative % to anchor/duplex
120
100
80
60
40
20
0
U-tail
No U-tail
uUtSE Sequence

MPE-Fe(II) data supports two double-stranded regions, one just below the terminal loop (black) and the other in the middle of the mRNA (dark gray, including the information from shorter run gels). The cleavages just below the terminal loop appear to be the result of a single intercalation site by MPE-Fe(II), producing a set of increased cleavages within a short region. The cleavages of the middle duplex were not confined to a short region, suggesting multiple intercalation sites, thus producing the broad shoulder of increased cuts (Figure 3-10 and data not shown). The two regions of double stranded RNAs identified by MPE-Fe(II) are in agreement with the predicted structure of mRNA alone.

Crosslinked RNAs showed two areas of increase MPE-Fe(II) cleavage, suggesting two regions of double stranded RNA. In the presence of the gRNA the anchor duplex, maintained by the crosslink, was highlighted by cleavage as expected (Figures 9). Figure 3-10B shows the quantitation of the 6 minute lane of 5'CYbUT+gCYb-558. The nucleotides examined represent the sequence of interest: predicted basepairs which must be disrupted in order for the gRNA to bind. Quantitation of cleavage upstream of the crosslink was not possible, due to the presence of the gRNA. Bases just 5' of the crosslink did not produce a quantifiable signal suggesting that the aryl group of the linkage might have sterically interfered with cleavage. The region of ES1 was much less reactive, again emphasizing that the first three editing sites were single stranded in nature. To compare MPE-Fe(II) probing of the uUtSE in the presence or absence of the U-tail, the intensities of cleavages were normalized against the strongest cleavage site in the anchor/duplex. We assumed that MPE-Fe(II) reactivity within the anchor/duplexes in the presence or absence of the U-tail would be equal as the sequences in this region are identical and the duplex is stabilized by the crosslink. The average of three experiments

was used to compare the crosslinks. In the presence of the U-tail, cleavage was strongest at A59, A60, and A61, with a reactivity approaching 80% of the anchor/duplex (Figure 3-10C). Crosslinks produced with gCYb-558sU (no U-tail) showed consistently weaker cleavage in the uUtSE with only 40% of the reactivity observed in the anchor duplex (Figure 3-10C). These data were consistent with basepairing of the uUtSE only in the presence of the U-tail, as predicted.

Predicted Structures

The data obtained from the nuclease and chemical probing support both predicted structures with only a minor modification to the crosslinked mRNA structure. 5'CYbUT alone, folds into a stem-loop, making the terminal loop, which contains ES1-3, particularily accessible to single strand specific nucleases. The stem of 5'CYbUT was particularily resistant to nuclease cleavage, suggesting that it is quite stable while MPE-Fe(II) reactivity outlined two helices consistent with the predicted structure.

Crosslinking the gCYb-558 to 5'CYbUT clearly changed the structure of the mRNA. The intense cleavage at ES1 through ES3 in the terminal loop of the free mRNA was significantly reduced in the presence of crosslinked gRNA. In addition, probing of crosslinked 5'CYbUT provided direct evidence for the U-tail/mRNA duplex. The presence of the U-tail showed a clear reduction in cleavage of the uUtSE as opposed to crosslinks without a U-tail. Furthermore, elevated MPE-Fe(II) cleavage was observed in the uUtSe in the presence of the U-tail in comparison with no U-tail. Both of these observations are indicative of a U-tail/mRNA duplex. This duplex cannot be characterized as a strong and stable interaction as the uUtSE can still be weakly cleaved

by single strand specific nucleases. In comparison, it is interesting to note that the duplex in free 5'CYbUT immediately below the terminal loop is much more resistant to nuclease attack, and involves sequences found in both the gRNA/mRNA duplex and the U-tail/mRNA duplex.

Within the predicted structure for the gRNA/mRNA complex, 12 nts (U41-G52) are predicted to be single stranded, however, this region had a distinct lack of cleavages (Figure 3-11A). T1 was able to cleave at two of the G's (G50, G52) in the sequence while only two additional weak cuts (MBN at A42 and A44) were observed. This suggested that the mRNA in this region was not single stranded. The lack of cleavage at nucleotides A46-A49 suggested that these four 4 nts were basepaired. Using this constraint, an alternative structure was predicted for this upstream region (Figure 3-11 B), which better fit the cleavages observed.

