

**CHAPTER 1 “NMR STUDIES OF THE SPLICED LEADER RNA  
FROM *CRITHIDIA FASCICULATA* AND *LEPTOMONAS COLLOSOMA*”**

## 1.1 Summary

This chapter presents nuclear magnetic resonance (NMR) studies on the Spliced Leader RNA (SL RNA) from two species of trypanosomes, *Leptomonas collosoma* and *Crithidia fasciculata*. Previous studies showed that the 5' half of the SL RNAs could possibly adopt two secondary structures, denoted form 1 and form 2 (LeCuyer and Crothers, 1993). Using NMR techniques, the *in vitro* secondary structure of the 5' half of both SL RNAs was determined. The *L. collosoma* was found to exist in the form 1 structure as previously proposed (LeCuyer and Crothers, 1993; Cload, et al., 1993). The *C. fasciculata* SL RNA was found to be in the form 2 structure, a surprising result given the high degree of sequence homology between the two RNAs.

The form 1 hairpin of the *L. collosoma* SL RNA was further examined by synthesizing smaller RNA fragments of the parent 55 nucleotide (nt) molecule. A twenty-five nt and a thirteen nt hairpin were studied. They demonstrate a remarkable feature of this class of RNA, the existence of a tri-uridine hairpin loop in which the first and third uridine are basepaired. Furthermore, two possibly monomeric conformations of the form 1 hairpin were found to exist and can be studied independently by appropriately adjusting the buffer salt conditions.

## 1.2 Introduction and background

The SL RNAs are found in a variety of lower eukaryotic organisms such as nematodes, euglena and trypanosomes. The two parent SL RNA sequences studied in this chapter are derived from two species of trypanosomes. These organisms are dangerous human pathogens that have an interesting molecular biology.

### 1.2.1 Trypanosome biology

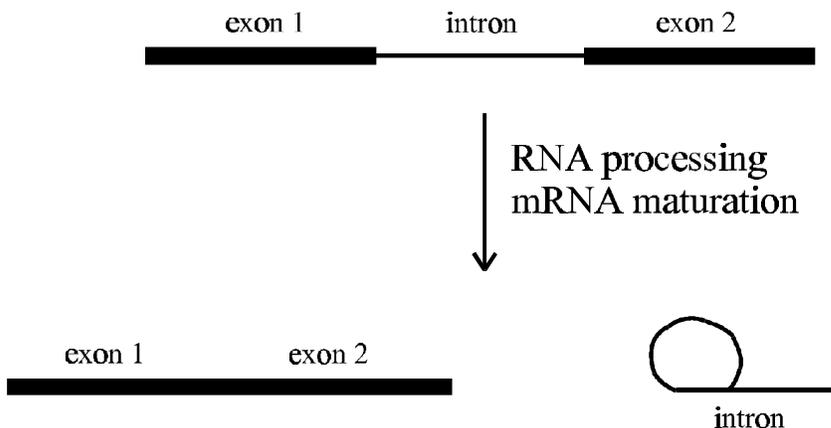
The trypanosomes are flagellated protozoans of the order *kinetoplastida*. The order is so named for the distinctive large mitochondrial kinetoplast found inside each organism. The trypanosomatids include monogenetic insect parasites (such as *Leptomonas* and *Crithidia*, among others) and digenetic parasites that cycle between insects and plants (*Phytomonas*) or insects and vertebrates (*Trypanosoma*, *Leishmania* and *Endotrypanum*).

The trypanosomes are important human pathogens. For example, transmitted via the bite of reduviid ("kissing") bugs in South America, *T. cruzi*, causes Chaga's disease, in which the invading trypanosome burrows into the heart muscle of the victim. Another trypanosome, *T. brucei gambiense*, is transmitted by the tsetse fly and causes trypanosomiasis (or sleeping sickness), in which the parasite develops in the bloodstream and eventually enters the nervous system. These pathogens can have very severe epidemic consequences; over 4 million people from the African country of Uganda alone were killed by an outbreak of trypanosomiasis that occurred in 1904.

### 1.2.2 Pre-mRNA processing in trypanosomes

Aside from their interesting pathogenic properties, trypanosomes also exhibit a unique molecular biology. Unlike other eukaryotes, trypanosomal genes that encode for nuclear proteins lack internal introns, rather, they are excised from polycistronic transcription units solely by *trans*-splicing to the SL RNA and poly-adenylation (Clayton, 1992; Ullu et al., 1995). This *trans*-splicing was first discovered in trypanosomes and has subsequently been found to occur in some of the lower eukaryotes such as

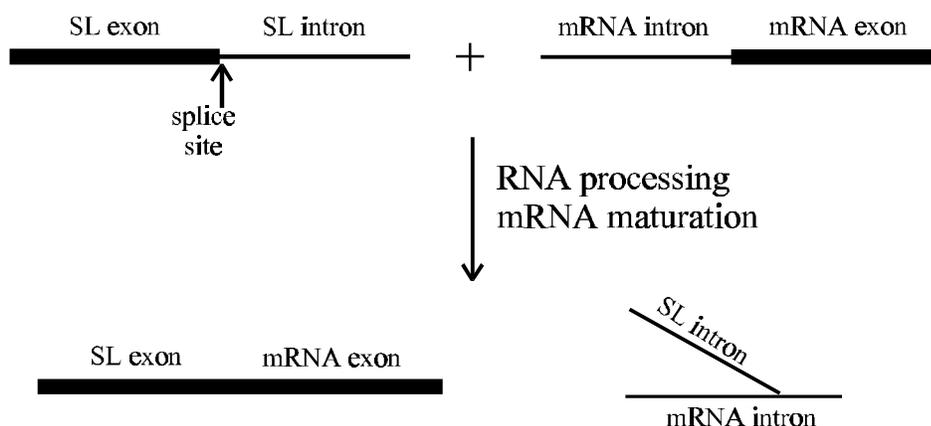
nematodes, euglena and trematodes (Kraus and Hirsh, 1987; Blumenthal and Thomas, 1988) as well.



**Figure 1. 1 mRNA processing by *cis*-splicing**

The molecular mechanism for *trans*-splicing is analogous to that of *cis*-splicing. Most eukaryotic cells, including mammalian, excise introns from mRNA via a *cis*-splicing mechanism to produce a mature mRNA message ready for translation (Fig 1.1) (Padgett et al., 1986; Maniatis and Reed, 1987). In this splicing scheme, the snRNP associates with the intron-exon boundaries of the mRNA and a branch point in the intron. The intron is looped out of the mRNA via a series of *trans*-esterification reactions and the two splice sites are ligated together (Gutherie, 1991). The substrate for this reaction is a single RNA molecule and excision proceeds to give a mature mRNA molecule in which the intron has been removed. An RNA "lariat" is formed when the 5' splice site phosphate is ligated to the 2' hydroxyl of the branch point nucleotide, which leaves a 3' hydroxyl at the 5' splice site nucleotide. The phosphate of the 3' end of the intron is then ligated to this 3' hydroxyl of the 5' exon, excising the intron RNA in the "lariat" shape.

In the *trans*-splicing reaction (Fig 1.2), a spliced leader (SL) exon is joined to the 5' end of a mRNA coding region on a separate transcript (Konarska, et al., 1985; Murphy, et al., 1986; Sutton and Boothroyd, 1986). The SL exon (20-35 nts) is derived from the full length SL RNA (130-220 nts), which exists as a small ribonucleoprotein particle (or snRNP) (Michaeli, et al., 1990; Cross, et al., 1991). The products of the reaction are a mature mRNA transcript "capped" by the SL RNA exon and a Y-branched mRNA/SL RNA intron molecule.

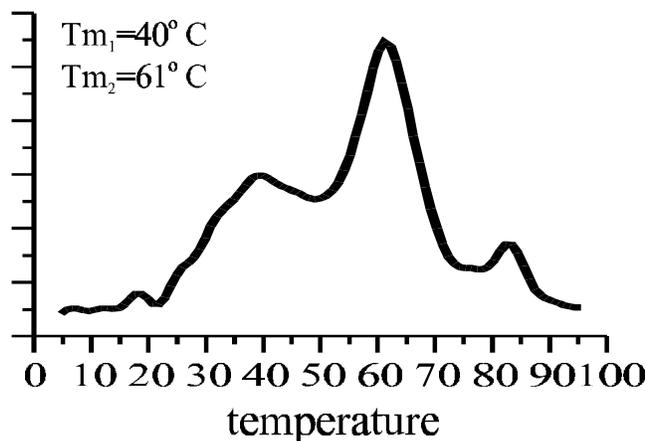


**Figure 1. 2 mRNA processing by *trans* splicing**

It is not well understood what is the functional role of the *trans*-splicing mechanism in the processing of typanosomal mRNA, or what role the post-spliced SL RNA exon plays. Since the same SL exon is spliced onto the 5' end of all mRNAs, it is speculated that the SL RNA could function to protect the mature transcript from degradation or is involved in signaling the cell to transport the message out of the nucleus. The argument that the SL RNA protects the mRNA from degradation is supported by the fact that the 5' end of SL RNA contains a number of methylated nucleotides,  ${}^7\text{Gpppm}_2{}^6\text{A}(2'\text{Om})\text{A}(2'\text{Om})\text{C}(2'\text{Om})\text{m}^3\text{U}(2'\text{Om})$  (Perry, et al., 1987;







**Figure 1. 5 Derivative UV melting curve of the 60 nt 5' half of the *L collosoma* SL RNA**

Sequence analysis of the SL RNAs of other trypanosomes (Fig 1.6) shows that this ability to adopt two secondary structures may be a common feature. If the ability to adopt two alternate secondary structures is a common feature of all trypanosomal SL RNAs; this raises the question of why. Steitz (1992) has proposed a model of *trans*-splicing that incorporates components of both structure models in which the structural switch between form 1 and 2 mimics the functions of the U1 and U5 RNAs found in *cis*-splicing. Also, it has been noted (LeCuyer and Crothers, 1993) that the SL exon is extensively basepaired to the intron while in form 2, but there is virtually no base pairing in form 1. Thus, form 1 may be a method of disrupting interactions between the SL RNA intron and a mRNA.

<u>AACUAA</u> AACA <u>UUUUUGAAGAA</u> <u>CAGUUUCUGUACU</u> U <u>CAUUGGUAU</u> GGUAUGUAGAG <u>CUUC</u>	<i>L. collosoma</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> A <u>UUAUUGGUAU</u> GAGAAG----- <u>CU</u>	<i>T. brucei</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> A <u>UUAUUGGUAU</u> GCGAAG----- <u>CUU</u>	<i>T. cruzi</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> A <u>UUAUUGGUAU</u> GCGAAG----- <u>CUUC</u>	<i>T. rangeli</i>
<u>AACUAA</u> AGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> A <u>UUAUUGGUAU</u> GAGAAG----- <u>CU</u>	<i>T. malayi</i>
<u>AACUAA</u> AGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> A <u>UUAUUGGUAU</u> GAGAAG----- <u>CU</u>	<i>T. vivax</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> U <u>UUAUUGGUAU</u> GCGAAA----- <u>CUU</u>	<i>L. enreittii</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> U <u>UUAUUGGUAU</u> GCGAAA----- <u>CUUC</u>	<i>L. mexicana</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> U <u>UUAUUGGUAU</u> GCGAAA----- <u>CUUC</u>	<i>L. donovani</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> U <u>UUAUUGGUAU</u> GAGAAG----- <u>CUUC</u>	<i>L. seymouri</i>
<u>AACUAA</u> CGCU <u>UUUUUAGAA</u> -- <u>CAGUUUCUGUACU</u> U <u>UUAUUGGUAU</u> AAGAAG----- <u>CUUC</u>	<i>C. fasciculata</i>

**Figure 1. 6** Sequence analysis of trypanosomal SL RNAs

Interestingly, *in vivo* analysis of the *L. collosoma* and *T. brucei* SL RNAs using water-soluble chemical modification probes has shown that the form 2 structure predominates (Harris, et al., 1995). Thus, in the context of the snRNP, form 2 seems to be favored, while the *L. collosoma* RNA alone *in vitro* favors the form 1 structure. Further, it was shown that the methylated nucleotides on the 5' end of the SL RNA do not play a structural role *in vivo* (Harris, et al., 1995); however, the methyl groups are required for the *trans*-splicing reaction to occur (Ullu and Tschudi, 1991, 1993; McNally and Agabian, 1992).

#### 1.2.4 Project goals

The goal of this project is to investigate the structural features of the SL RNAs from two species of trypanosomes, *L. collosoma* and *C. fasciculata*, by NMR. Two main interests were pursued in the NMR investigations. The first was in determining the *in vitro* secondary structures of the SL RNAs derived from the two species. The second was in finding and characterizing smaller structural fragments derived from the parent RNAs, in the hope that these smaller fragments might prove to be structurally interesting and tractable by NMR methods.

In this report, NMR techniques were used to determine the *in vitro* secondary structure of the 5' halves of both SL RNAs. The secondary structure of the *L. collosoma* sequence was confirmed to be in the form 1 secondary structure as predicted by the previous biophysical studies. The SL RNA from *C. fasciculata* was found to be in the form 2. This result is interesting in that most of the SL RNAs have a closer sequence homology to the *C. fasciculata* SL RNA, and possibly the *L. collosoma* SL RNA is the only one found in the form 1 structure *in vitro*.

Further studies were carried out on the form 1 SL RNA from *L. collosoma*, including  $^{15}\text{N}/^{13}\text{C}$  isotope labeling the 55 nt 5' half and characterization of smaller fragments that contain only the form 1 hairpin. The form 1 hairpin was found to contain an unusual feature, a three-uridine loop, with the first and third uridine base paired. This is a surprising result in that it would require the hairpin loop to be spanned by a single nucleotide. Because of this, the possibility of dimerization of the RNA was investigated. The evidence favors that the U=U basepair is found in monomeric RNA, but more work needs to be done to prove that the molecularity is one.

The smallest fragment of the *L. collosoma* SL RNA studied, a thirteen-nucleotide hairpin, was found to exist in two conformations in slow exchange. The ratio of the concentrations of the two conformations was found to be a function of the ionic strength of the solution. Clearly, the possibility exists that the conformational change may be a monomer-dimer exchange, and both biophysical and NMR methods were utilized to investigate this possibility. Both of the two conformations were studied individually and characterized by NMR.

### 1.3 Results

Five NMR samples were synthesized for study. The names of the samples are derived from the species from where they came, and the length of the RNA. Thus, the sample "rLC55" is an RNA derived from the *L. collosoma* sequence and is 55 nts long. Figure 1.7 shows a complete listing of the samples, their names and the numbering scheme used in identifying the nucleotides

#### 1.3.1 Choice of *L. collosoma* experimental samples

The full length (130 nt) and the 5' half (52 nt) of the SL RNA from the species of trypanosome, *L. collosoma*, (Fig. 1.3 and 1.4) have been extensively studied by biophysical methods. These studies demonstrated that the 5' half of the RNA is structurally independent of the 3' half (LeCuyer and Crothers, 1993) and that the 5' half is still characterized by the biphasic UV melt (Fig. 1.5). Given that the "tertiary" structural elements exist on the 5' half of this SL RNA, the wild type 52 nt 5' half of the *L. collosoma* SL RNA has been selected for studies by NMR. An additional 3 guanine residues were added to the 5' end of the RNA to increase the yield on the transcription reactions, as has been suggested previously (Milligan, et al, 1987).

The parent rLC55 sample is interesting because it represents the SL RNA before the splicing event and spans the splice site. Also of interest is what structure the SL RNA exon (30-40 nts) will adopt after the *trans*-splicing event. The SL exon can only adopt the form 1 hairpin, because the form 2 base pairing occurs on the 3' side of the splice site (Figs 1.3 and 1.4). This form 1 hairpin is presumably the structural element that may be recognized by cellular machinery responsible for transport of the mature mRNA out of



the nucleus. For this reason, smaller 25 and 13 nt (rLC25 and rLC13) form 1 hairpins were constructed for more detailed spectroscopic study. While these are derived from the *L. collosoma* sequence, they are fairly well representative of all the trypanosome form 1 hairpins given the high level of sequence conservation in this region of the SL RNAs. These samples also have the advantage of being much smaller than the parent SL RNA molecules, which makes them better suited for high resolution NMR characterization.

### 1.3.2 Choice of *C. fasciculata* experimental samples

When analyzing the sequence homology between the known trypanosome SL RNA sequences (Fig. 1.6) it is clear that two major "sequence classes" exist, one representative of the *L. collosoma* sequence and one representative of all the other species of trypanosomal SL RNAs. For this reason, it seemed appropriate to investigate the SL RNA of the sequences in this latter class. Thus, a second sequence of the SL RNA was chosen for study, the 52 nt of the 5' half of the wild type SL RNA from *C. fasciculata*, rCF55. As with the rLC55 sample, three guanine nucleotides were added to the 5' end of this RNA to increase the transcription yield. While no biophysical studies have been performed on the RNA, it was inferred from the sequence homology with the *L. collosoma* SL RNA that the properties of the two sequences would be similar.

The *C. fasciculata* form 2 hairpin, rCF30, was found serendipitously during the construction of a "segmentally" labeled version of the rCF55 sample. This is discussed in greater detail in the next section of this chapter.

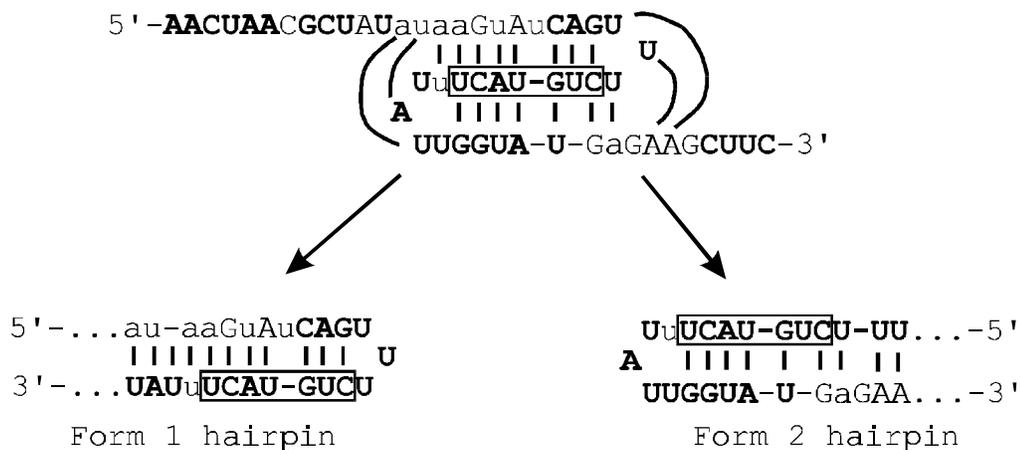
### 1.3.3 Determining the secondary structure of nucleic acids by NMR

Determination of the secondary structure of nucleic acids by NMR generally involves analysis of the solvent-exchangeable imino proton spectra. These experiments are conducted in H<sub>2</sub>O, where the imino is observable only when it is protected from fast exchange with bulk solvent, such as when it is involved in a hydrogen bond in a standard Watson-Crick base pair. Thus, the existence of an imino proton may indicate that there exists some form of a secondary structure for that region of the molecule.