DISCUSSION

The process by which active editosomes assemble onto gRNA/mRNA complexes is still unknown. Both gRNAs and mRNAs alone are able to support formation of RNP complexes that may represent specific steps in the assembly process (Göringer et al., 1994; Koslowsky et al., 1996; Read et al., 1994). However, we know that editing requires the interaction of both RNAs, arguing that final editosome assembly requires specific aspects of this interaction. Our previous crosslinking study mapped the position of the 5' and 3' ends of gRNAs along their cognate mRNAs. These data led to a predicted common structure for interacting gRNA/mRNA pairs involving a gRNA/mRNA anchor duplex, an upstream U-tail/mRNA duplex and a gRNA stem-loop. This suggests that

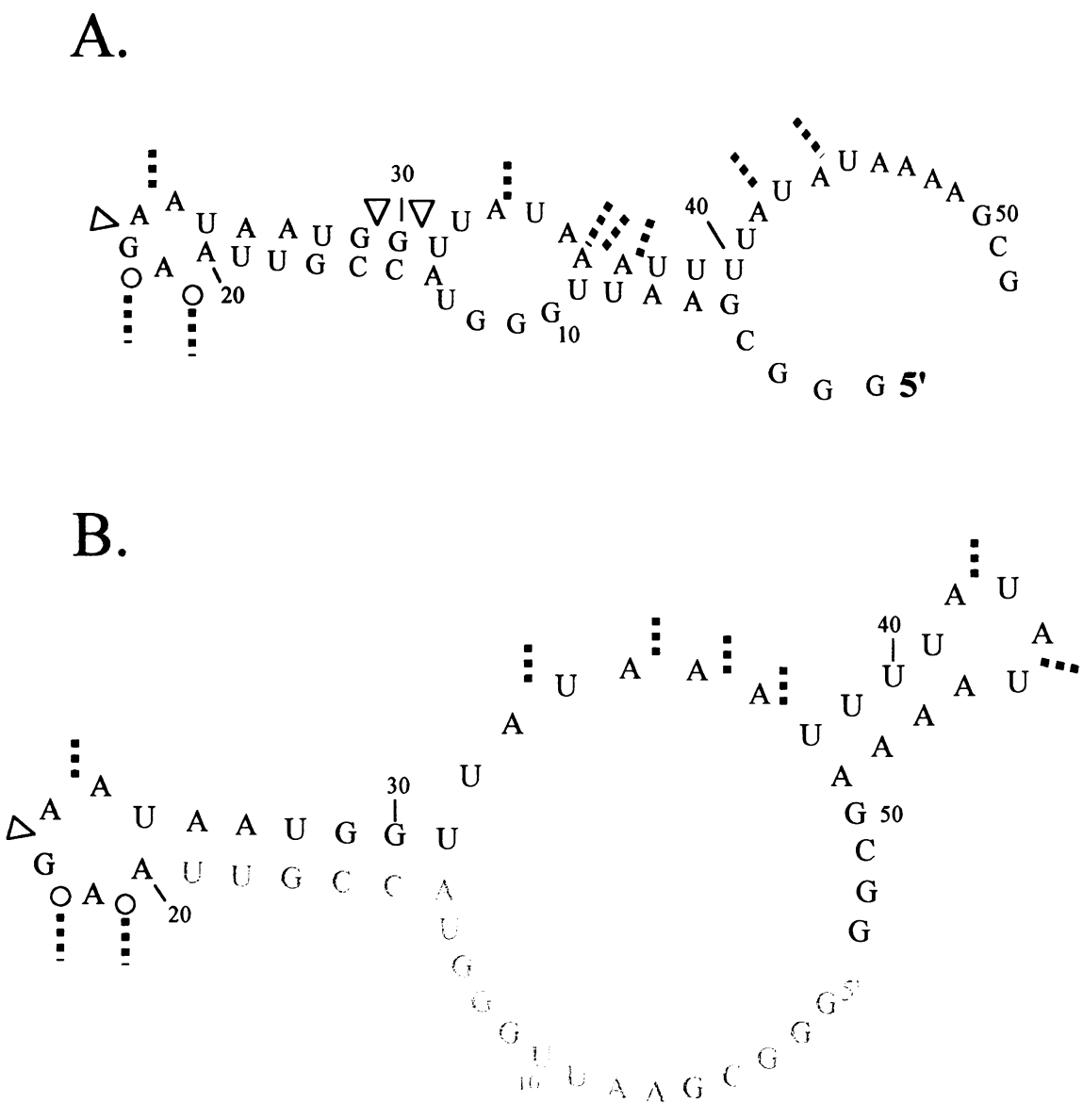


Figure 3-11. An alternative structure for the 5' end of 5'CYbUT crosslinked to gCYb-558. (A) The original predicted structure of the region. (B) An alternative structure that is a better fit for the probing data. No data for nucleotides below A20 was collected. See Figure 3-4 for the legend.

structural aspects of the gRNA/mRNA interaction could play a role in editosome assembly. To provide additional evidence in support of our predicted structures, we solution probed 5'CYbUT crosslinked to gCYb-558. We chose to crosslink the two RNAs together as probing of two interacting free RNAs introduces several technical problems.

5' crosslinked molecules are biologically relevant as they support activities connected with editing. Accurate cleavage of all crosslinked substrates was observed at the appropriate editing sites demonstrating that the RNAs supported gRNA directed endoribonuclease activity. Furthermore, the UTP-dependent addition of a nucleotide to the 5' cleavage product of crosslinked 5'CYbUT was indicative of TUTase activity, while at sites of U deletion, U-specific exonuclease activity on 5' cleavage products was observed. These enzymatic activities have been shown to associate together in a complex using glycerol fractionation and chromatography (Corell et al., 1996; Pollard et al., 1992; Rusché et al., 1997). This argues that the editing complex interacts and assembles correctly on the crosslinked RNAs.