Two NMR experiments are primarily used to analyze the imino protons. The two-dimensional (2D) <sup>1</sup>H-<sup>1</sup>H H<sub>2</sub>O NOESY experiment is used to give the connectivities from an imino to its nearest neighbors, possibly to the imino in the next basepair. The second experiment is the 2D <sup>1</sup>H-<sup>15</sup>N HMQC (Szewczak, et al., 1993). This experiment correlates the imino proton to the chemical shift of the nitrogen to which it is directly attached. This is important in the assignment of the imino protons since the nitrogen of the purines (guanine) and pyrimidines (uridine) have unique chemical shifts. Thus, the base-identity of each imino can be established based solely on the distribution of the <sup>15</sup>N chemical shifts.

This commonly used approach of analyzing the imino NOESY pattern to determine the secondary structure of nucleic acids failed to work for the SL RNAs studied. The problem lies in the fact that the information derived from the aforementioned NMR experiments is an imino proton pattern such as "GUUGU". If this pattern can exist in more than one region of the RNA, it is difficult to unambiguously make an assignment of the secondary structure. The SL RNAs can possibly adopt either the form 1 or form 2 secondary structures, as mentioned before, both of which share a

common stretch of the RNA as shown below (Fig. 1.8). For this reason, depending on what nucleotides are bulged out of the helix, the connectivities of the imino NOESY experiment could not uniquely identify one of the two possible secondary structures.



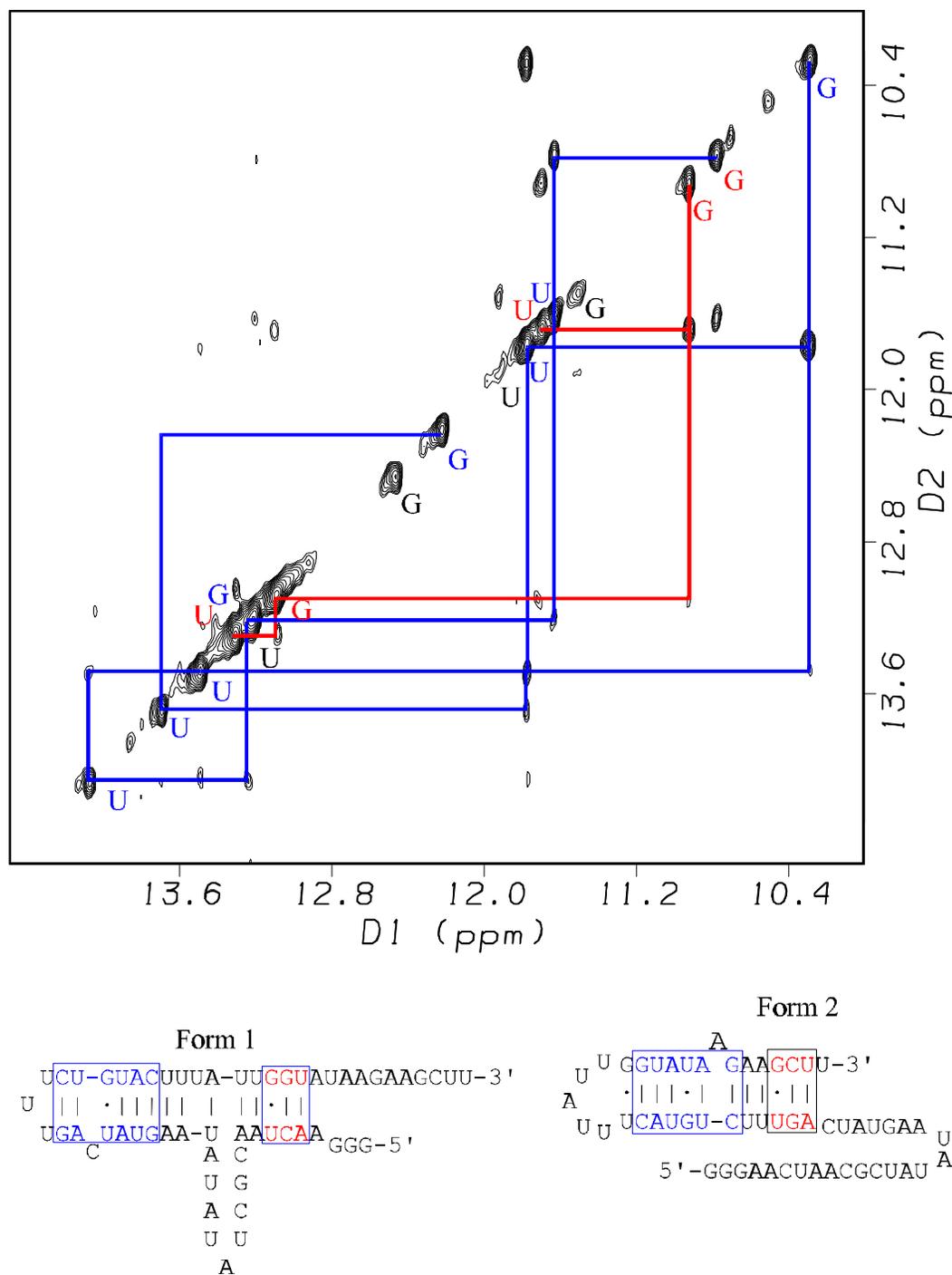
**Figure 1.8** Consensus “central core” nucleotides in form 1 and form 2

Because of this problem of assigning the iminos in the “central core” region of the SL RNAs, other methods were used to determine the secondary structures, such as comparison of the spectra of RNA fragments with that of the parent RNA.

#### 1.3.4 The in vitro secondary structure of the *C. fasciculata* SL RNA

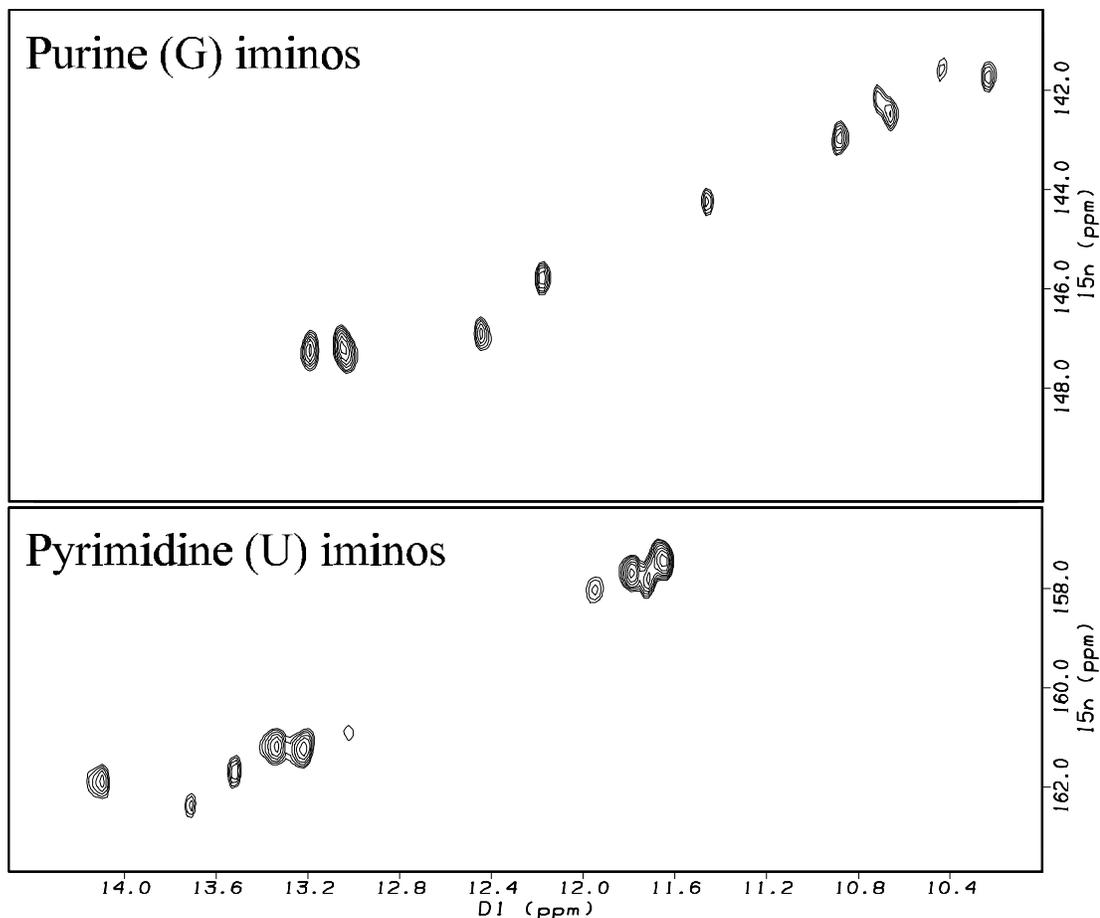
As was discussed for the *L. collosoma* SL RNA, the *C. fasciculata* SL RNA can adopt both the form 1 and form 2 secondary structure (Fig 1.9).





**Figure 1.10**  $H_2O$  NOESY spectrum of rCF55

The blue and red lines show the two main imino-imino crosspeak connectivity patterns from the JRSE  $H_2O$  NOESY experiment. Both the form 1 and the form 2 secondary structure of rCF55 could satisfy these imino crosspeak patterns. One cytosine must be bulged out of the form 1 helix, and one adenine must be bulged out of the form 2 helix to satisfy the connectivities. The experiment was performed at 25°C with a 250 ms mixing time.



**Figure 1.11** rCF55  $^1\text{H}$   $^{15}\text{N}$  HMQC

Assignment of the base identity of the imino protons shown in figure 1.10 for rLC55 was based on the imino nitrogen chemical shifts from this HMQC experiment. In total, 5 Watson-Crick base paired uridines, 4 Watson-Crick base paired guanines and 4 guanine-uridine wobble base pairs appear. The rCF55 sample was in 20 mM sodium phosphate buffer (pH 6.5), 150 mM sodium chloride and 1 mM EDTA. The data were collected at 25° C.

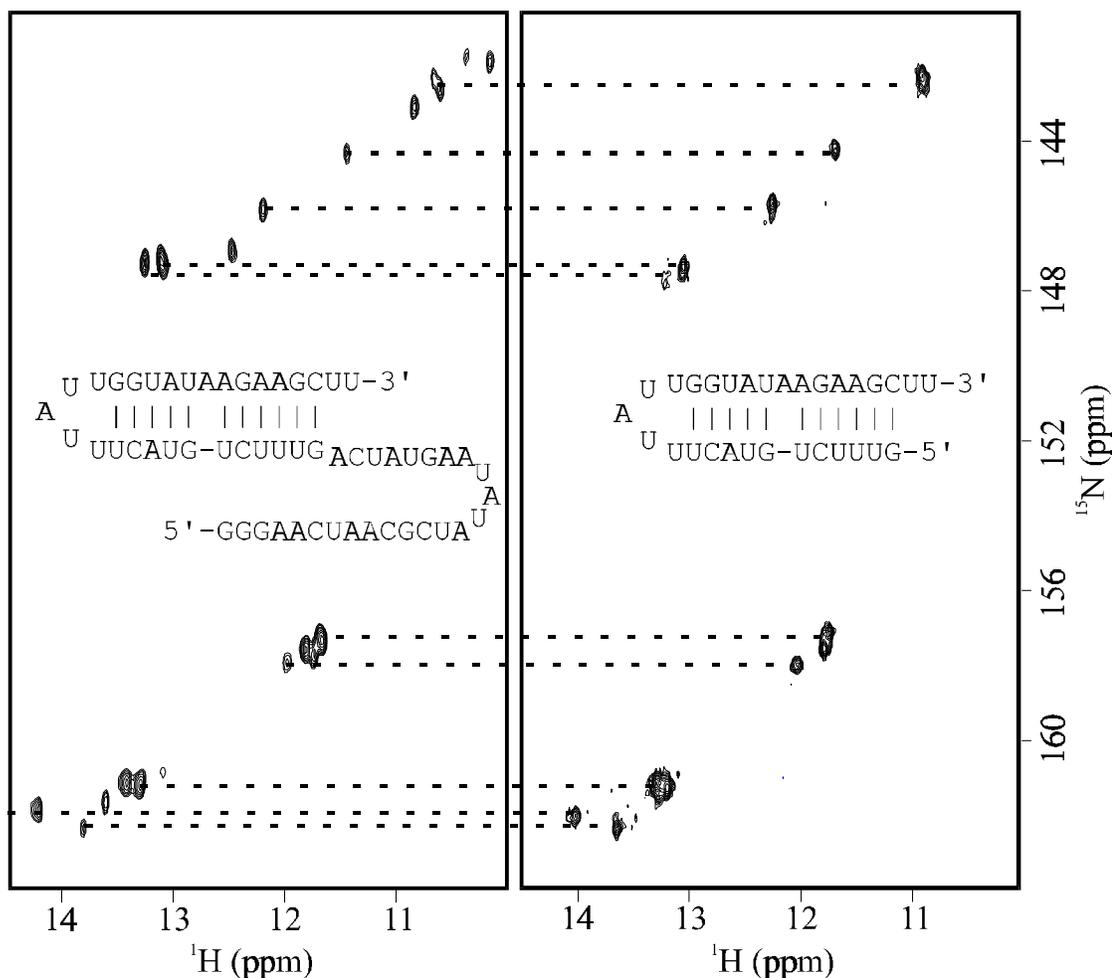
utilized and shown to be effective in our lab in the analysis of the secondary structure of the SL RNA from *C. elegans* (Xu, et al., 1996).

It became apparent that it would not be necessary to make the segmentally labeled rCF55 RNA during the process of analyzing one of the segments. The 3' end  $^{15}\text{N}$  labeled segment of the RNA, rCF30, contained the entire form 2 hairpin and showed nearly an identical 2D  $^1\text{H}$ - $^{15}\text{N}$  HMQC spectrum to that of the full length RNA. Figure 1.12 shows the comparison of the rCF55 and rCF30 HMQC data. A few iminos found in the rLC55 spectrum are absent in the rLC30 spectrum, the third G imino from the left, the third U imino from the left and one of the G=U base pairs.

Since there is considerable sequence homology between the *C. fasciculata* and the *L. collosoma* SL RNA (Fig. 1.6) we wanted to further confirm the hypothesis that the SL RNA of *C. fasciculata* exists in the form 2 structure *in vitro*. With that goal in mind, constant temperature native gel analysis (see materials and methods) of form 1 and form 2 mutants was performed. The results shown in figure 1.13 confirm that the wild type sequence runs with the same mobility as the form 2 mutant RNA. Oddly, the relative mobility of the form 1 and form 2 at 25° C appears to be the inverse of what is seen for the *L. collosoma* SL RNA (LeCuyer and Crothers, 1993). However, running the gel at 10° C inverts the relative mobilities of the form 1 and form 2 *C. fasciculata* SL RNAs.

### 1.3.5 The *in vitro* secondary structure of the *L. collosoma* SL RNA

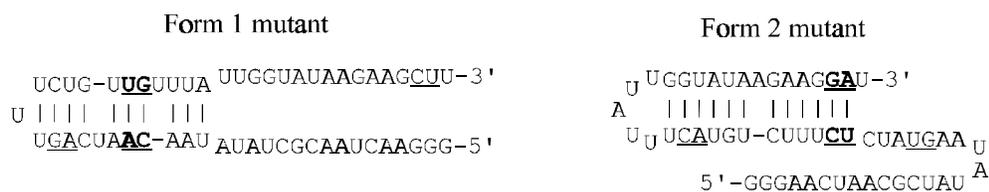
The secondary structure of the *L. collosoma* SL RNA has been well characterized and has been shown to exist *in vitro* as form 1 (LeCuyer and Crothers, 1993; Cload, et al., 1993). The buffer conditions of these studies was typically from pH 6 to 7.5 and between 50 to 200 mM sodium chloride. Choosing optimal NMR buffer conditions is important

A) WT *C. fasciculata* SL RNAB) Form 2 hairpin *C. fasciculata*

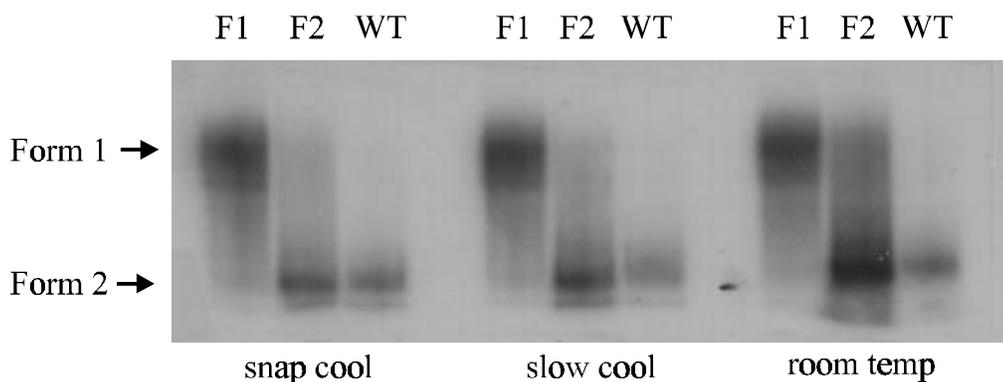
**Figure 1.12 HMQC comparison between rCF55 and form 2 hairpin**

A) The  $^1\text{H}$ - $^{15}\text{N}$  HMQC spectrum of the rCF55 sample and B) the spectrum of the rCF30 sample (form 2 hairpin). The buffer conditions were identical for both sample, 20 mM sodium phosphate (pH 6.5), 150 mM sodium chloride and 1 mM EDTA. The secondary structures of each sample are shown above. The comparison of the chemical shifts of both the imino protons and the nitrogens in these two spectra allowed for the unambiguous assignment of the *in vitro* secondary structure of the *C. fasciculata* SL RNA to the form 2. Both spectra were collected at 25°C.

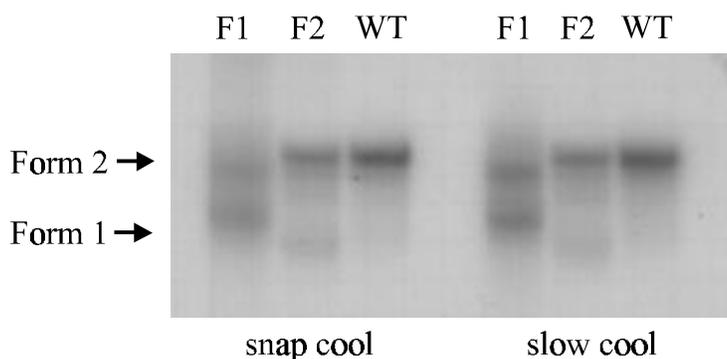
## A) form 1 and form 2 mutants



## B) Native gel 25° C



## C) Native gel 10° C



**Figure 1.13** *C. fasciculata* SL RNA form 1 and form 2 mutant native gels

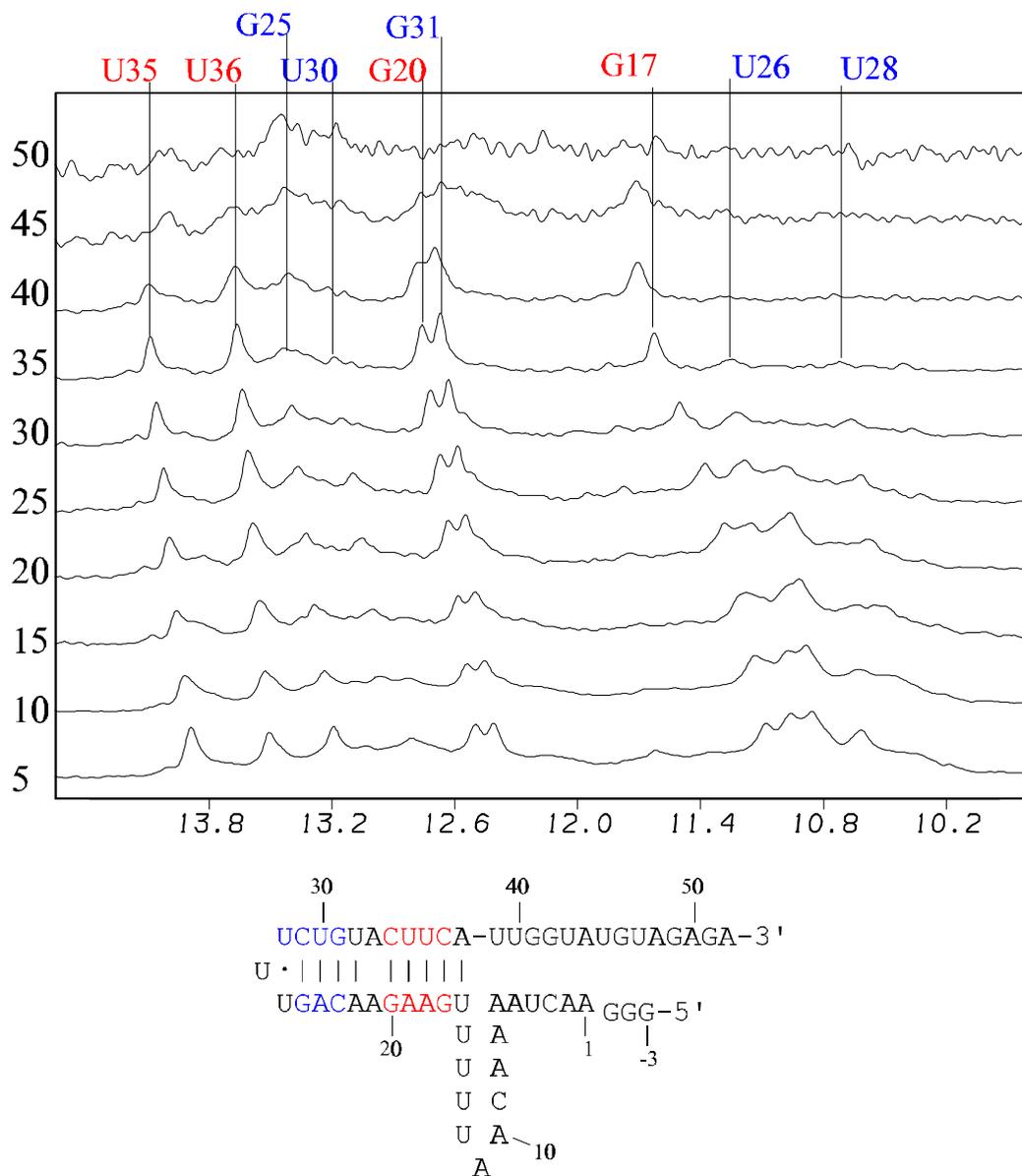
A) The form 1 and form 2 mutant sequences used in the native gel mobility study. The bold lettered regions of the sequence represent a position where the basepairs were inverted from their wild type positions. This should have no effect on the desired secondary structure, while inhibiting formation of the other structure. B) Room temperature native gel and C) 10°C native gel. The titles on the lanes represent either the form 1 or form 2 mutant sequence or the wild type rCF55 sequence. Samples were annealed using either snap cooling or slow cooling techniques (see materials and methods), with no effect on the results.

for obtaining well resolved (and meaningful) spectra. The buffer used in these NMR studies was chosen to be sodium phosphate, because it lacks any protons to interfere with the spectra and is slightly acidic, pH 6.5, to favor slower imino proton exchange. Two different sodium chloride salt concentration buffers were studied, a low salt (~30 mM [NaCl]) and a high salt (~130 mM [NaCl]).

The 1D imino proton temperature melt data are shown for both the low and high salt buffers (Figs. 1.14 and 1.15). A number of features can be seen in comparing the two experiments. The iminos in the low salt buffer begin to exchange broaden at approximately 45 degrees, while those in the high salt buffer are still strong at the same temperature. The iminos later identified to be G25, U30, U26 and U28 disappear in the low salt buffer at approximately 30 degrees, while they remain intense in the high salt buffer. The high salt buffer is clearly the better NMR candidate, at least in terms of the spectroscopy of the imino protons, and 130 mM [NaCl] was consequently chosen for further studies.

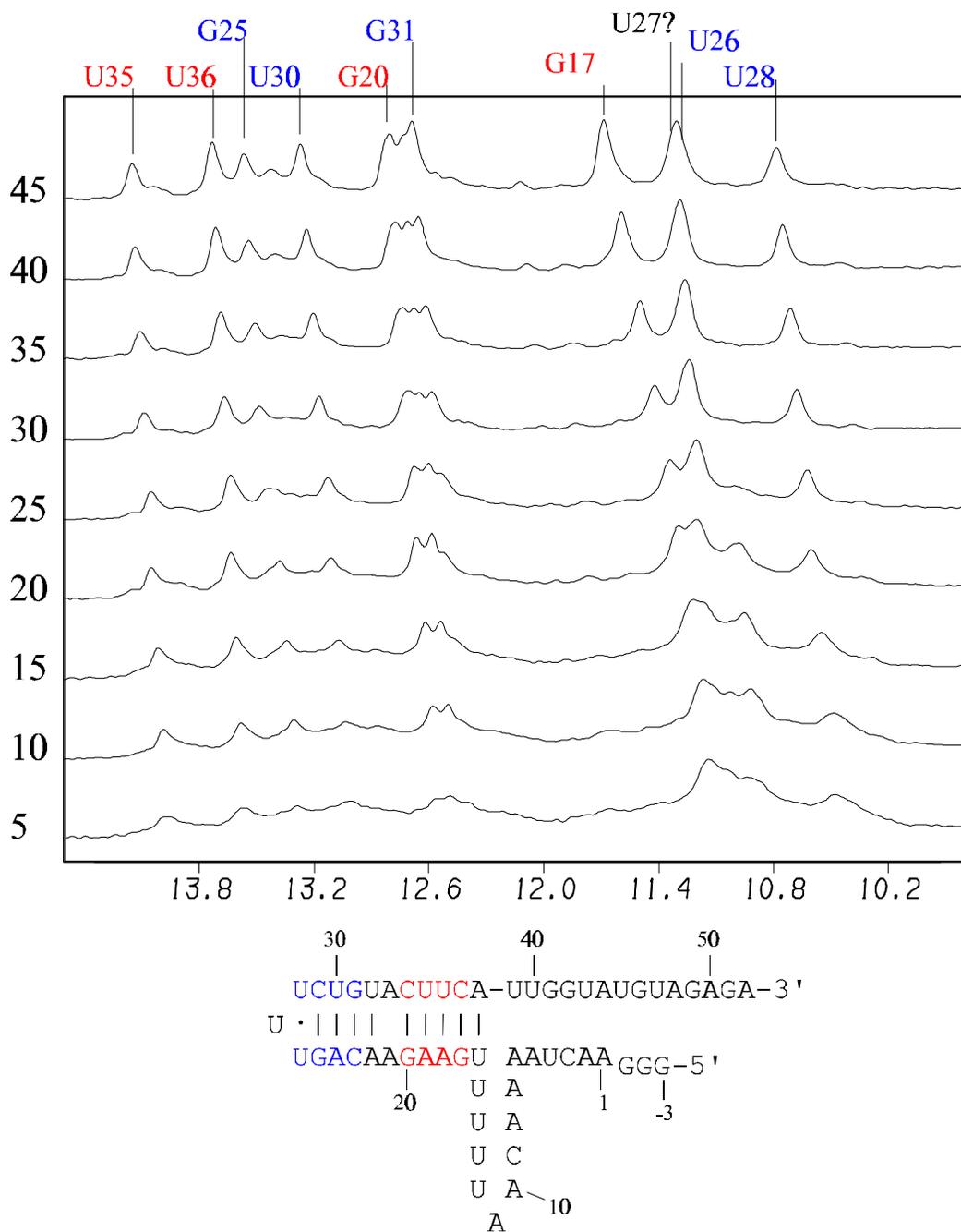
UV melts were performed on the rLC55 sample using same NMR buffer conditions (Fig. 1.16) to determine whether it displays a similar biphasic melting profile as seen by LeCuyer and Crothers. The concentration of the RNA was varied from 0.6  $\mu$  M to 120  $\mu$  M to look for signs of concentration dependent aggregation effects. The UV melts appear to be similar and the measured  $T_m$  for both transitions are the same at 40° and 61° C.

The 2D H<sub>2</sub>O NOESY (Fig 1.17) and the <sup>1</sup>H-<sup>15</sup>N HMQC (Fig. 1.18) spectra of rLC55 are in agreement with the assignment of the form 1 secondary structure, and the imino proton assignments are shown. Two main regions of imino proton connectivities



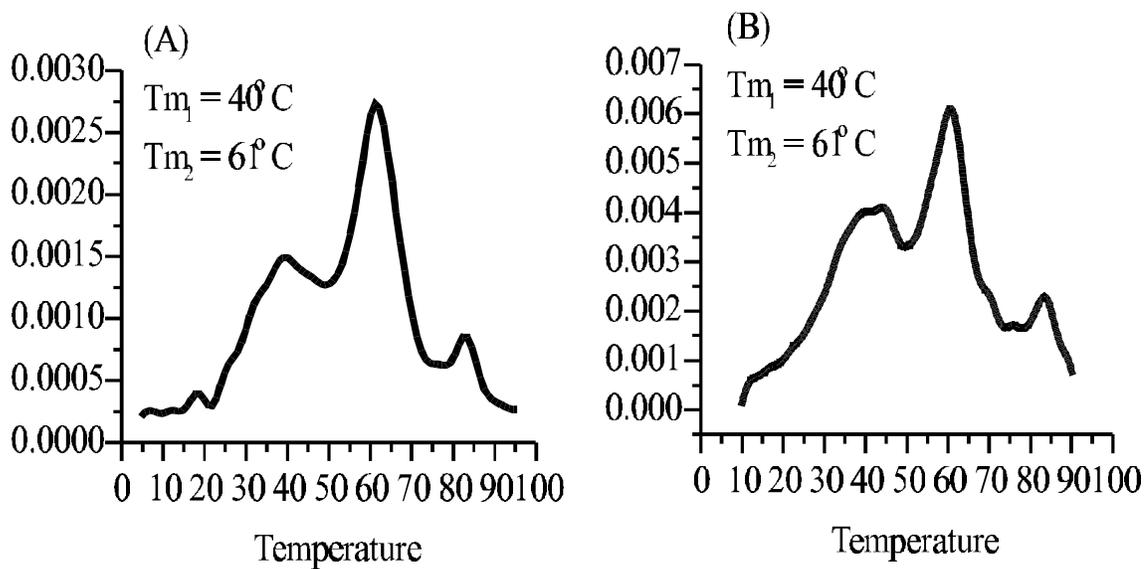
**Figure 1.14 Low salt rLC55 imino proton temperature study**

The low salt buffer conditions used in this imino proton temperature study were 10 mM sodium phosphate buffer (pH 6.5) and 1 mM EDTA, which is approximately 30 mM in sodium ions. The iminos G25, U30, G17, U26 and U28 appear broad and featureless.



**Figure 1.15 High salt rLC55 imino proton temperature study**

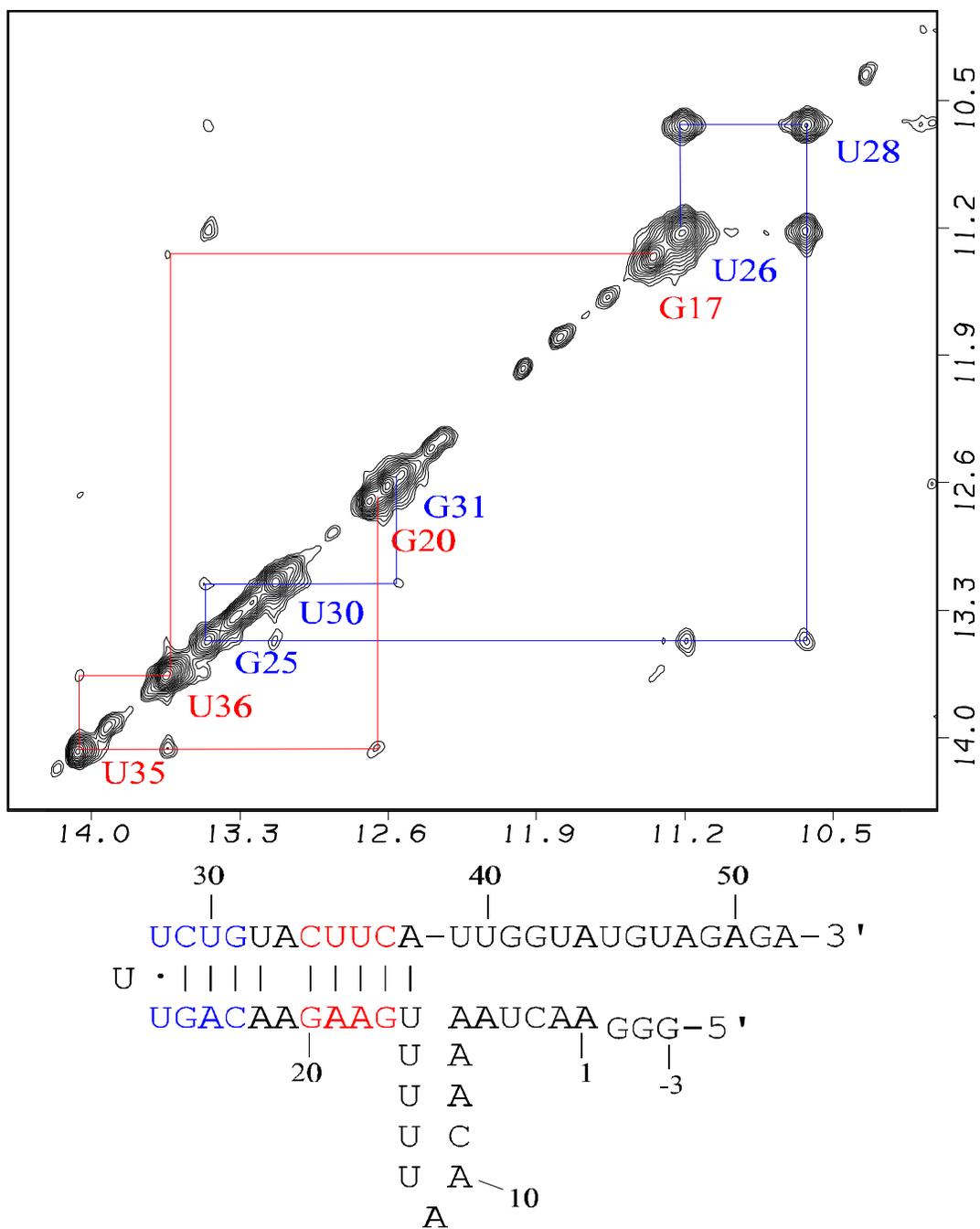
The high salt buffer conditions used in this imino proton temperature study was 20 mM sodium phosphate buffer (pH 6.5), 100 mM NaCl, and 1 mM EDTA, giving approximately 160 mM in sodium ions. The G25, U30, G17, U26 and U28 iminos appear much more intense and sharp as compared to the iminos found in the low salt buffer.



**Figure 1.16 High salt rLC55 derivative UV melts**

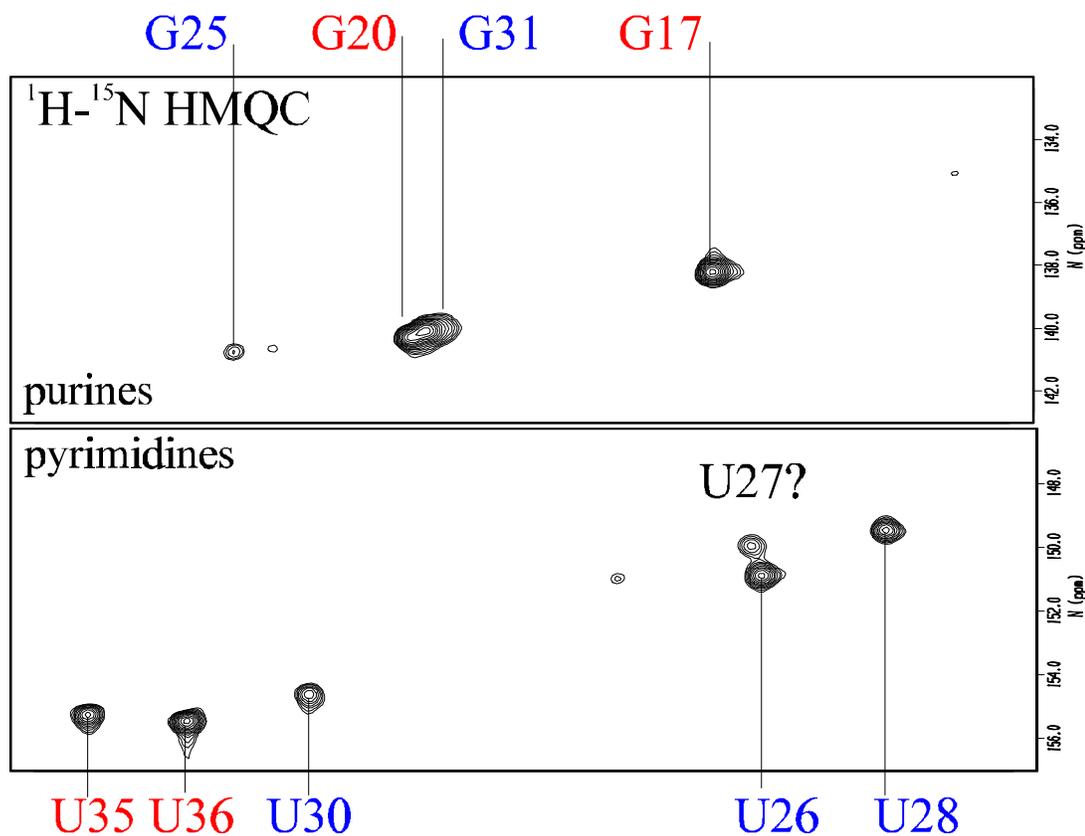
The UV melting curves of rLC55 are presented in the buffer conditions chosen for the NMR studies, 20 mM sodium phosphate (pH 6.5), 100 mM NaCl and 1 mM EDTA. More complete studies, with varied buffer conditions, have been performed (LeCuyer, 1992; Harris, 1994) and the data will not be duplicated here.