In the solution probing of crosslinked 5'CYbUT, single-stranded and double-stranded regions were identified using enzymatic and chemical probes, respectively. The structure of free 5'CYbUT was also investigated as gCYb-558 must disrupt the structure of the mRNA in order to bind. We were interested in observing the changes required to bind gCYb-558 as the predicted structure of 5'CYbUT is a large stable stem-loop (Piller et al., 1995). Overall the probing data supported the predicted stem-loop structure of free 5'CYbUT.The terminal loop containing the first three editing sites was clearly identified by single-strand specific nucleases and within the stem, two predicted double-stranded

regions showed increased sensitivity to MPE-Fe(II). Minor nuclease cleavages were also observed highlighting several internal loops

The presence of crosslinked gRNA drastically changed the structure of 5'CYbUT as expected. Although the initial three editing sites were cleaved by the single strand specific nuclease, the sensitivitiy of this region was significantly reduced in comparison with the terminal loop of free 5'CYbUT. It was interesting to note that none of the nucleases showed a strong preference for ES1 perhaps suggesting that gRNA/mRNA pairs interacts with the editosome so as to present the correct editing site to the endoribonuclease component. Just upstream of this region in the mRNA, the presence of the U-tail/mRNA duplex was supported by both the single-strand specific nucleases as well as the double-strand specific MPE-Fe(II). Enhanced cleavage in the uUtSE (A54, A56-A61) by MBN and T2 in the absence of the U-tail contrasted with very weak cleavages in the presence of the U-tail. This indicated that the U-tail basepairs with the uUtSE, providing protection to this sequence. However, cleavage is not completely abolished in the presence of the U-tail suggesting that this interaction is not very stable. The presence of the U-tail also correlated with increased MPE-Fe(II) cleavage in the uUtSE, indicative of MPE-Fe(II) intercalation. The only modification to the predicted structure of the gRNA/mRNA pair suggested by the probing data was immediately upstream of the uUtSE. A region of 12 nucleotides (U41-G52) was initially predicted to be single stranded. However, it was weakly cleaved by T1 and MBN at only 4 of these nucleotides. An alternative structure (Figure 3-11B) better fits the observed cleavages in the 5' half of the mRNA although the biological significance of this structure is questionable as the mRNA is truncated in this region. In general, the structure probing

data from this study correlated well with our predicted secondary structure of gRNA/mRNA pairs.

The presence of the U-tail/mRNA duplex supports the U-tail's role as a tether to hold on to the 5' cleavage mRNA product during editing. The fact that this interaction is not very stable is also in agreement with experiments demonstrating that modified gRNA sequences that interact more stably with the uUtSE function as better tethers (Burgess et al., 1999; Kapushoc and Simpson, 1999; Seiwert et al., 1996). Despite being presented with the full range of available upstream purines, the U-tail basepairs to mRNA sequences just upstream of the initial editing sites. The formation of two duplexes that flank the first few initial editing sites raises the possibility that the two duplexes function in concert to correctly present the initial editing sites to the editing complex.

In addition, the data suggests that the disruption of the stable stem-loop structure of 5'CYbUT by gCYb-558 is most likely energetically unfavorable. Most of the mRNA anchor nucleotides that gCYb-558 would initially basepair with, are part of a strong duplex within free 5'CYbUT. Indeed the K_d of this interaction is very high in comparison with other gRNA/mRNA interactions (D. Koslowsky, unpublished results). The duplexed stem in free mRNA is quite resistant to nuclease activity and appears to be very stable. Although no conclusions of the stability of the gRNA/mRNA anchor duplex could be made due to the introduced crosslink, the nuclease probing data of the U-tail/mRNA duplex indicates that this is not a very stable interaction. This suggests that efficient interaction of the two RNAs could be affected by environmental conditions such as ionic strength or could require additional factors. Reaction conditions that affect gRNA/mRNA association are currently being investigated in our laboratory.

The results of this study indicate that 5' crosslinked RNAs are ideal substrates to pursue the structural aspects of gRNA/mRNA interactions that could play a role in full editosome assembly.

REFERENCES

Bhat G. J., Koslowsky D. J., Feagin J. E., Smiley B. L., Stuart K. 1990. An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. Cell **61**(5):885-94.

Blum B., Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell **62**(2):391-7.

Burgess M. L., Heidmann S., Stuart K. 1999. Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit in vitro U insertion. RNA **5**(7):883-92.

Burgin A. B., Pace N. R. 1990. Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. EMBO J. **9**(12):4111-8.

Byrne E. M., Connell G. J., Simpson L. 1996. Guide RNA-directed uridine insertion RNA editing in vitro. EMBO J. **15**(23):6758-65.

Christiansen J., Egeberg J., Larsen N., Garrett R. A. 1990. In: Speding G., editor. Ribosomes and Protein Synthesis: A Practical Approach. New York: Oxford University Press. p 229-252.

Corell R. A., Read L. K., Riley G. R., Nellissery J. K., Allen T. E., Kable M. L., Wachal M. D., Seiwert S. D., Myler P. J., Stuart K. D. 1996. Complexes from Trypanosoma brucei that exhibit deletion editing and other editing-associated properties. Mol. Cell. Biol. **16**(4):1410-8.

Cruz-Reyes J., Piller K. J., Rusché L. N., Mukherjee M., Sollner-Webb B. 1998a. Unexpected electrophoretic migration of RNA with different 3' termini causes a RNA sizing ambiguity that can be resolved using nuclease P1- generated sequencing ladders. Biochemistry **37**(17):6059-64.

Cruz-Reyes J., Rusché L. N., Piller K. J., Sollner-Webb B. 1998b. T. brucei RNA editing: adenosine nucleotides inversely affect U-deletion and U-insertion reactions at mRNA cleavage. Mol Cell **1**(3):401-9.

Cruz-Reyes J., Sollner-Webb B. 1996. Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. Proc. Natl. Acad. Sci. USA **93**(17):8901-6.

Estevez A. M., Simpson L. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria--a review. Gene **240**(2):247-60.

Feagin J. E., Abraham J. M., Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. Cell **53**(3):413-22.

Feagin J. E., Jasmer D. P., Stuart K. 1987. Developmentally regulated addition of nucleotides within apocytochrome b transcripts in Trypanosoma brucei. Cell **49**(3):337-45.

Göringer H. U., Koslowsky D. J., Morales T. H., Stuart K. 1994. The formation of mitochondrial ribonucleoprotein complexes involving guide RNA molecules in Trypanosoma brucei. Proc. Natl. Acad. Sci. USA **91**(5):1776-80.

Hajduk S. L., Sabitini R. S. 1998. Mitochondrial mRNA Editing in Kinetoplastid Protozoa. In: Grosjean H., Benne R., editors. Modification and Editing of RNA. Washington, D.C.: ASM Press. p 377-393.