A) Low concentration UV melt,  $[\text{rLC55}] = 0.6 \mu\text{M}$  in a 10 mm cuvette. B) High concentration UV melt,  $[\text{rLC55}] = 120 \mu\text{M}$  in a 1 mm "etched quartz" cuvette. There appears to be no appreciable change in the melting profile, indicating that at these concentrations the rLC55 sample is not involved in duplex aggregation.



**Figure 1.17** rLC55 H<sub>2</sub>O NOESY

Jump-Return Spin Echo water suppressed H<sub>2</sub>O NOESY spectrum on the rLC55 SL RNA at 25° C. The pattern of iminos fits with the form 1 secondary structure as shown above. The imino protons most stabilized by the higher salt conditions (see Figs. 1.15 and 1.16) are mapped to the loop region of the hairpin.



**Figure 1.18** rLC55  ${}^1\text{H}$ - ${}^{15}\text{N}$  HMQC

This HMQC was used to identification of the base type of the imino protons. The most striking feature of this spectrum is the U26 and U28 iminos, which have a strong imino-imino NOESY crosspeak signature. They are involved in a U=U wobble. The U27 imino is tentatively assigned as the weak third upfield shifted uridine imino. Notice that the G25 and G17 imino are shifted downfield and upfield, respectively, from the region a "normal" Watson-Crick G:C base-paired imino would appear.

can be identified. The first (shown in blue) is U<sub>26</sub>=U<sub>28</sub>, G<sub>25</sub>, U<sub>30</sub>, G<sub>31</sub> and the second (shown in red) is G<sub>20</sub>, U<sub>35</sub>, U<sub>36</sub>, G<sub>17</sub>.

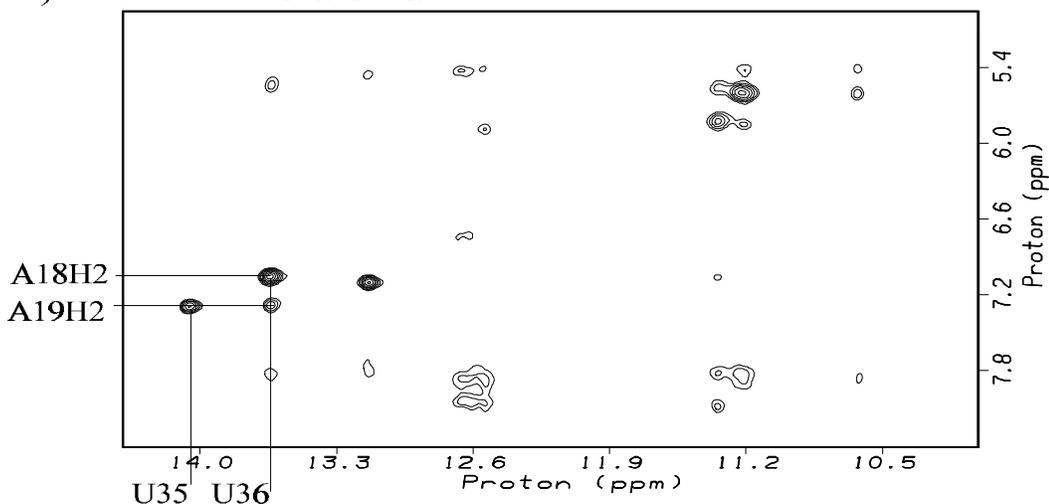
The orientation of the second stretch of iminos (G<sub>20</sub> – G<sub>17</sub>) was identified by observing the NOESY crosspeak between the U<sub>36</sub> imino and A<sub>19</sub>H2 proton, but not between the U<sub>35</sub> imino and A<sub>18</sub>H2 proton (Fig. 1.19). This pattern can only be explained by the assignment of the iminos in the orientation shown.

Unambiguous confirmation of the secondary structure assignment was accomplished by comparison of the imino NOESY crosspeak patterns of rLC55 with that of the smaller rLC25 RNA (Fig. 1.20). The rLC25 RNA can only adopt the form 1 hairpin because it is missing the nucleotides required for the form 2 base pairing, consequently, comparison of the imino proton spectra between these two samples will prove whether the rLC55 RNA is in the form 1 structure. The spectra are nearly identical, with only the G<sub>17</sub> imino proton shifting slightly, which can be explained by its proximity to the form 1 hairpin termini where one would expect a slight structural difference between the two samples.

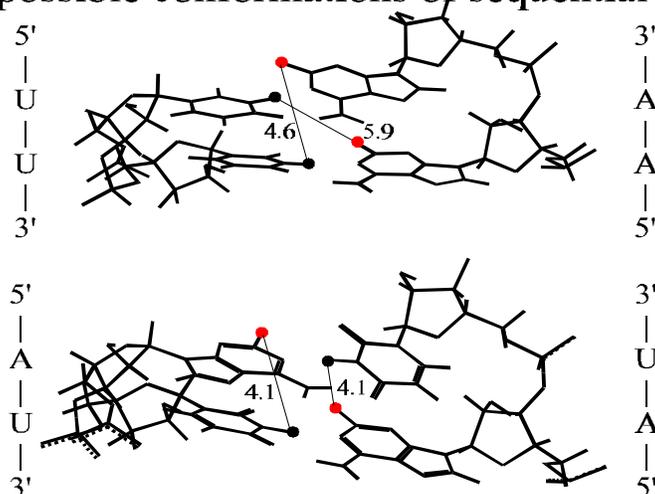
High-resolution studies on the non-exchangeable protons for rLC55 were attempted. The standard experiments, such as the D<sub>2</sub>O NOESY and DQFCOSY experiments were conducted with little success (data not shown). The spectral resolution of the data were poor with many overlapped resonances. Qualitatively, the T2 relaxation properties of this large RNA made most of the resonances broad and difficult to assign.

In an attempt to solve the enormous spectral overlap problem, the six cytosines found in rLC55 were selectively <sup>15</sup>N/<sup>13</sup>C isotope labeled. This cytosine labeled sample

## A) rLC55 imino-aromatic



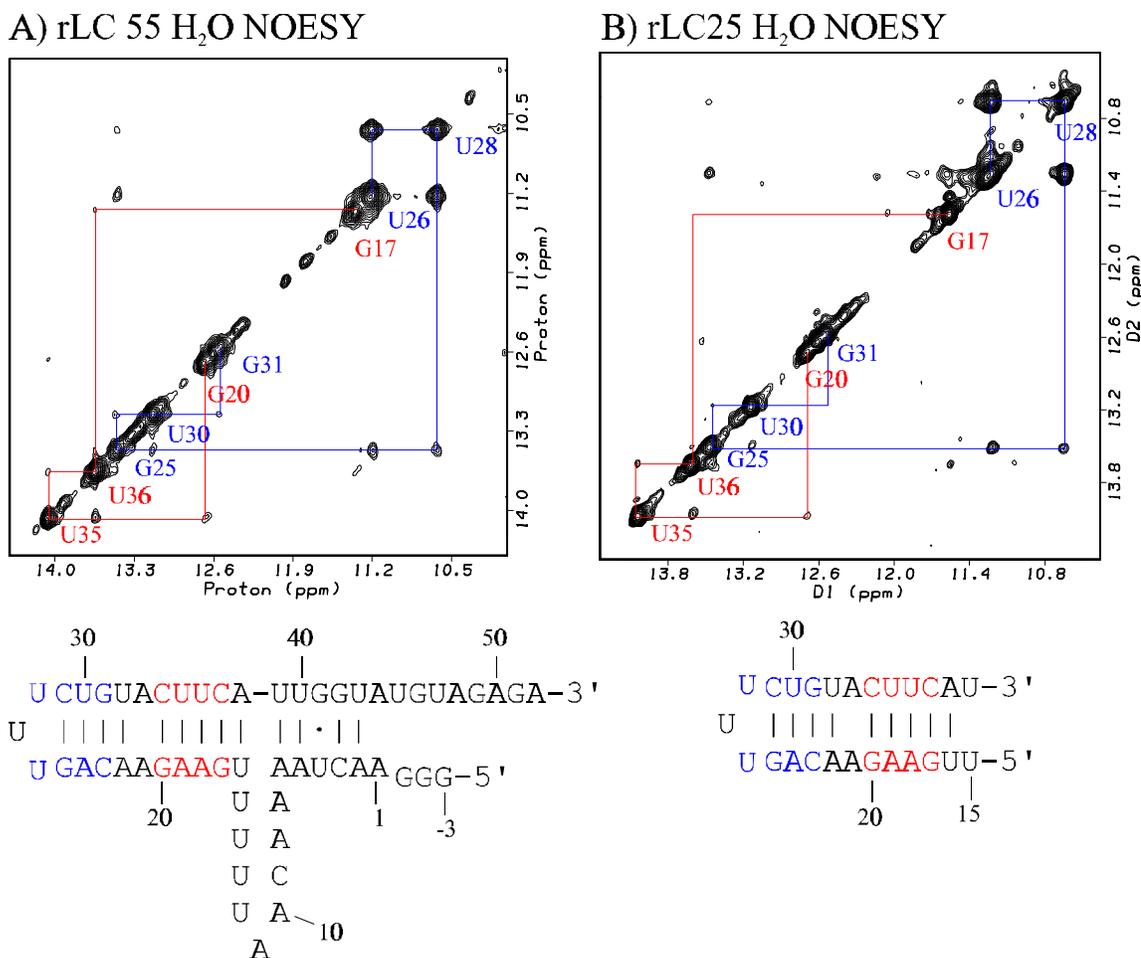
## B) Two possible conformations of sequential AU base pairs



**Figure 1.19** Assignment of the rLC55 U35 and U36 iminos

The assignment of the orientation of the AU to AU base pairing was accomplished by observing that the U imino NOESY crosspeak to the neighboring AH2 proton was asymmetric. If one builds a 5'-AU-3' duplex (B), by the nature of the symmetry, the crosspeak intensities will be nearly identical between each U imino to the neighboring AH2 proton. This would appear as a "box" of four crosspeaks in the region of the NOESY spectrum. This symmetric imino-AH2 crosspeak pattern has been observed in other RNAs in which there exists a 5'-AU-3' region of the sequence (personal communication, Dave Schweisguth).

Thus, the assignment of the U35 imino and the U36 imino followed from the observation that the imino-neighboring AH2 proton crosspeak pattern was asymmetric (see A above), and that the 3'-most uridine will have the stronger imino-neighboring AH2 proton crosspeak due to the 1.5 Å closer distance.



**Figure 1.20 Comparison of the H<sub>2</sub>O NOESY for rLC55 and rLC25**

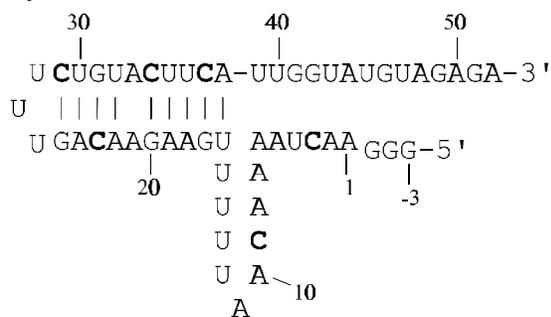
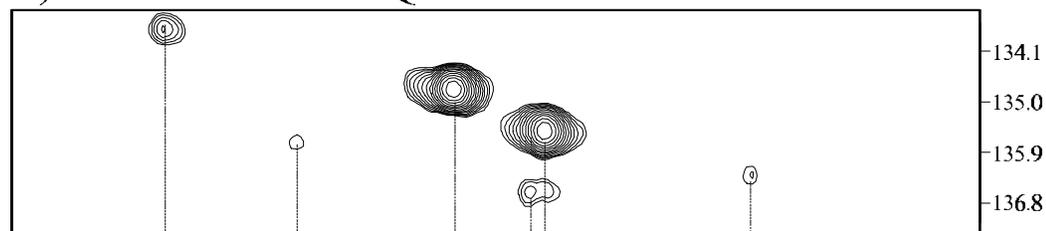
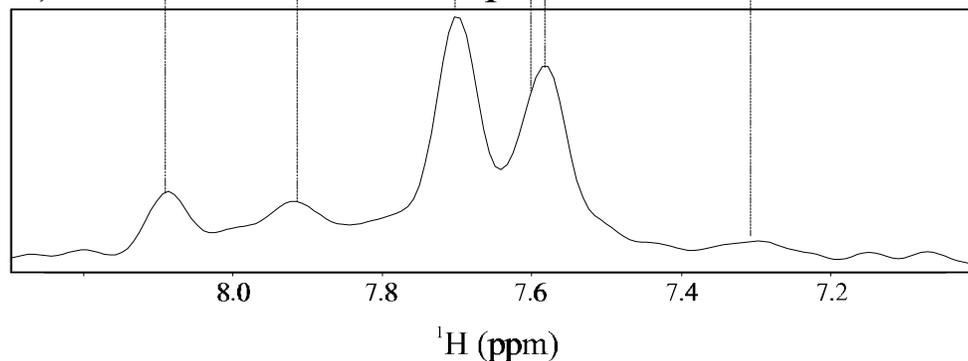
The two spectra were collected under identical conditions, 25°C and 250 ms mixing time. A) the data from the rLC55 sample, B) the rLC25 form 1 hairpin sample. The chemical shifts of the iminos are nearly identical between the rLC55 and rLC25 sample, indicating that the secondary structure of the rLC55 RNA is represented well by the form 1 hairpin. The G17 imino does shift slightly, which is not unexpected because it is near the termini of the form 1 hairpin, where one would expect differences between the rLC55 and rLC25 samples.

could then be explored using isotope-filtered experiments (see chapter 3 for a discussion). While this did solve the problem of spectral overlap, the broad linewidths due to short T2 relaxation times became even worse. The 2D constant time  $^1\text{H}$   $^{13}\text{C}$  HSQC and the 1D filtered spectrum are shown in figure 1.21. The two intense resonances are assigned to the C<sub>3</sub> and C<sub>9</sub>, and the broadened resonances (C<sub>23</sub>, C<sub>29</sub>, C<sub>34</sub> and C<sub>37</sub>) are those involved in the form 1 hairpin base-pairing region. A 2D isotope filtered NOESY experiment (see chapter 3) was also collected on this sample (data not shown) and the  $^{12}\text{C}$  subspectrum was characterized by broad overlapped peaks, and the  $^{13}\text{C}$  subspectrum had very little signal because of the extra proton relaxation by the  $^{13}\text{C}$ . High-resolution characterization of the non-exchangeable protons on rLC55 was not successful and any further studies would have to utilize the smaller rLC25 and rLC13 form 1 hairpin fragments.

### 1.3.6 Salt mediated conformational change in the *L. collosoma* form 1 hairpin

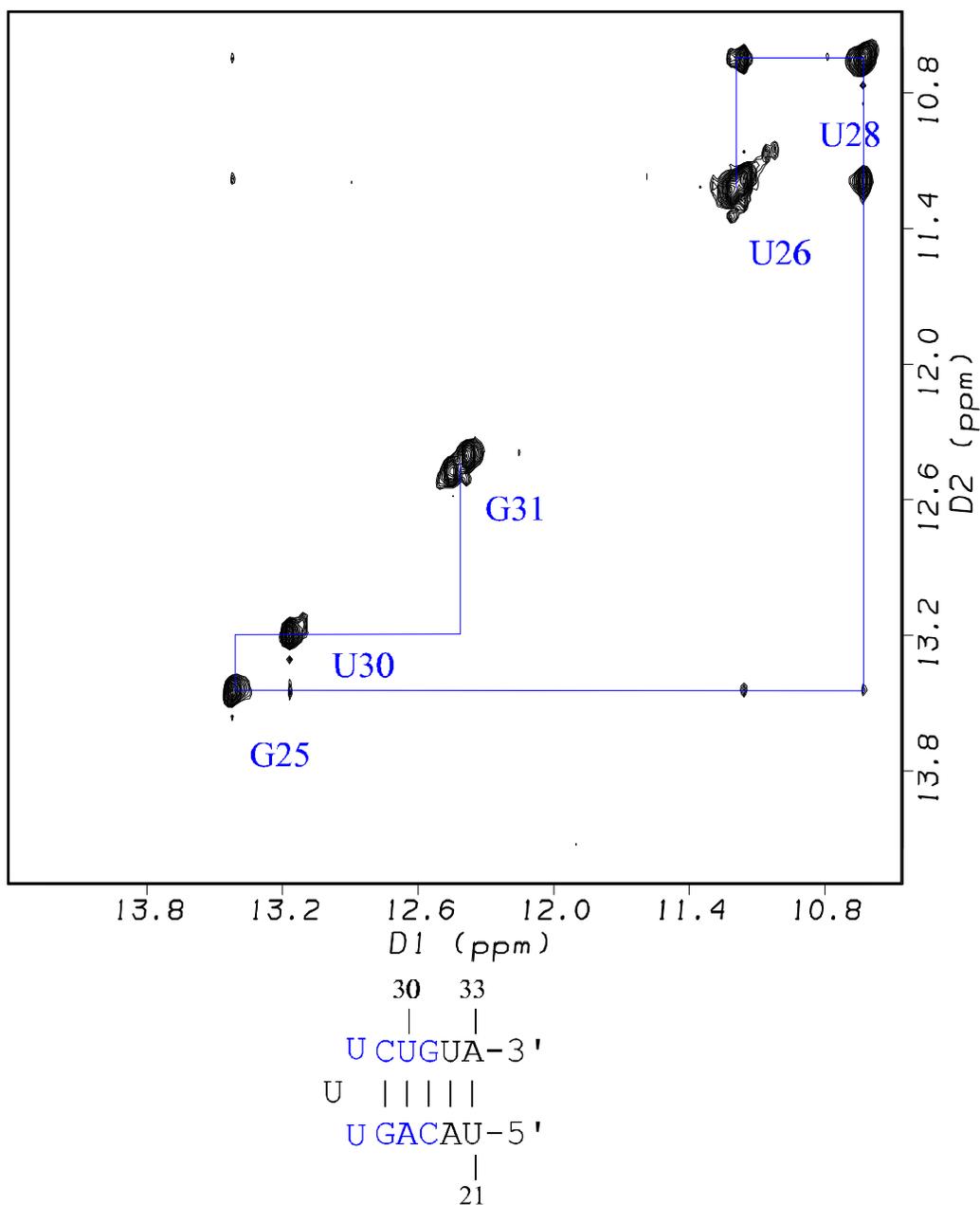
The smaller fragments, rLC25 and rLC13, provided a means to study the form 1 hairpin at higher resolution than was possible with the large rLC55 sample. One intriguing structural feature of these form 1 hairpins is the strong imino-imino crosspeak between the two uridines (Fig. 1.22). This is intriguing because the only position in the sequence where this U=U base pair can form is between U<sub>26</sub> and U<sub>28</sub> in the hairpin loop. Requiring that the hairpin loop be spanned by a single nucleotide, U<sub>27</sub>, unless the RNA is a duplex. Further characterization of this hairpin loop was necessary.