Hertzberg R. P., Dervan P. B. 1984. Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses. Biochemistry **23**(17):3934-45.

Igo R. P., Jr., Palazzo S. S., Burgess M. L., Panigrahi A. K., Stuart K. 2000. Uridylate addition and RNA ligation contribute to the specificity of kinetoplastid insertion RNA editing [In Process Citation]. Mol. Cell. Biol. **20**(22):8447-57.

Kable M. L., Seiwert S. D., Heidmann S., Stuart K. 1996. RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA [see comments] [published erratum appears in Science 1996 Oct 4;274(5284):21]. Science **273**(5279):1189-95.

Kapushoc S. T., Simpson L. 1999. In vitro uridine insertion RNA editing mediated by cis-acting guide RNAs. RNA **5**(5):656-69.

Kean J. M., White S. A., Draper D. E. 1985. Detection of high-affinity intercalator sites in a ribosomal RNA fragment by the affinity cleavage intercalator methidiumpropyl-EDTA- iron(II). Biochemistry **24**(19):5062-70.

Köller J., Norskau G., Paul A. S., Stuart K., Göringer H. U. 1994. Different Trypanosoma brucei guide RNA molecules associate with an identical complement of mitochondrial proteins in vitro. Nucleic Acids Res. **22**(11):1988-95.

Koslowsky D. J., Kutas S. M., Stuart K. 1996. Distinct differences in the requirements for ribonucleoprotein complex formation on differentially regulated pre-edited mRNAs in Trypanosoma brucei. Mol. Biochem. Parasitol. **80**(1):1-14.

Leung S. S., Koslowsky D. J. 1999. Mapping contacts between gRNA and mRNA in trypanosome RNA editing. Nucleic Acids Res. **27**(3):778-87.

Leung S. S., Koslowsky D. J. 2001. RNA editing in Trypanosoma brucei: characterization of gRNA U-tail interactions with partially edited mRNA substrates. Nucleic Acids Res. **29**(3).

Mathews D. H., Sabina J., Zuker M., Turner D. H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. **288**(5):911-40.

Matzura O., Wennborg A. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. Comput Appl Biosci **12**(3):247-9.

Milligan J. F., Groebe D. R., Witherell G. W., Uhlenbeck O. C. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res. **15**(21):8783-98.

Piller K. J., Decker C. J., Rusché L. N., Harris M. E., Hajduk S. L., Sollner-Webb B. 1995. Editing domains of Trypanosoma brucei mitochondrial RNAs identified by secondary structure. Mol. Cell. Biol. **15**(6):2916-24.

Piller K. J., Rusché L. N., Cruz-Reyes J., Sollner-Webb B. 1997. Resolution of the RNA editing gRNA-directed endonuclease from two other endonucleases of Trypanosoma brucei mitochondria. RNA **3**(3):279-90.

Pollard V. W., Harris M. E., Hajduk S. L. 1992. Native mRNA editing complexes from Trypanosoma brucei mitochondria. EMBO J. **11**(12):4429-38.

Read L. K., Göringer H. U., Stuart K. 1994. Assembly of mitochondrial ribonucleoprotein complexes involves specific guide RNA (gRNA)-binding proteins and gRNA domains but does not require preedited mRNA. Mol. Cell. Biol. **14**(4):2629-39.

Riley G. R., Corell R. A., Stuart K. 1994. Multiple guide RNAs for identical editing of Trypanosoma brucei apocytochrome b mRNA have an unusual minicircle location and are developmentally regulated. J. Biol. Chem. **269**(8):6101-8.

Rusché L. N., Cruz-Reyes J., Piller K. J., Sollner-Webb B. 1997. Purification of a functional enzymatic editing complex from Trypanosoma brucei mitochondria. EMBO J. **16**(13):4069-81.

Schmid B., Riley G. R., Stuart K., Göringer H. U. 1995. The secondary structure of guide RNA molecules from Trypanosoma brucei. Nucleic Acids Res. **23**(16):3093-102.

Schultz P. G., Dervan P. B. 1983. Sequence-specific double-strand cleavage of DNA by penta-N- methylpyrrolecarboxamide-EDTA X Fe(II). Proc. Natl. Acad. Sci. USA **80**(22):6834-7.