To examine if the rLC25 and rLC13 RNA existed in a single conformation on the NMR time scale, the "double quantum-filtered COSY" (DQFCOSY) NMR experiment was performed. This experiment correlates protons that are three bonds away from each

A) Cytosine  $^{13}\text{C}$  labeled rLC55 NMR sampleB) 2D  $^1\text{H}$ - $^{13}\text{C}$  CT-HSQCC) 1D  $^{13}\text{C}$  selected  $^1\text{H}$  subspectrum

**Figure 1.21** rLC55 cytosine  $^{13}\text{C}$  labeled spectra

A) The rLC55 sample were synthesized with only the six cytosines  $^{13}\text{C}$  isotope labeled, shown in bold. B) The constant time HSQC showing the 6 cytosine correlation between the H6 proton and the C6 carbon. The two strong intensity peaks are probably from the non-base paired C3 and C9 nucleotides that are experiencing a faster local correlation time due to local dynamical movement. C) The 1D  $^{13}\text{C}$  subspectrum (see chapter 3) from the same sample.

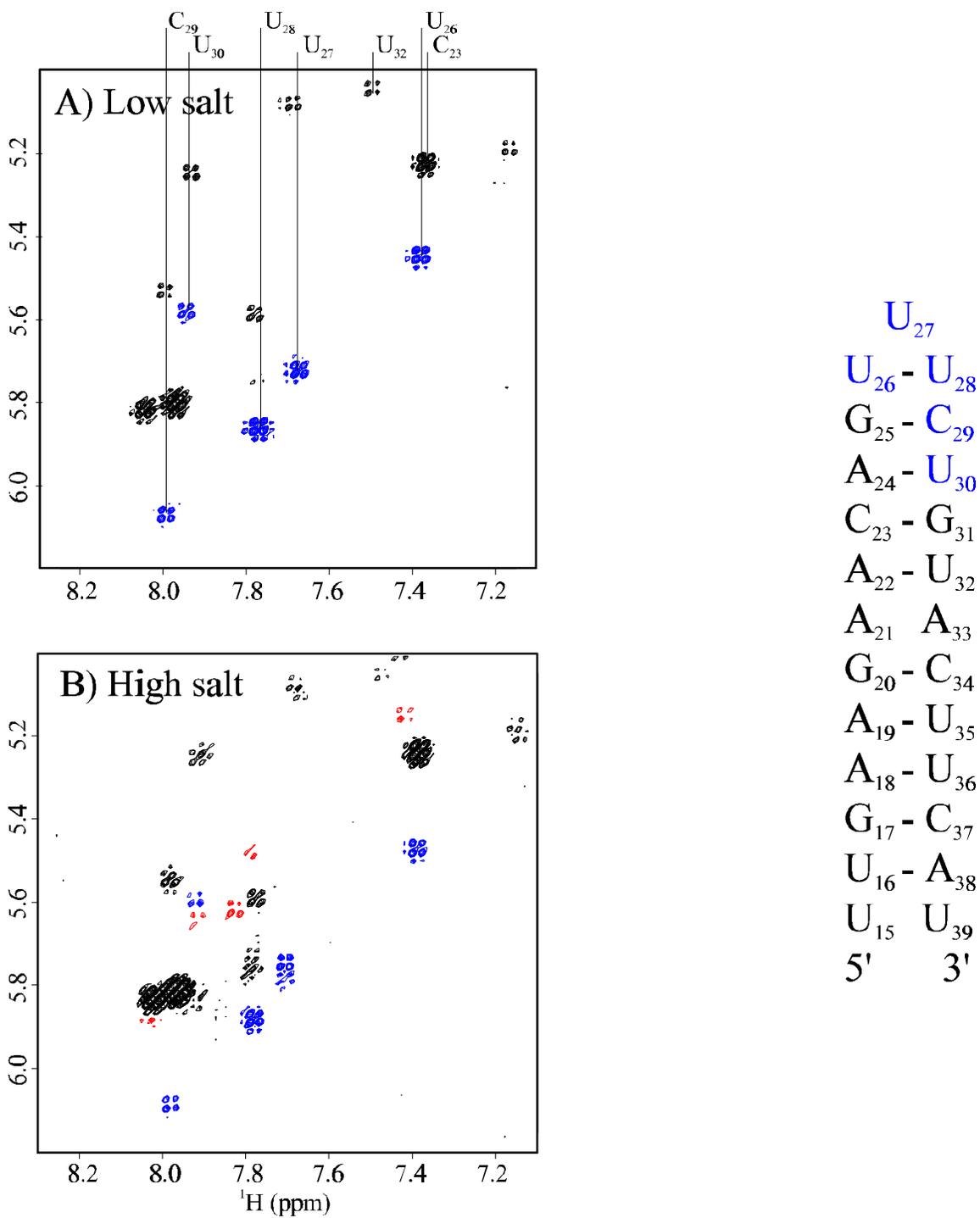


**Figure 1.22 2D H<sub>2</sub>O NOESY of rLC13**

The imino protons for the rLC13 RNA are a subset of those for rLC55 and rLC25. The crosspeak between U30 and G31 has never been seen in a H<sub>2</sub>O NOESY experiment for rLC13, but the connectivity is drawn in above, by inference from the other NOESY data for rLC25 and rLC55. The absence of the cross peak is probably due to a higher exchange rate of the G31 imino with H<sub>2</sub>O since it is near the terminus of the helix and not because of some major structural change.

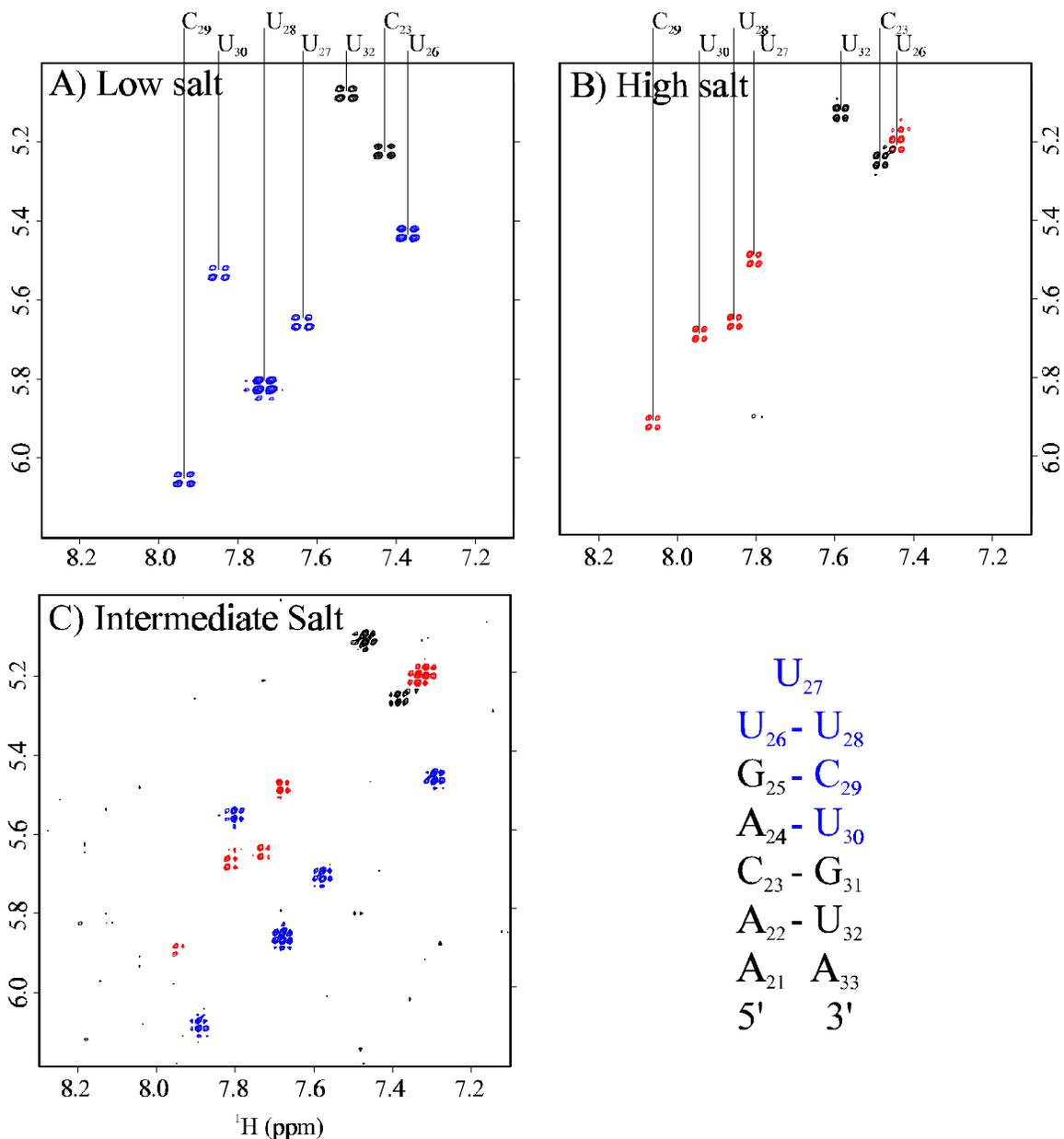
other. The intensity of the crosspeak is dependent on the magnitude of the vicinal  $^3J$ -coupling constant between the protons. Because of the anti-phase nature of the crosspeak quartets, small coupling causes the crosspeaks to cancel out. The  $^3J$ -coupling intensity between 3 bond distant protons follows the Karplus relationship (1959), with a maximum coupling at 0 and 180 degrees and a minimum at 90 and 270 degrees. The H5 and H6 protons in the pyrimidine bases of nucleic acids are ideal protons to observe with this experiment because they are fixed in position relative to each other at 0 degrees. Thus they have a large  $^3J$ -coupling constant (~10-12 Hz) and are strong crosspeaks in the dqfcosy experiment. For determining if an NMR sample exists in a single conformation, one only has to count the number of H5-H6 DQFCOSY crosspeaks. If they add up the same number as what is expected, then the sample is in a single time-averaged fast-exchange conformation.

DQFCOSY spectra for rLC25 and rLC13 were collected for different buffer salt conditions, and the results are shown in figures 1.23 and 1.24, respectively. We observed more H5-H6 crosspeaks in the spectrum corresponding to the "intermediate" salt conditions (50 mM NaCl) than can be accounted for by the sequence. This could be explained by the existence of two structural conformations in slow-exchange. If the ionic strength of the buffer was lowered (<30 mM NaCl), one of the two conformations was favored (state A) and if the ionic strength of the buffer was raised (>150 mM NaCl), the other conformation was favored (state B). This ability to favor one conformation over the other by adjusting the salt concentration was seen for both the rLC25 and rLC13 samples.



**Figure 1.23** rLC25 salt dependence DQFCOSY data

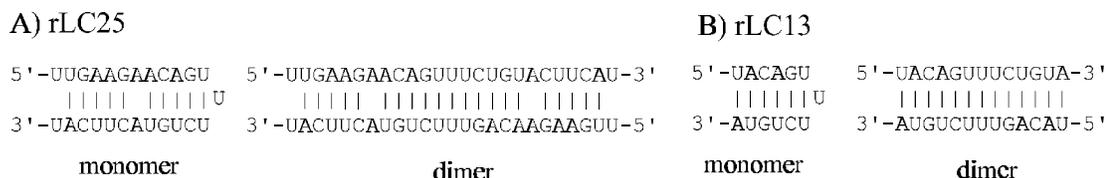
Low salt conditions: 10 mM sodium phosphate buffer (pH 6.5), 1 mM EDTA. B) High salt conditions: 10 mM sodium phosphate (pH 6.5), 1 mM EDTA and 100 mM NaCl.



**Figure 1.24 rLC13 salt dependence DQFCOSY data**

All RNA samples contained the rLC13 sample dialyzed against A) 10 mM phosphate buffer (pH 6.5) and 1 mM EDTA, B) 10 mM phosphate buffer (pH 6.5), 1 mM EDTA and 100 mM NaCl. C) The intermediate buffer condition contained 50 mM NaCl. dimeric duplexes as shown in figure 1.25.

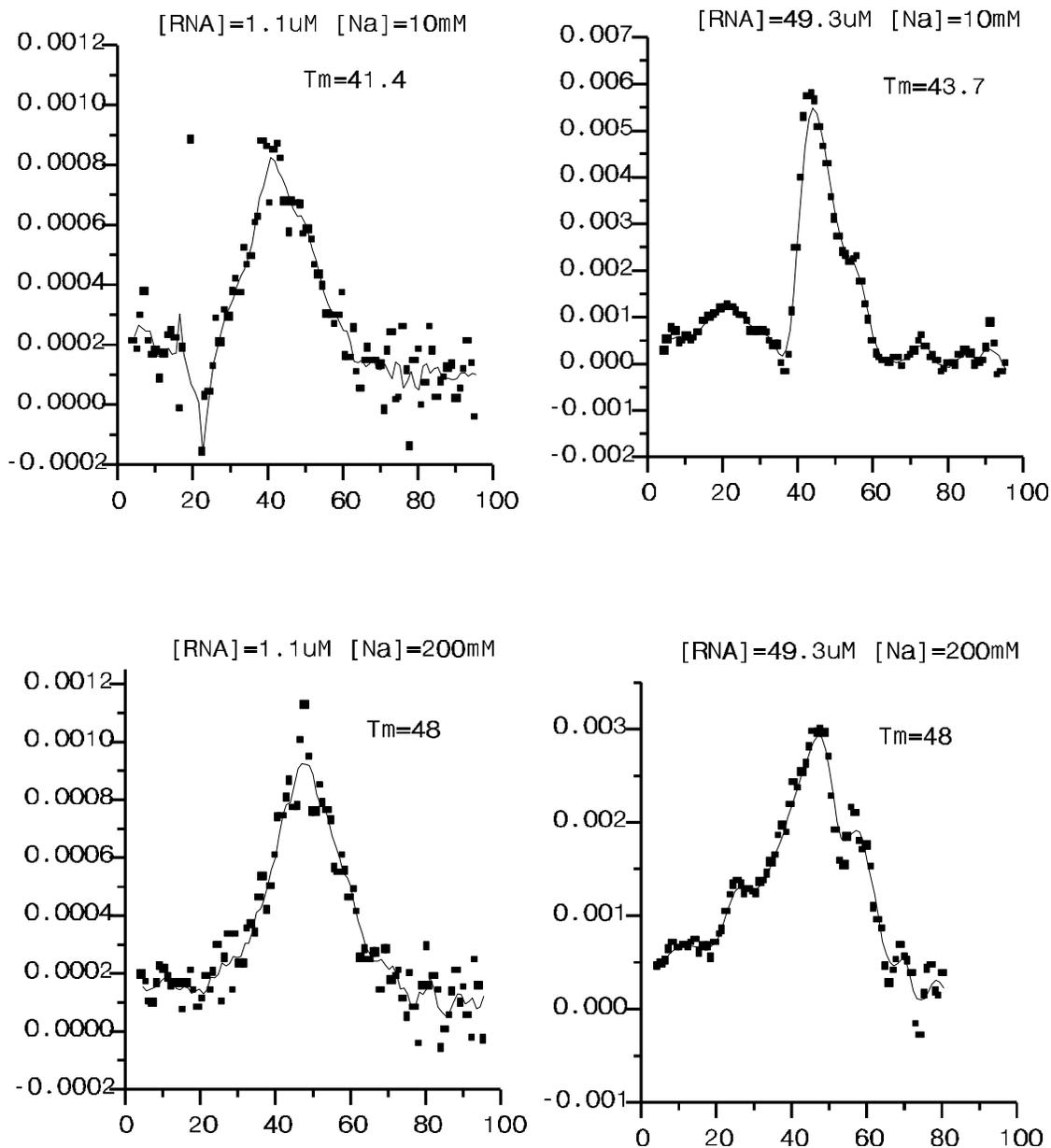
One explanation for the ion strength dependent conformations would be a hairpin-duplex transition. Small RNAs such as these can form either monomeric hairpins or dimeric duplexes as shown in figure 1.25. These monomer-dimer structures are difficult to differentiate by NMR. Because of the inherent dyad symmetry of the dimer they can appear nearly identical to the monomer hairpin spectroscopically. A number of biophysical and NMR techniques can be used to determine if a sample is a monomer or dimer; see chapter 4 for a full discussion of the methods. We used the techniques of optical UV melts and translational diffusion constant measurements to clarify this issue.



**Figure 1.25 rLC13 and rLC25 basepairing possibilities for a monomer or dimer**

### 1.3.6 Evidence that the rLC13 low and high salt samples are monomeric

Optical UV melting curves for the rLC13 samples are shown in figure 1.26 for both the low and high salt conditions, and at two different RNA concentrations. The stability of a hairpin is independent of its concentration in solution, while the stability of a dimer is dependent on its concentration. Thus, UV melting curves will show a concentration dependence to the measured melting temperature ( $T_m$ ) for a dimer and not for a monomer. Another variable is the ionic strength of the solution. High salt stabilizes the hairpin (or base paired) structures and as expected the  $T_m$  of the high salt samples were higher (6 degrees). However, there was no perceptible RNA concentration dependence to the  $T_m$  of either the low or high salt sample.