Seiwert S. D., Heidmann S., Stuart K. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. Cell **84**(6):831-41.

Stuart K., Allen T. E., Heidmann S., Seiwert S. D. 1997. RNA editing in kinetoplastid protozoa. Microbiol Mol Biol Rev **61**(1):105-20.

Wower J., Aymie M., Hixson S. S., Zimmermann R. A. 1989. Photochemical labeling of bovine pancreatic ribonuclease A with 8-azidoadenosine 3',5'-bisphosphate. Biochemistry **28**(4):1563-7.

CHAPTER 4

SUMMARY AND FUTURE PROSPECTS

SUMMARY

The process of kinetoplastid mitochondrial RNA editing requires the coordination of proteins, gRNAs and mRNAs in a ribonucleoprotein complex termed the editosome. Therefore, editosome function more than likely requires the interaction of the editing proteins with both RNAs. However, a survey of gRNAs and mRNAs has yet to reveal a common sequence motif that could act as a recognition site for the editing complex. This led to our overall hypothesis that the editing complex is able to assemble on hundreds of gRNA/mRNA pairs by interacting with the RNAs through recognition domains provided through a common structural core. In this thesis I describe three studies that support a common structure for interacting gRNAs and mRNAs, an integral part of our overall hypothesis.

To begin the search for structural elements, the 5' and 3' ends of three gRNAs were mapped along their cognate mRNAs. We began here as the 5' end of the gRNA was involved in anchor duplex formation with the mRNA and the 3' U-tail was hypothesized to interact with upstream purine sequences (Blum and Simpson, 1990; Seiwert et al., 1996). Using photoaffinity crosslinkers, the 5' and 3' ends of the three gRNAs were mapped. The 5' crosslinks confirmed that the anchor duplex formed as predicted while the 3' crosslinks provided the first direct evidence that the U-tail interacted with upstream purines near the initial editing site. The 3' crosslinks were in agreement with the hypotheses that the U-tail could be involved in stabilizing the gRNA/mRNA interaction and could function to hold on to the 5' mRNA cleavage fragment to prevent its loss from the active site of the editosome. Using this crosslinking data, it was exciting to see that similar secondary structures were predicted for the three gRNA/mRNA pairs. The

common structural elements include a gRNA/mRNA anchor duplex, a U-tail/mRNA duplex and a gRNA stem-loop. From the predicted structure it appears that with the two duplexes flanking the initial editing site, any secondary structure in the mRNA in this region is removed. This could provide the editosome access to the proper editing site.

The U-tail/mRNA duplex was particularly intriguing as the U-tail was predicted to basepair with mRNA sequences within the editing domain. Using gCYb-558 modified with a photoaffinity crosslinker at the 3' end and partially edited 5'CYb substrates, the U-tail interaction was examined as editing proceeded from ES1 to ES3. Remarkably, the 3' end of the U-tail basepaired with the same sequence despite the insertion of 6 U's, that doubling the length of the anchor duplex. Structure predictions using this crosslinking data indicate that the gRNA stem-loop is maintained by the U-tail as editing proceeds enabling the 3' end of the U-tail to interact with the same sequence. This function of the U-tail has not been previously proposed and the maintenance of the gRNA stem-loop strongly suggests that this predicted structural element is an important feature of interacting gRNAs and mRNAs.

To provide additional support for the predicted secondary structure of gRNA/mRNA pairs, 5' crosslinks were structure probed. The crosslinks in these molecules preserved the gRNA/mRNA anchor duplex allowing for further analysis of the U-tail/mRNA interaction. A comparison of crosslinks containing gRNAs with or without the U-tail directly demonstrated the presence of a U-tail/mRNA duplex. Only in the presence of the U-tail was protection of the uUtSE from nuclease activity observed. Furthermore, increased cleavage of the uUtSE by the intercalator MPE-Fe(II), correlated with the presence of the U-tail. Therefore, the enzymatic and chemical probing data

support the presence of the U-tail/mRNA duplex, in agreement with our predicted structure of gRNA/mRNA pairs.

The initial crosslinking studies led to three predicted structural elements: the gRNA/mRNA anchor duplex, the U-tail/mRNA duplex and the gRNA stem-loop. Examining the U-tail interaction as editing proceeds provided additional structural predictions that supported the presence of the gRNA stem-loop. Finally, structural probing directly established the existence of a U-tail/mRNA duplex. Therefore, the data from all three chapters support a common structure for interacting gRNAs and mRNAs.

Whether these structures are essential for interaction with the editosome remains to be determined. However, it has been established that both the gRNA/mRNA anchor duplex and U-tail/mRNA duplex play critical roles in the editing process (Burgess et al., 1999; Kapushoc and Simpson, 1999; Seiwert et al., 1996). In addition both 5' and 3' crosslinked gRNA/mRNA pairs are accurately cleaved by the gRNA directed endoribonuclease. The position of the 5' covalent link (gRNA orientation) between gRNA and mRNA also enabled observation of both TUTase and U-specific exonuclease activity on the resulting 5' cleavage mRNA fragment. All three of these enzymatic activities associate with the editosome, indicating that the crosslinked RNAs interact correctly with the editing complex. Therefore, the structural elements of the crosslinks are ideal candidates for recognition domains of the editing complex and warrant further investigation.

The function of the gRNA U-tail has not been clearly defined, despite considerable controversy over its role in the process of editing. Using crosslinking and structure probing techniques to investigate the interaction of the U-tail with the mRNA,

direct evidence was obtained that demonstrated the U-tail did indeed basepair with upstream purine rich mRNA sequences. As a result the U-tail was capable of acting as an upstream anchor and a tether for the 5' fragment created by endoribonuclease cleavage during editing. Experiments in our lab are underway to characterize the contribution of the U-tail to the association of gRNAs and mRNAs, and the stabilization of this interaction.

Interestingly, by studying the U-tail/mRNA interaction as editing proceeds, a potentially new role for the U-tail has been revealed. The successive insertion of U's into CYb mRNA, extends the anchor duplex using nucleotides involved in the gRNA stem-loop. The U-tail is able to maintain the stem-loop by feeding into the structure and basepairing with the guiding sequence of the gRNA.

These multiple roles may provide an explanation for why the gRNA has a U-tail. Upstream mRNA sequences are purine rich as editing most often directs the insertions of uridines. Therefore a U-tail is able to basepair with upstream mRNA sequences to act as an anchor and tether. The same holds true for the guiding sequence of the gRNA, it is also purine rich as it is the complement of fully edited mRNA and directs the insertion of U's. Thus, as editing proceeds and the role of the U-tail changes, the U-tail is able to bind to the information sequence of the gRNA to maintain a stem-loop in the gRNA. Taken together, this suggests that due to the ability of U's to basepair with both purines (A and G), a poly-U sequence is the best universal sequence to carry out all of these tasks.

FUTURE PROSPECTS

Short-term objectives

The presence of the gRNA stem-loop in the predicted secondary structure of gRNA/mRNA pairs has not been established. Interest in this element arises from its similarity to stem-loop structures defined by structure probing in free gRNAs and the ability of gBP21 to bind to this structure. To determine if the gRNA stem-loop is present in gRNA/mRNA pairs, structure probing of 5' crosslinked substrates could be performed. The methodology used to create a gRNA with a U_{15}-tail involves phosphorylating the 5' end of the U_{15} RNA oligonucleotide to allow it to be ligated to the rest of the gRNA. End-Labeling the U_{15} RNA oligonucleotide to a high specific activity would enable structure probing of the gRNA. No data for the U-tail interaction could be collected but sequences upstream of the U-tail (specifically the labeled phosphate) could be examined.

If structure probing supports a gRNA stem-loop it would be informative to analyze the relationship between the gRNA stem-loop and the U-tail/mRNA duplex. The U-tail interacts close to the initial editing site despite upstream sequences that would provide a more stable interaction. The question of whether the stem-loop affects the position of the U-tail interaction could be examined using mutations within the gRNA that disrupt the stem-loop as well as compensatory mutations. The U-tail interaction could be followed using standard crosslinking techniques.