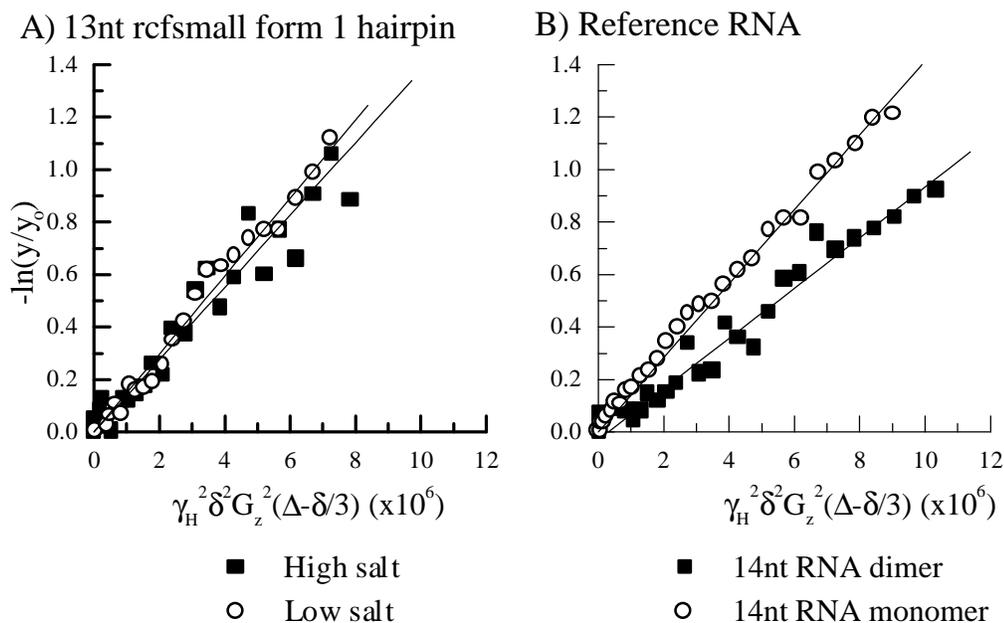
**Figure 1.26** rLC13 UV melts

Equilibrium melting curves for rLC13 in the low and high salt buffer used in the NMR experiments. The top two graphs are the low salt buffer (10 mM sodium phosphate pH 6.5, 1 mM EDTA) with [RNA]=1.1 μM and 49.3 μM respectively. The bottom two graphs are the high salt buffer conditions (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 1 mM EDTA) with [RNA]=1.1 μM and 49.3 μM.

The melting temperature is ~42° C for the low salt conditions and ~48° C for the high salt conditions with no appreciable RNA concentration effects.

Unfortunately, this result alone does not guarantee that the same is true for the NMR samples. The major problem with reliance on UV melting curves in this type of analysis is that it is difficult to perform the melts at RNA concentrations high enough to perform NMR experiments (millimolar). The highest concentrations of RNA that can be used for UV melting curves is ~50-200  $\mu\text{M}$  (depending on the size of the RNA) and requires the use of special short path length cuvettes, such as the 1 mm cuvettes used in this study. Thus, even though our results indicate both samples were monomeric, other methods must be employed to secure our conclusion at NMR concentrations. For that, the NMR based method of measuring the translational self-diffusion constants (see chapter 4 for a full discussion) was used.

The translational self-diffusion constants for the low and high salt rLC13 sample were measured as  $1.40 \times 10^{-6} \text{ cm}^2/\text{s}$  and  $1.45 \times 10^{-6} \text{ cm}^2/\text{s}$ . The data are shown in figure 1.27. The results are compared to those from a 14 nt reference RNA which can be examined as either a monomer or a dimer (Lapham, et al., 1997). Hydrodynamics theory predicts that for RNAs of this size, the dimer:monomer ratio of the diffusion constants should be approximately 0.65. This was exactly what was observed when we measured the 14 nt reference RNA in its two conformations. The ratio obtained for the diffusion constant of both the low and high salt rLC13 samples was approximately 1, suggesting that they are similar hydrodynamically. Furthermore, the absolute diffusion rate measured for the rLC13 samples ( $\sim 1.4 \times 10^{-6} \text{ cm}^2/\text{s}$ ) is what is expected of a 13 nt monomer (see chapter 4 for discussion on predicting diffusion constants). Therefore, the diffusion rate measurements predict that both the low and high salt rLC13 samples are monomeric.



**Figure 1.27 Diffusion rate of rLC13 low and high salt conformations**

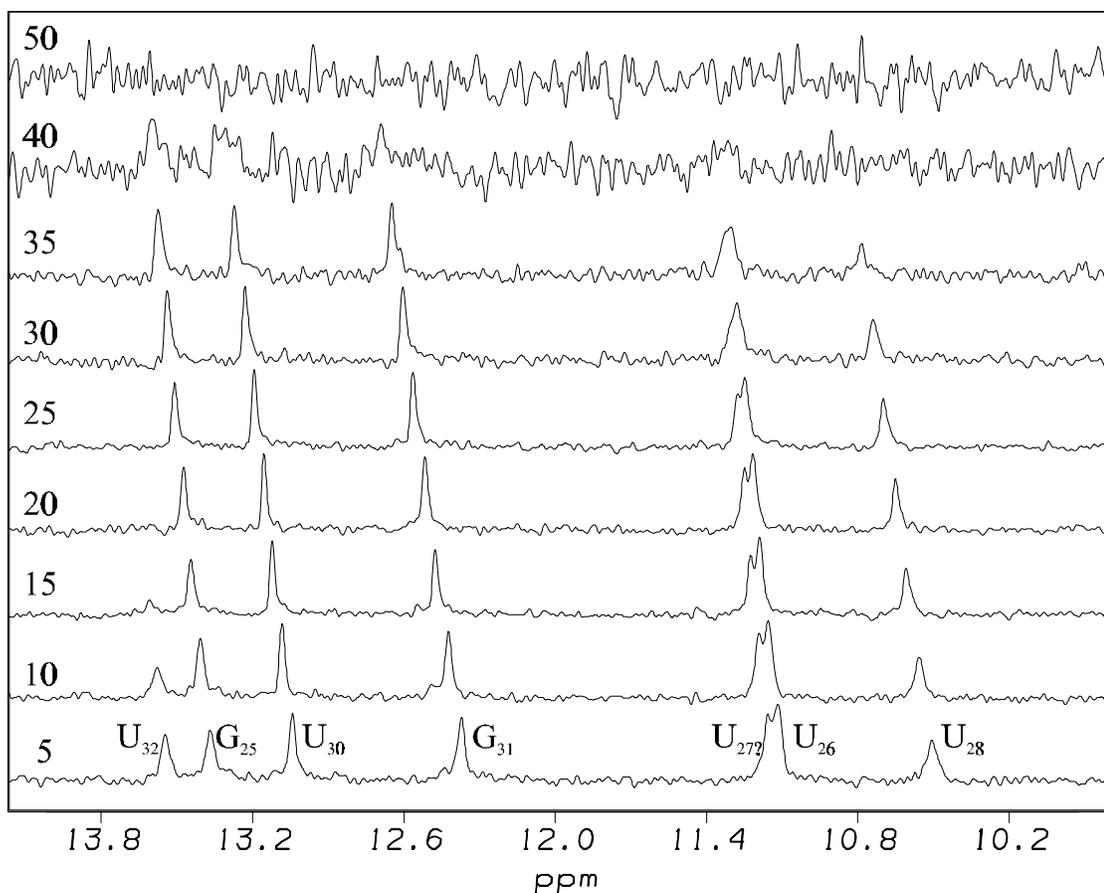
The post-processed (see materials and methods) data from the pulsed field-gradient stimulated echo (pfg-STE) experiment.  $\Delta=0.1s$  and  $\delta=0.004s$ , the gradient was varied from 0 to 32 Gauss/cm in steps of 1 Gauss per increment. The experiments were conducted at 25°C. A) The 13 nt rLC13 RNA with a measured diffusion rate of  $1.40 \times 10^{-6}$  and  $1.45 \times 10^{-6} \text{ cm}^2/s$  for the low and high salt sample respectively. B) Reference data from a 14 nt RNA (see chapter 4) in either a monomeric hairpin or a duplex form.

The 1D imino proton melts of rLC13 are shown in figure 1.28 for the high salt sample. In the low salt melting experiment (data not shown) all the imino protons disappear by 10° C except for the imino from G31, which melts out at 30° C.

### *1.3.7 Assignments of the non exchangeable protons for rLC13*

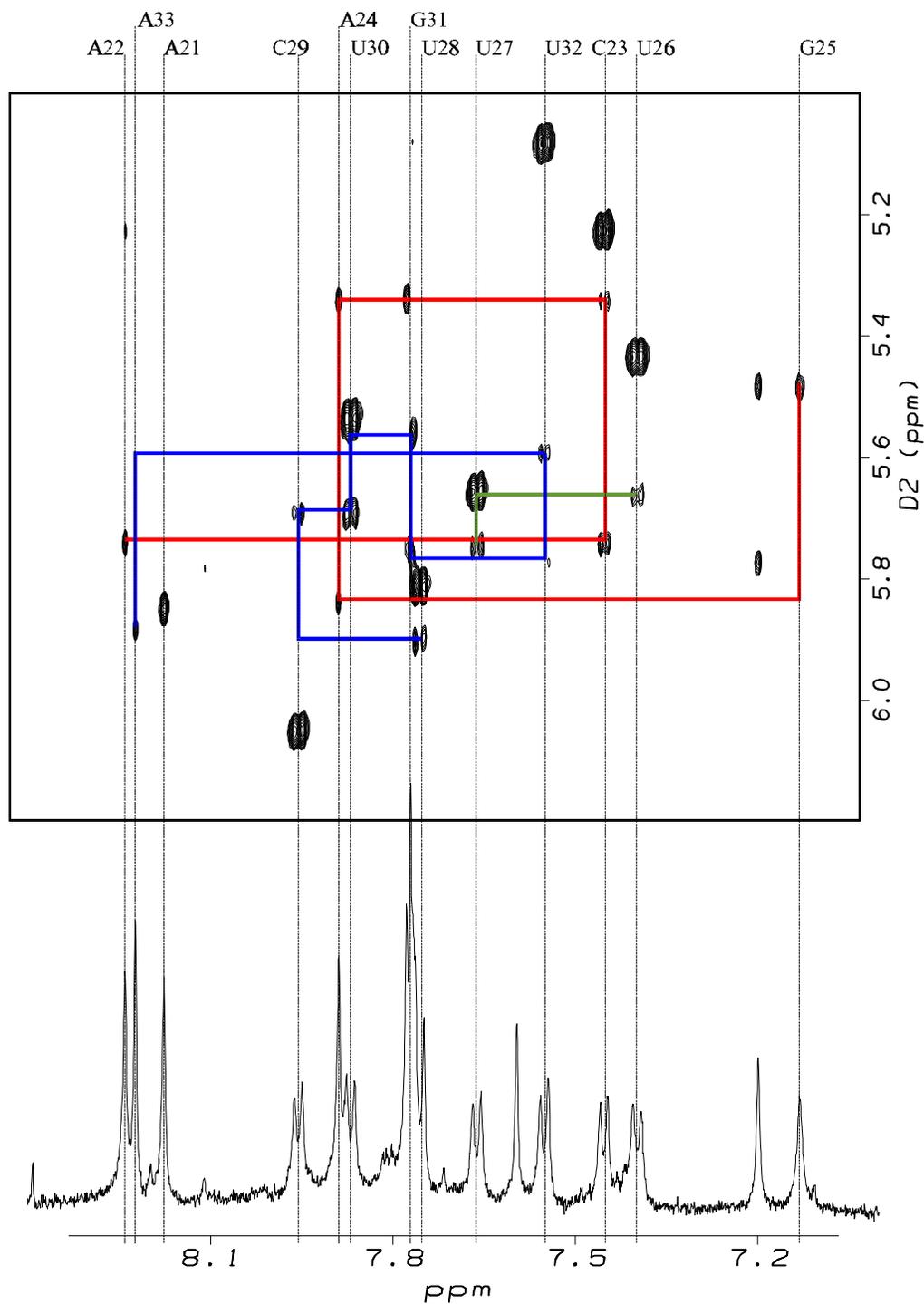
The 2D NOESY spectra of the low salt rLC13 sample in D<sub>2</sub>O are shown in figures 1.29 and 1.30. The NOESY spectra from the high salt rLC13 sample in D<sub>2</sub>O are shown in figures 1.31 and 1.32. The non-exchangeable protons were assigned using the "anomeric-aromatic walk" in which the H6/H8 base proton of a nucleic acid is correlated to its own H1' and the H1' in the 5' direction.

Aside from the anomeric-aromatic walk, additional connectivities confirmed the assignments of the protons. As an example, the A<sub>24</sub> H2 proton cross-strand and same-strand NOEs confirmed the assignments to the G<sub>31</sub> H1' and G<sub>25</sub> H1'. The exchangeable imino protons were correlated to the non-exchangeable protons by the 2D watergate NOESY experiment (Fig. 1.33). This experiment allows for the observation of exchangeable to non-exchangeable NOE crosspeaks that appear close to the water resonance. As an example, the G<sub>31</sub> imino has a strong NOE connection to the amino protons on C<sub>23</sub>, which then show a strong crosspeak to the C<sub>23</sub> H5. In this manner, the assignments of the cytosine H5 protons could be reaffirmed.



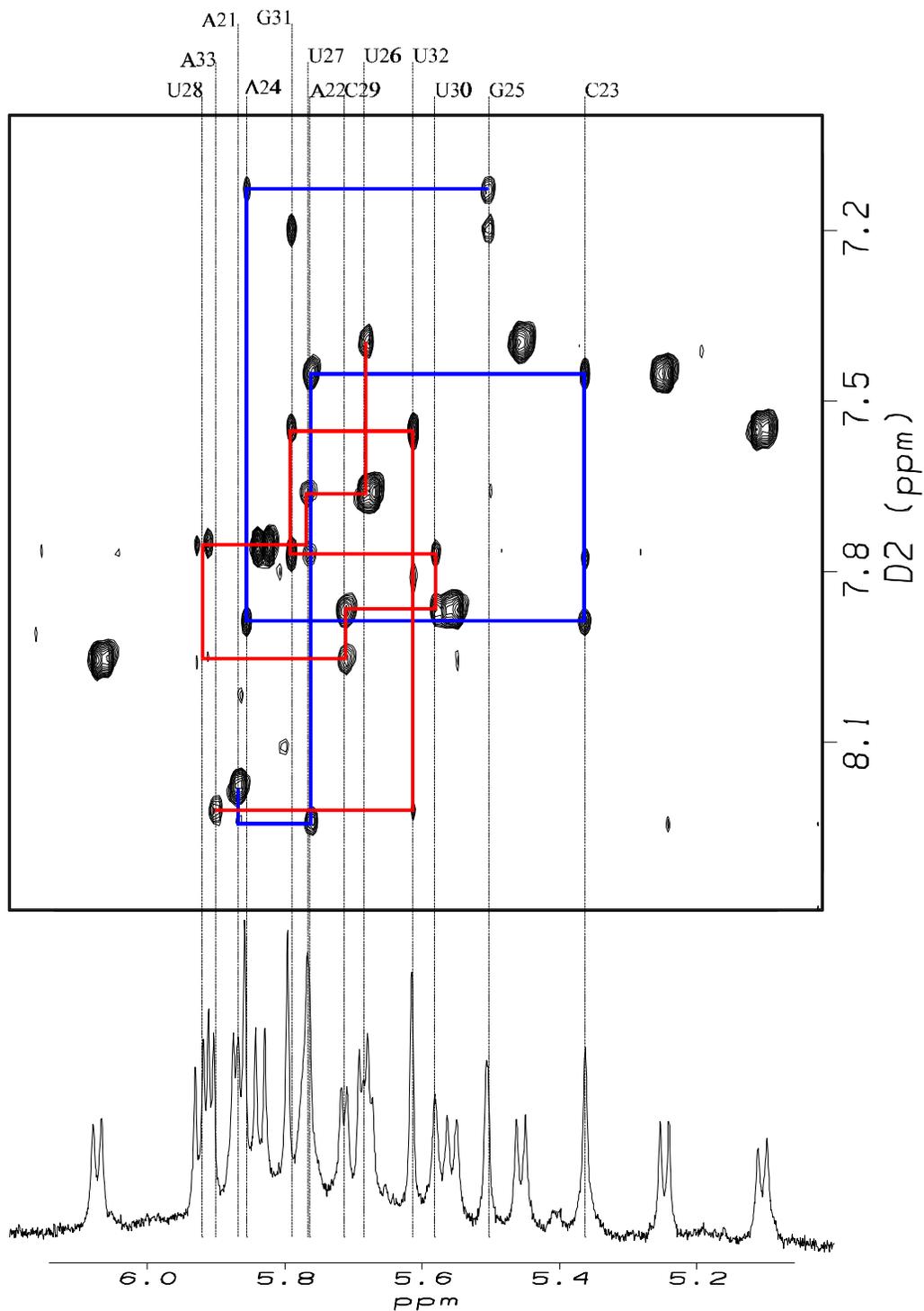
**Figure 1.28 Temperature study of the imino proton from high salt rLC13**

The imino protons from the rLC13 RNA sample in a high salt buffer. The profile is very similar in terms of the chemical shifts seen for rLC55, but these imino protons have much sharper line widths. The U32 imino (near the terminus) melts out at a low temperature (15°), while the other iminos melt out at 40° C. The U27/U26 and U28 iminos seem to exchange broaden before the stem iminos do, which might indicate they are involved in a more solvent accessible conformation. The low salt temperature study on rLC13 showed all the same iminos at 5° C, but only the G31 imino was visibly above 10° C.



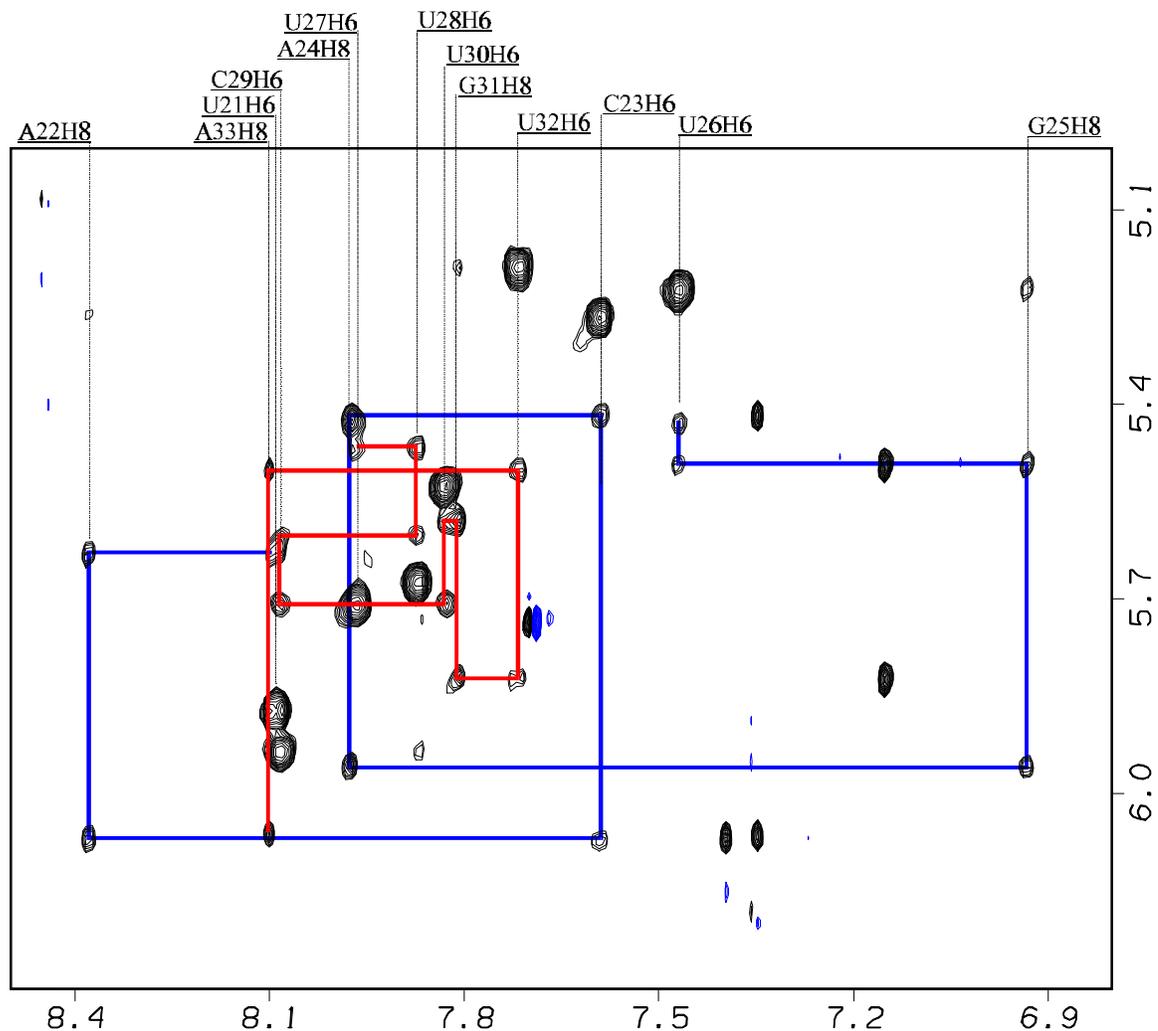
**Figure 1.29** NOESY spectrum of the low salt rLC13

The mixing time was 250 and the temperature was 20° C. The anomeric-aromatic walk is demonstrated with the overlay lines. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAA-3'.



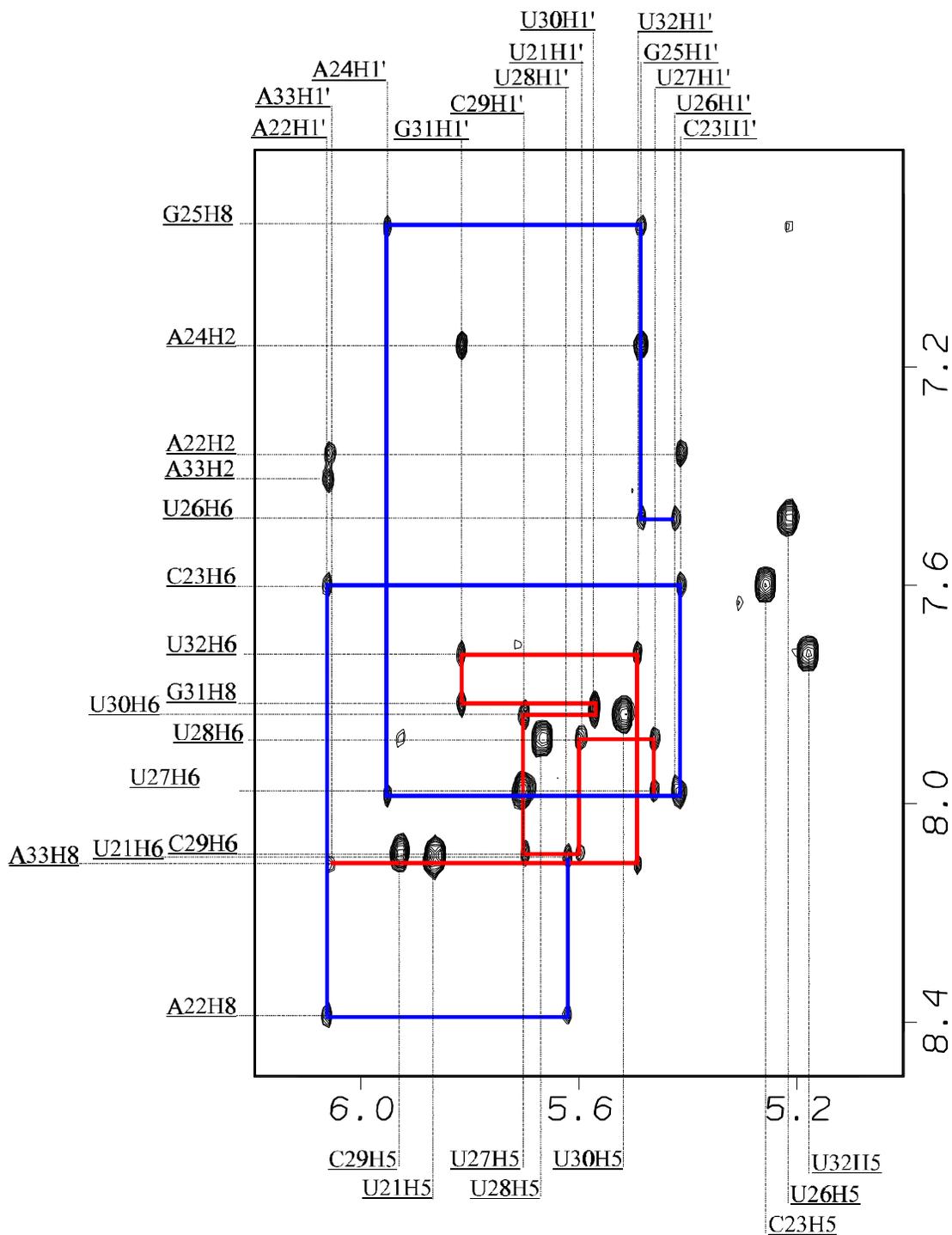
**Figure 1.30** NOESY spectrum of the low salt rLC13

The same experiment as shown in figure 1.29, but with the limits transposed. Some of the connectivities are better seen in this region of the spectrum. The mixing time was 250 ms and the temperature was 20° C. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAA-3'.



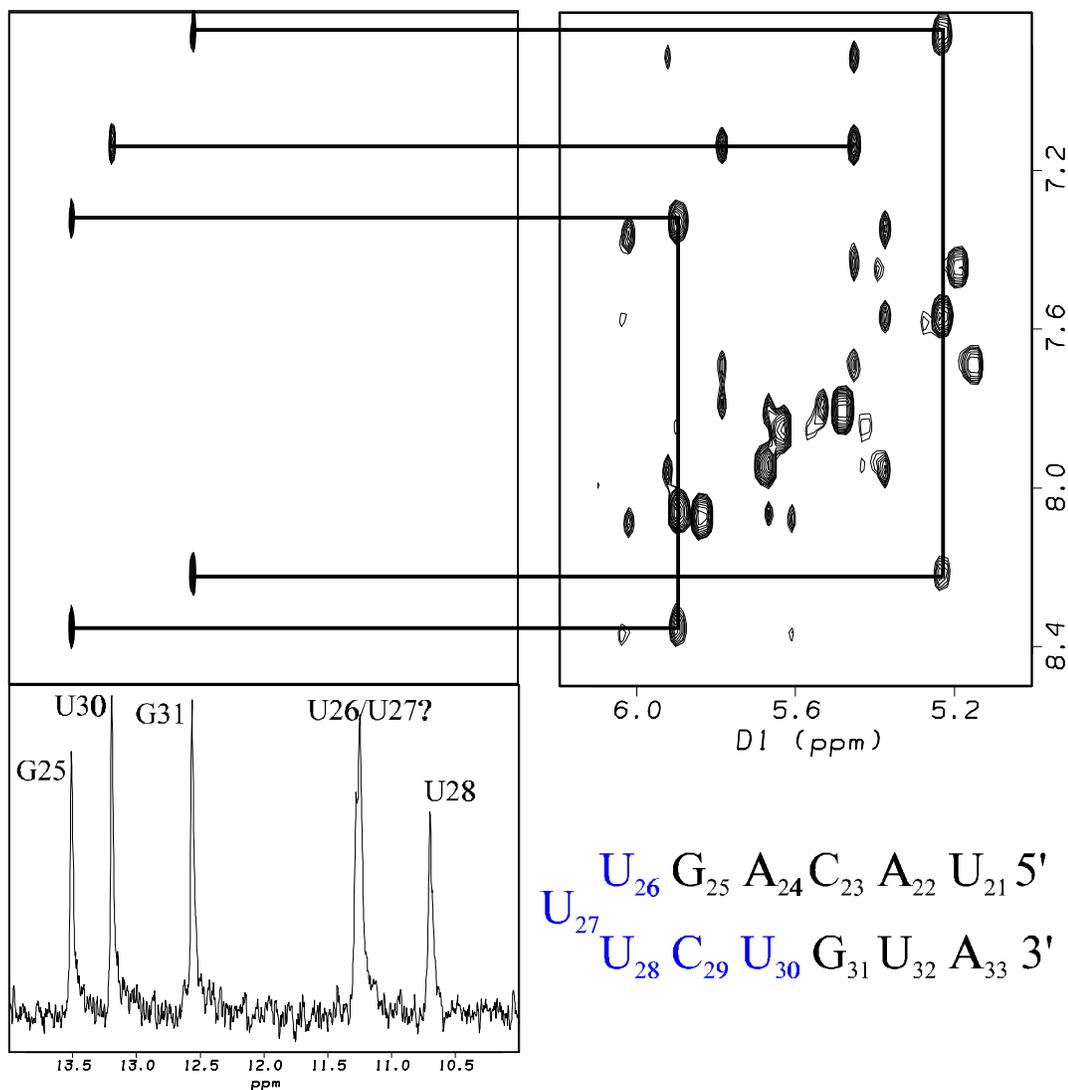
**Figure 1.31** NOESY spectrum of the high salt rLC13

The mixing time was 250 ms and the temperature was 25° C. The assignments are shown with the dotted lines and the solid lines represent the anomeric-aromatic walk. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAU-3', there is an extra uridine on the 3' end as compared to the data shown for the low salt sample.



**Figure 1.32** NOESY spectrum of the high salt rLC13

The mixing time was 250 and the temperature was 25° C. Assignments are shown with the dotted lines and the anomeric-aromatic walk is shown with the solid lines. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAU-3', there is an extra uridine on the 3' end as compared to the data shown for the low salt sample.



**Figure 1.33** Imino to non-exchangeable of high salt rLC13

Watergate NOESY spectrum with 300 ms mixing time at 25° C. Many of the non-exchangeable assignments can be confirmed with this experiment. For instance, the G31 and G25 imino protons have strong NOE cross peaks to the amino protons on their base pair partner cytosines. These aminos then have a strong connectivity to the H5 proton. Additionally, the U30 imino has a strong cross peak to the AH2 proton from A24.

Chemical shift data

	H8/H6	H5/H2	H1'	H2'
U21	8.09	5.88	5.63	4.61
A22	8.38	7.35	6.07	4.59
C23	7.60	5.27	5.42	
A24	7.98	7.16	5.96	4.72
G25		n/a	5.50	4.48
U26	7.47	5.23	5.44	
U27	7.98	5.71	5.47	4.22
U28	7.88	5.68	5.60	4.46
C29	8.03	5.94	5.71	4.43
U30	7.84	5.53	5.59	4.53
G31	7.82	n/a	5.83	4.48
U32	7.72	5.20	5.50	4.27
A33	8.10	7.40	6.07	4.15

**Table 1.1 rLC13 High salt assignments**

Chemical shift data

	H8/H6	H5/H2	H1'	H2'
A21	8.17		5.87	
A22	8.24	7.77	5.76	
C23	7.45	5.25	5.36	
A24	7.88	7.19	5.86	
G25	7.12	n/a	5.51	
U26	7.39	5.45	5.68	
U27	7.65	5.67	5.77	
U28	7.75	5.83	5.92	
C29	7.95	6.07	5.71	
U30	7.86	5.55	5.58	
G31	7.77	n/a	5.79	
U32	7.54	5.11	5.61	
A33	8.22		5.90	

**Table 1.2 rLC13 Low salt assignments**

## 1.4 Discussion

The goals of this project were to characterize the *in vitro* secondary structures of the SL RNA from the trypanosomes *L. collosoma* and *C. fasciculata*, and to determine what structural element was responsible for the form 1 low temperature UV melt transition found in the *L. collosoma* SL RNA.

### 1.4.1 The secondary structures of the *C. fasciculata* and *L. collosoma* SL RNAs

The *in vitro* secondary structures of the two SL RNAs have been identified using NMR techniques. The rCF55 RNA was found to exist in the form 2 structure by comparison of its  $^1\text{H}$ - $^{15}\text{N}$  HMQC to the form 2 fragment hairpin, rLC30. The rLC55 RNA was confirmed to exist in the form 1 secondary structure, as previously proposed (Lecuyer and Crothers, 1993) by comparing the  $\text{H}_2\text{O}$  NOESY spectrum of the parent RNA to that of the form 1 fragment hairpin, rLC25.

Because the *C. fasciculata* SL RNA has a closer sequence homology to all the other non-*L. collosoma* SL RNAs, we speculate that most of the trypanosome SL RNAs probably exist, *in vitro*, in the form 2 secondary structure. The biological significance of this is unclear, as it has been shown that the *in vivo* secondary structure of a SL RNA may be different to the *in vitro* secondary structure (Harris, et al., 1995).

### 1.4.2 The *L. collosoma* "tertiary" structure

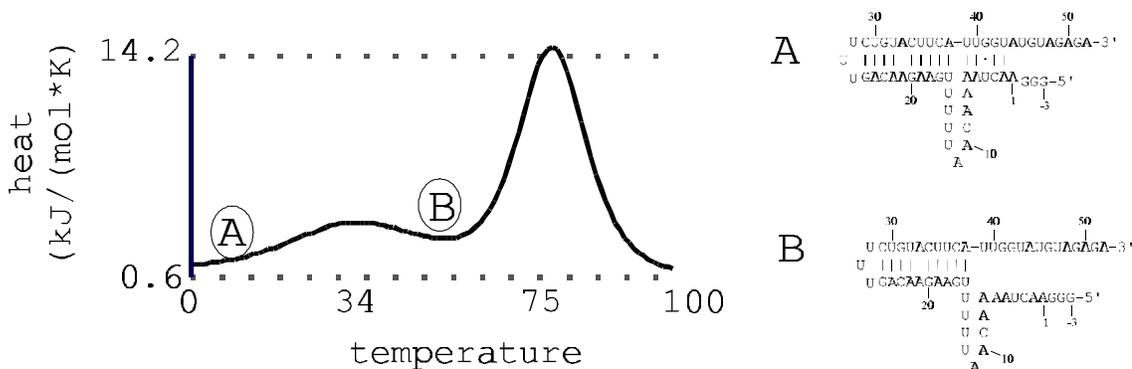
The low temperature melting transition has been speculated to be due to a "tertiary" structural element, which may exist between the form 1 hairpin and one of the two RNA "tails" that extends off the terminus of the hairpin. A number of the results presented in this chapter seem to contradict this possibility.

Figure 1.20 demonstrates that the chemical shifts of the imino protons of rLC55 and rLC25 are nearly identical, except for the G<sub>17</sub> imino proton. This G<sub>17</sub> imino proton is located on the terminus of the form 1 hairpin, and would be expected to experience a chemical shift change. The correlation of the chemical shifts of all other imino protons indicates that the form 1 hairpins are in similar environments in both samples. If there was some type of tertiary interaction between the hairpin and the rest of the RNA, one might expect the interaction to affect the environment of the exchangeable protons in the rLC55 sample. Thus, the 25 nucleotides of the form 1 hairpin are probably structurally independent of the rest of the SL RNA.

Another indication that the form 1 hairpin is not interacting with the rest of the SL RNA can be inferred by observing the linewidths in the 2D <sup>1</sup>H-<sup>13</sup>C CT-HSQC experiment (Fig. 1.21). The linewidths of the cytosines in the form 1 hairpin (C<sub>23</sub>, C<sub>29</sub>, C<sub>34</sub> and C<sub>37</sub>) and those in the 5' end of the RNA (C<sub>3</sub> and C<sub>9</sub>) suggest that the two regions of the RNA are characterized by different T<sub>2</sub> relaxation times. This can be explained if two regions of the RNA experience different effective correlation times, as would be the case if the 5' end of the RNA is unstructured. The cytosines in the form 1 hairpin would then experience the actual correlation time of the molecule, and C<sub>3</sub> and C<sub>9</sub> would experience a faster effective correlation time. As an example, it is often seen that the terminal nucleotides of a nucleic acid duplex have narrow very intense peaks because they experience a faster effective correlation time.

Another explanation for the origin of the low temperature UV melting transition is simply the melting out of the base pairing across the splice site. This theory is based on specific heat calculations using the "rnadraw" computer program (Matzura and

Wennborg, 1996), which was derived from the "rnaheat" program (Hofacker, et al., 1994). The program utilizes the partition function algorithm by McCaskill (McCaskill, et al., 1990) and energy parameters from Turner (Turner, et al., 1988), Freier (Freier, et al., 1986) and Jaeger (Jaeger, et al., 1989). The calculations for the rLC55 RNA are shown in figure 1.34 below and support the hypothesis that the splice site base pairing is present, but melts earlier than the rest of the RNA.



**Figure 1.34** Specific heat calculation for rLC55

H<sub>2</sub>O NOESY experiments were collected on the rLC55 sample at cold temperatures (data not shown), in an attempt to see the imino proton resonances of the splice site base pairs. No new iminos were observed at the colder temperatures. While this may suggest that the splice site helix never forms, it may also be that the helix forms transiently and the resonances cannot be seen because the imino protons are exchanging rapidly with the solvent.