Furthermore, if a gRNA stem-loop is present within a gRNA/mRNA complex, it would be logical to determine whether gBP21 is able to bind to this stem-loop. gBP21 has been shown to bind a stem-loop in free gRNAs suggesting that one of its roles involves stabilization of this structure. A similar role in gRNA/mRNA complexes could

suggest that gBP21 affects the stem-loop stability and thus affect the position of the U-tail/mRNA interaction. gBP21 may also function to anneal gRNAs and mRNAs together overcoming unfavorable conditions as observed for gCYb-558 and 5'CYbUT (personal communication, Uli Göringer). However, it is not known whether gBP21 remains bound once the two RNAs come together. One method that could be used to determine if gBP21 bound to crosslinks is a label-transfer technique. This technique was originally used to identify gBP21 along with several other proteins that bound specifically to gRNAs (Köller et al., 1994; Leegwater et al., 1995). This would involve crosslinking gRNA transcribed as a high specific activity probe (a riboprobe) and unlabeled mRNA. These crosslinks would be recovered and incubated in the presence of excess gBP21 (footprinting of gND7-506 was achieved using 10 molar excess of gBP21) (Hermann et al., 1997). After incubation the reaction would be exposed to UV irradiation (254 nm) followed by ribonuclease digestion. Any gBP21 bound to the gRNA would be crosslinked to the gRNA during UV irradiation. Crosslinked protein would protect a region of the gRNA from ribonuclease activity, hence tagging the protein with a small labeled RNA fragment. This could be visualized using SDS-PAGE. One would be looking for an approximately 21 kDa (depending on the size of the RNA fragment) radioactive band. The identity of the band should be confirmed via a western blot using an antibody to gBP21. The position of the RNA-protein interaction could be further narrowed down by site specific incorporation of labeled nucleotides into the gRNA and using the label transfer technique.

Long-term directions

The presence of three helices (the anchor duplex, U-tail/mRNA duplex and the stem of the predicted stem-loop in the gRNA) indicates that there is the potential for helical stacking. Helical stacking is a common strategy for RNAs to form tertiary structures. The gRNA/mRNA tertiary structure could be the core architecture that the editing complex recognizes and binds. Tertiary interactions could be investigated using standard site-specific 4-thiouridine crosslinking techniques. Due to geometry 4-thiouridine does not theoretically crosslink to bases it is basepaired with (Dubreuil et al., 1991). Therefore 4-thiouridine crosslinks are the result of tertiary interactions. Three-dimensional modeling of the two interacting RNAs could reveal sites ideal for RNA-protein binding.

Using the information from the above study, one could then incorporate 4-thiouridines into promising sites for RNA-protein interactions such as exposed and accessible structures. The thiol group of 4-thiouridine can act as a nucleophile enabling the attachment of a longer linker with a photoaffinity group. This would enable one to crosslink proteins interacting with this region of the RNA. The RNA-protein interactions identified in this manner could be further examined by disrupting the structure of this region, to determine if this inhibited protein binding as well as whether this affected editing activities. This method could help to elucidate the important structures of gRNA/mRNA pairs required for active editosome assembly.

These studies together with the identification of proteins involved in editing will provide a more complete understanding of the process of mitochondrial RNA editing in kinetoplastids.

REFERENCES

Blum B., Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. Cell **62**(2):391-7.

Burgess M. L., Heidmann S., Stuart K. 1999. Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit in vitro U insertion. RNA **5**(7):883-92.

Dubreuil Y. L., Expert-Bezancon A., Favre A. 1991. Conformation and structural fluctuations of a 218 nucleotide long rRNA fragment: 4-thiouridine as an intrinsic photolabelling probe. Nucleic Acids Res. **19**:3653-3660.

Hermann T., Schmid B., Heumann H., Göringer H. U. 1997. A three-dimensional working model for a guide RNA from Trypanosoma brucei. Nucleic Acids Res. **25**(12):2311-8.

Kapushoc S. T., Simpson L. 1999. In vitro uridine insertion RNA editing mediated by cis-acting guide RNAs. RNA **5**(5):656-69.

Köller J., Norskau G., Paul A. S., Stuart K., Göringer H. U. 1994. Different Trypanosoma brucei guide RNA molecules associate with an identical complement of mitochondrial proteins in vitro. Nucleic Acids Res. **22**(11):1988-95.

Leegwater P., Speijer D., Benne R. 1995. Identification by UV cross-linking of oligo(U)-binding proteins in mitochondria of the insect trypanosomatid Crithidia fasciculata. Eur. J. Biochem. **227**(3):780-6.

Seiwert S. D., Heidmann S., Stuart K. 1996. Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. Cell **84**(6):831-41.