### 1.4.3 Form 1 hairpin structure

All three 2D H<sub>2</sub>O NOESY spectra from rLC55, rLC25 and rLC13 show a strong uridine to uridine imino proton crosspeak which has been assigned to the imino protons

of U<sub>26</sub> and U<sub>28</sub>. The <sup>1</sup>H <sup>15</sup>N HMQC from rLC55 (Fig. 1.18) shows a weak extra uridine imino, which we tentatively assigned to U<sub>27</sub>. The existence of that imino proton would strongly argue that the rLC55 RNA (and by association, rLC25 and rLC13) is a duplex RNA, rather than a monomeric hairpin. If the RNAs are duplexes, then the appearance of the U<sub>27</sub> imino is easily accounted for. Due to the symmetry of the duplex, only one of the two U<sub>27</sub> iminos would be visible.

However, other than this extra imino proton and common sense, all the other evidence suggests that the RNAs are monomeric hairpins. The UV melting curves of rLC55 (Fig. 1.16) and rLC13 (Fig. 1.26) do not show any perceptible RNA concentration dependence for the melting temperature and the translational self-diffusion rates of the RNAs are consistent with monomeric hairpins.

If the form 1 hairpin is shown to be a monomer in the high salt conditions, it is intriguing to imagine how the tri-uridine hairpin loop would form. A uridine-uridine base pair will bring the helix backbone closer together as compared to a Watson-Crick base pair because both pairing partners are pyrimidines. This would make it easier for the middle uridine to extend across the phosphates to close the loop. Some simple model building has shown that it is possible to maintain the U=U base pair in this manner. However, it is difficult to imagine how the U<sub>27</sub> imino proton would be protected from solvent exchange in this model.

The experimental evidence that would unambiguously answer the question of whether the RNAs are monomers or dimers is to use the NMR method proposed by Pardi and coworkers (Aboul-ela, et al., 1994). In this scheme, <sup>15</sup>N labeled rLC13 RNA is mixed at a 1:1 ratio with <sup>14</sup>N labeled rLC13. A ½-X-filtered NOESY experiment is used

to collect  $^{14}\text{N}$ - $^{14}\text{N}$ ,  $^{15}\text{N}$ - $^{15}\text{N}$  and  $^{14}\text{N}$ - $^{15}\text{N}$  NOESY subspectra on the mixture. If crosspeaks are found connecting a  $^{14}\text{N}$  labeled imino with a  $^{15}\text{N}$  labeled imino, the sample must exist as a dimer. If the only crosspeaks found connect  $^{14}\text{N}$  with  $^{14}\text{N}$  iminos and  $^{15}\text{N}$  with  $^{15}\text{N}$  iminos, then the sample must be monomeric.

In conclusion, we have shown that the *in vitro* secondary structures of the *L. collosoma* and the *C. fasciculata* SL RNAs exist in the form 1 and form 2, respectively. The tertiary structure from the *L. collosoma* SL RNA is most likely due to melting out of transiently formed base pairing across the splice site, and does not involve interactions between the form 1 hairpin and the rest of the RNA. The form 1 hairpin from *L. collosoma* contains a U=U base pair, and is well behaved spectroscopically. The assignments of both the exchangeable and non-exchangeable protons for the low and high salt forms of the rLC13 RNA have been determined. Further work needs to be done to determine whether this RNA is a monomer or a dimer.

## 1.5 Materials and Methods

A number of methods were employed in the synthesis of RNA molecules discussed in this chapter. Each method will be described in this section, and table 1.1 lists each sample and what method was used in its synthesis.

### 1.5.1 Chemical synthesis of RNA

The HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory provided the chemically synthesized RNAs discussed in this chapter. Each 1  $\mu$ mole RNA synthesis was deprotected by suspension in a 2 mL solution of 1M Tributylammonium fluoride in THF for 48 hrs. This solution was concentrated by speed-vac to a volume of less than 0.5 mL and desalted on a size exclusion column. This desalted solution of RNA was then ethanol precipitated and resuspended in a minimum volume of aqueous 8M urea and purified by standard denaturing poly-acrylamide gel electrophoresis (PAGE).

### 1.5.2 Enzymatic synthesis of RNA

All enzymatically synthesized RNAs were produced from a transcription reaction which utilized a bottom strand DNA template coding for the RNA plus a 5' 17 nucleotide T7 RNA polymerase promoter sequence (Milligan, et al, 1987). The top strand DNA template was complementary to the 17 nucleotide promoter sequence. The T7 RNA polymerase was overexpressed and purified as described previously. All transcription reactions were conducted under identical conditions, except that the magnesium ion concentration was optimized independently for each reaction. The reaction conditions for the transcriptions were typically 40 mM Tris HCl (pH 8.3 @ 20° C), 5 mM DTT, 1mM

spermidine, 20 mM MgCl<sub>2</sub>, 0.01% NP-40, 50 mg/ml PEG 8000, 4 mM in each rNTP (1 mM for <sup>15</sup>N/<sup>13</sup>C labeled), 200nM DNA template, and 0.1 mg/ml T7 RNA polymerase.

All reactions were carried out at 37° C for 4-8 hours and the products of the transcriptions were purified by 15% denaturing PAGE.

### 1.5.3 RNaseH cleavage

One of the major disadvantages of the enzymatic synthesis method of producing RNA using T7 RNA polymerase is that the reaction yields are highly dependent on the 5' end sequence. A method for avoiding this problem is to synthesize an RNA with a high yield 5' end sequence, and use RNase H and a 2'-O-methyl RNA/DNA chimera to direct a site-specific cleavage of the RNA. This reaction is described in detail in chapter 2 of this thesis.

### 1.5.4 List of samples

Table 1.3 lists each of the RNA samples discussed in this chapter and describes which method of synthesis was used to produce it. Molecule names beginning with "rCF" are from the *C fasciculata* SL RNA and those beginning with "rLC" are from the *L. collosoma* SL RNA. Methods of synthesis are abbreviated E=enzymatic, C=chemical and R=RNase H cleavage and are described in other sections of the materials and methods. The sequences of the DNAs used to transcribe the enzymatically synthesized RNA are not given, as they can be inferred from the cDNA sequence to the RNA.

Name	Description	Synthesis Method	Sequence (5'-3')
rCF55	wild type	E	GGGAACU AACGCUAUUAAGUAUCAGUUUC- UGUACUUUAUUGGUAUAAGAAGCUU
rCF30	F2 hairpin	E, R	GUUUCUGUACUUUAUUGGUAUAAGAAGCUU
rCFf1m	F1 mutant <sup>1</sup>	E	GGGAACU AACGCUAUUA <b>ACA</b> AUCAGUUUC- UGUUGUUUAUUGGUAUAAGAAGCUU
rCFf2m	F2 mutant <sup>1</sup>	E	GGGAACU AACGCUAUUAAGUAUC <b>UCUUUC</b> - UGUACUUUAUUGGUAUAAGAAGGAU
rCFf2hp <sup>2</sup>	F2 hairpin and <b>rLDR</b> <sup>2</sup>	E	<b>GGGAUCACACAAUAC</b> GUUUCUGUACUUUAU- UGGUAUAAGAAGCUU
rLC55	wild type	E	GGGAACUAAAACAAUUUUUGAAGAACAGUU- UCUGUACUUCAUUGGUAUGGUAUGUAGAGA
rLC25	F1 hairpin	C	UUGAAGAACAGUUUCUGUACUUCAU
rLC13	F1 hairpin <sup>3</sup>	C	AACAGUUUCUGUA
or rLC13	F1 hairpin <sup>3</sup>	C	UACAGUUUCUGUA

<sup>1</sup> Mutations from wild type are shown in bold.

<sup>2</sup> rLDR portion of the sequence shown in bold, see Chapter 2 for a discussion of the rLDR sequence.

<sup>3</sup> Notice that this RNA was synthesized with a "A" and a "U" at the 5' end. The sequence change had no noticeable affect on any of the RNAs characteristics.

### Table 1. 3 RNA samples

#### 1.5.5 Optical equilibrium-melting curves

Nucleic acids have a strong UV absorbance at 260 nm because of the conjugated ring structures in the bases. The extinction coefficient for each nucleotide is somewhat dependent on the local environment in which the nucleotide exists. It so happens that the UV absorbance of a nucleotide is lower when it is involved in an RNA double strand helix, because of the tight stacking of the bases. When the nucleic acid is thermally induced to "melt" out of the helix to an unstructured single strand the UV absorbance rises in what is known as a hypochromic shift. The deflection point of the derivative of the change in UV absorbance with respect to temperature is known as the melting temperature ( $T_m$ ). The  $T_m$  and the shape of the melting profile may be used to calculate

the thermodynamic parameters of the nucleic acid (Gralla and Crothers, 1973; Puglisi and Tinoco, 1989).

All equilibrium-melting curves were collected on a Varian Cary 1 UV spectrophotometer. The samples were heated to above their melting temperature and either snap-cooled (by placing in ice/water or dry ice/isopropanol) or slow-cooled. The melts were carried out by first cooling the sample to 5 °C, then the temperature was raised by 0.5 to 1 °C per minute. The UV absorbance was collected every 1 °C. Data were processed and analyzed statistically using the software package Origin v4.1 (Microcal Software Inc, USA).

#### 1.5.6 NMR Methods

Homonuclear and heteronuclear NMR data presented in this chapter were collected on either a Varian Unity 500 or Unity+ 600 spectrometer. Most samples were dialyzed extensively against 20 mM sodium phosphate buffer (pH 6.5) and 1 mM EDTA. The high salt buffers typically included 100-200 mM sodium chloride, the low salt buffers typically included 0-50 mM sodium chloride. The D<sub>2</sub>O experiments were conducted using 99.996% D<sub>2</sub>O and the H<sub>2</sub>O experiments used 15% D<sub>2</sub>O. Unless otherwise stated, 1024 complex points were collected in the direct dimension, and 300 points in the indirect dimension. Quadrature in the indirect dimension was accomplished using the States method. Data processing was performed using the software package Felix (Biosym Inc.). Unless stated differently a 90 degree shifted sine-bell was used to apodize the FIDs before Fourier transformation.

Spectra from the H<sub>2</sub>O NOESY experiments were collected using either the Jump Return Spin-Echo NOESY (JRSE-NOESY) pulse sequence or the Watergate

NOESY (WNOESY) pulse sequence (Piotto, et al., 1992; Lippens, et al., 1995; Sich, et al., 1996). Typically, the sweep width was set to 10,000 Hz on a 500 MHz spectrometer to insure complete coverage of the imino protons (14-10 ppm) and the offset frequency was centered on the H<sub>2</sub>O line (4.75 ppm).

Spectra from the DQFCOSY experiments were collected using the canned `dqfcosy.c` pulse sequence supplied with the Varian spectrometers. The sweep width was set to 5000 Hz in each dimension on the 500 MHz spectrometers to insure coverage of the aromatic region of the spectrum (8-7 ppm) and the offset was centered on the residual HDO line (4.75 ppm). Data were processed by apodizing the FID with a zero degree shifted sine-bell.

Spectra from the D<sub>2</sub>O NOESY experiments were collected using a modified version of the canned `noesy.c` pulse sequence supplied with the Varian spectrometers. The modification was to add a gradient pulse to the mixing time, to destroy any transverse magnetization because of single or double-quantum coupling. A low power 0.5 second water presaturation pulse was used to remove the residual HDO line. Typically, 2-10 second recycle delays were utilized. The sweep width was set to 5000 Hz in each dimension on a 500 MHz spectrometer to insure coverage of the aromatic region of the spectrum (8-7 ppm) and the offset was centered on the residual HDO line (4.75 ppm).

The HMQC pulse sequence was derived from that published by Szewczak (1993). The <sup>15</sup>N carrier frequency was set to 150 ppm to center on the imino proton nitrogens.

The translational self-diffusion experiments were performed (and the gradients were calibrated) as discussed in chapter 4 of this thesis. The pfg\_diffusion.c pulse sequence (see 4.6.1) was used to collect the data, setting  $\Delta=0.1\text{s}$  and  $\delta=0.004\text{s}$  (other values were examined as well, with no effect on the results). 32 experiments were collected arraying the gradient strength from 0 to 31 G/cm. The processing of the data was performed using the Felix95 software package using the macro diffusion.mac (see 4.6.2). The resultant "xy" file was further processed using the xy2xm script (see 4.6.3) using a maximum gradient strength value of 32 G/cm. The final "xm" file was then graphed using the Origin v4.1 statistical software package (Microcal Software Inc, USA). Reported values of the translational self-diffusion rate and the error in the measurement come directly from the built-in linear regression package.

## 1.6 References

- Aboul-ela F, Nikonowicz EP, Pardi A. 1994. Distinguishing between duplex and hairpin forms of RNA by  $^{15}\text{N}$ - $^1\text{H}$  heteronuclear NMR. *FEBS Lett* 347:261-264.
- Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothroyd JC. 1992. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem* 267:9805-9815.
- Blumenthal T, Thomas J. 1988. *Cis* and *trans* mRNA splicing in *C. elegans*. *Trends in Genetics* 4:305-308.
- Brucoleri R, Heinrich G. 1988. An improved algorithm for nucleic acid secondary structure display. *Computer Applications in the Biosciences* 4:167-173.
- Bruzik JP, Doran KV, Hirsh D, Steitz JA. 1988. *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature* 335:559-562.
- Clayton C. 1992. Developmental regulation of nuclear gene expression in *Trypanosoma brucei*. *Prog Nucleic Acid Res Mol Biol* 43:37-66.
- Cload ST, Richardson PL, Huang YH, Schepartz A. 1993. Kinetic and thermodynamic analysis of RNA binding by tethered oligonucleotide probes: Alternative structures and conformational changes. *JACS* 115:5005-5014.
- Cross M, Günzl A, Palfi Z, Bindereif A. 1991. Analysis of small nuclear ribonucleoprotein (RNPs) in *Trypanosoma brucei*: Structural organization and protein components of the spliced leader RNP. *Mol Cell Biol* 11:5516-5526.
- Freier SM, Kiezerk R, Jaeger JA, Sugimoto N, Caruthers MH, Nelson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *PNAS* 83:9373-9377.
- Freistadt MS, Cross GAM, Robertson HD. 1988. Discontinuously synthesized mRNA from *Trypanosoma brucei* contains the highly methylated 5' cap structure,  $\text{m}^7\text{GpppA}^*\text{A}^*\text{C}(2'\text{-O})\text{mU}^*\text{A}$ . *J Biol Chem* 263:15071-15075.
- Gralla J, Crothers DM. 1973. Free Energy of Imperfect Nucleic Acid Helices. III. Small Internal Loops Resulting from Mismatches. *J. Mol. Biol.* 78:301-319.
- Harris K. 1995. Biochemical analysis of trypanosomatid SL RNA structure.
- Harris KA, Crothers DM, Ullu E. 1995. In Vivo Structural Analysis of Spliced Leader RNAs in *Trypanosoma brucei* and *Leptomonas collosoma*: A Flexible Structure that Is Independent of Cap4 Methylations. *RNA* 1:351-362.

- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. *Chemical Monthly* 125:167-188.
- Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. *PNAS* 86:7706-7710.
- Karplus M. 1959. *J. Chem. Phys.* 30:11.
- Konarska MM, Padgett PA, Sharp PA. 1985. *Trans* splicing of mRNA precursors *in vitro*. *Cell* 42:165-171.
- Krause M, Hirsh D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49:753-761.
- Lapham J, Rife J, Moore PB, Crothers DM. 1997. Measurement of diffusion constants for nucleic acids by NMR. *J. Biomolecular NMR* 10:255-262.
- LeCuyer KA. 1992. Conformational dynamics of the *L. collosoma* spliced leader RNA.
- LeCuyer KA, Crothers DM. 1993. The *Leptomonas collosoma* Spliced Leader RNA Can Switch between Two Alternate Structural Forms. *Biochemistry* 32:5301-5311.
- LeCuyer KA, Crothers DM. 1994. Kinetics of an RNA Conformational Switch. *PNAS* 91:3373-3377.
- Lippens G, Dhalluin C, Wieruszski JM. 1995. Use of the water flip-back pulse in the homonuclear NOESY experiment. *J Biomol NMR* 5:327-331.
- Maniatis TM, Reed R. 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* 325:673-678.
- Matzura O, Wennborg A. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Computer Applications in the Biosciences* 12:247-249.
- McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 29:1105-1119.
- McNally KP, Agabian N. 1992. *Trypanosoma brucei* spliced leader RNA methylations are required for *trans*-splicing *in vivo*. *Mol Cell Biol* 12:4844-4851.
- Michaeli S, Roberts TG, Watkins KP, Agabian N. 1990. Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. *J Biol Chem* 265:10582-10588.

- Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide Synthesis using T7 RNA Polymerase and Synthetic DNA Templates. *NAR* 15:8783-8798.
- Murphy WJ, Watkins KP, Agabian N. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell* 47:517-525.
- Padgett PA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. 1986. Splicing of messenger RNA precursors. *Ann Rev Biochem* 55:1119-1150.
- Perry KL, Watkins KP, Agabian N. 1987. Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *PNAS* 84:8190-8194.
- Piotto M, Saudek V, Sklenar V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661-665.
- Puglisi JD, Ignatio Tinoco J. 1989. Absorbance melting curves of RNA. *Meth in Enzymology* 180:304-325.
- Sich C, Flemming J, Ramachandran R, Brown LR. 1996. Distinguishing Inter- and Intrastrand NOEs Involving Exchangeable Protons in RNA Duplexes. *J Mag Res Series B* 112:275-281.
- Sutton PE, Boothroyd JC. 1986. Evidence for *trans* splicing in trypanosomes. *Cell* 47:527-535.
- Szewczak AA, Kellogg GW, Moore PB. 1993. Assignment of NH Resonances in Nucleic Acids Using Natural Abundance  $^{15}\text{N}$ - $^1\text{H}$  Correlation Spectroscopy with Spin-Echo and Gradient Pulses. *FEBS Lett* 327:261-264.
- Turner DH, Sugimoto N, Freier SM. 1988. RNA Structure Prediction. *Ann Rev Biophys Chem* 17:167-192.
- Turner DH, Sugimoto N, Jaeger JA, Longfellow CE, Freier SM, Kierzek R. 1987. Improved parameters for prediction of RNA structure. *Cold Spring Harb Symp Quant Biol* 52:123-133.
- Ullu E, Tschudi C. 1991. *Trans* splicing in trypanosomes requires methylation of the 5' end of the spliced leader RNA. *PNAS* 88:10074-10078.
- Ullu E, Tschudi C. 1993. 2'-O-methyl RNA oligonucleotides identify two functional domains in the trypanosome spliced leader ribonucleoprotein particle. *J Biol Chem* 268:13068-13073.

Ullu E, Tschudi C, Günzl A. 1995. *Trans-splicing in trypanosomatid protozoa*. Oxford, UK: Oxford University Press.

Xu J, Lapham J, Crothers DM. 1996. Determining RNA Solution Structure by Segmental Isotopic Labeling and NMR: Applications to *Caenorhabditis elegans* Spliced Leader RNA. *PNAS* 93:44-48.

Zuker M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* 244:48-52.

Zuker M, Stiegler P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *NAR* 9:133-148.