

# The structure of the L3 loop from the hepatitis delta virus ribozyme: a *syn* cytidine

Stephen R. Lynch and Ignacio Tinoco Jr\*

Department of Chemistry, University of California and Structural Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720-1460, USA

Received October 17, 1997; Revised and Accepted December 15, 1997

## ABSTRACT

The structure of the L3 central hairpin loop isolated from the antigenomic sequence of the hepatitis delta virus ribozyme with the P2 and P3 stems from the ribozyme stacked on top of the loop has been determined by NMR spectroscopy. The 26 nt stem-loop structure contains nine base pairs and a 7 nt loop (5'-UCCUCGC-3'). This hairpin loop is critical for efficient catalysis in the intact ribozyme. The structure was determined using homonuclear and heteronuclear NMR techniques on non-labeled and <sup>15</sup>N-labeled RNA oligonucleotides. The overall root mean square deviation for the structure was 1.15 Å (± 0.28 Å) for the loop and the closing C-G base pair and 0.90 Å (± 0.18 Å) for the loop and the closing C-G base pair but without the lone purine in the loop, which is not well defined in the structure. The structure indicates a U-C base pair between the nucleotides on the 5'- and 3'-ends of the loop. This base pair is formed with a single hydrogen bond involving the cytosine exocyclic amino proton and the carbonyl O4 of the uracil. The most unexpected finding in the loop is a *syn* cytidine. While not unprecedented, *syn* pyrimidines are highly unusual. This one can be confidently established by intranucleotide distances between the ribose and the base determined by NMR spectroscopy. A similar study of the structure of this loop showed a somewhat different three-dimensional structure. A discussion of differences in the two structures, as well as possible sites of interaction with the cleavage site, will be presented.

## INTRODUCTION

The self-cleaving RNA sequences isolated from the genomic and anti-genomic RNAs of hepatitis delta virus (HDV) (1,2), which act during rolling circle type replication of the virus, adopt a pseudoknot-like secondary structure (3,4), as shown in Figure 1a. The sequence shown is a catalytically active *trans*-acting ribozyme based upon the genomic and anti-genomic ribozyme sequences (5). This secondary structure is well supported by analysis of sequence variation between the genomic and anti-genomic sequences in order to maintain base pairing (3,4). In addition, mutations of the proposed double-helical regions demonstrates a dependence of the activity of the ribozyme that can be reversed with compensatory mutations in both the genomic and antigenomic

sequences (3,5-7). The pyrimidine-rich loop or L3 loop shown in Figure 1b is a section of the anti-genomic sequence of the hepatitis delta virus ribozyme critical for catalytic activity. Changing the first 2 nt of the loop (U11 and C12) results in a >1000-fold decrease in activity (8). A change in the one purine in the loop also results in a significant decrease in activity (~50-fold), though not as dramatic as changes in the nucleotides on the 5'-side of the loop (8). Thus the L3 loop is essential for activity, either as an integral part of the active site or through interactions with other sequences in the ribozyme.

The sequence requirements, metal ion requirements, catalytic function and structural features of the ribozyme have been well characterized through chemical and biochemical means, but a high resolution structure of it has yet to be determined. While the 85 nt *cis*-acting ribozyme or the 60 nt *trans*-acting ribozyme (9) is small in comparison with other RNA molecules, it is still large for high resolution structural studies by NMR spectroscopy. Our goal is to first determine a structure for the L3 loop and then use the NMR assignments and its three-dimensional structure to aid in structure determination of the entire ribozyme. In the process we will identify possible effects of tertiary interactions on the three-dimensional structure of this loop. In this study we applied NMR spectroscopy followed by restrained molecular dynamics to determine the conformation in solution of the 7 nt L3 loop from the anti-genomic sequence of the hepatitis delta virus ribozyme with the 3 nt P3 stem stacked on the loop and the 6 nt P2 stem stacked on the P3 stem.

## MATERIALS AND METHODS

### RNA synthesis and purification

The L3 loop from the antigenomic sequence of the hepatitis delta virus ribozyme, 5'-GGGUCCAGCCUCCUCGCGGCUGGA-UC-3', was enzymatically synthesized and purified as described previously (10). <sup>15</sup>N-Labeled NMPs were isolated from *Escherichia coli* grown on minimal medium containing <sup>15</sup>NH<sub>4</sub>Cl as the only source of nitrogen. The total cellular RNA was purified (11), then digested into NMPs (12,13). The NMPs were phosphorylated to form NTPs for *in vitro* transcription (14).

### NMR spectroscopy

RNA samples were dialyzed extensively against 2 l buffer. The buffer used in most NMR experiments was 10 mM sodium phosphate, 100 mM NaCl, 0.05 mM EDTA, pH 6.2. NaCl concentrations were varied from 20 to 200 mM and tested by 1D

\*To whom correspondence should be addressed. Tel: +1 510 642 3038; Fax: +1 510 643 6232; Email: intinoco@lbl.gov

NMR and by DQF-COSY to obtain optimal experimental conditions.  $MgCl_2$  was added to 10 mM to the NMR sample after dialysis for some NMR experiments, but changes in the NMR spectra were not observed. Structure calculations were done on spectra acquired in the absence of  $MgCl_2$ . Samples in  $H_2O$  and  $D_2O$  were prepared as described (10). The final sample concentration was 2–2.5 mM for non-labeled samples and 1.5 mM for the  $^{15}N$ -labeled sample. NMR experiments (10) were acquired on either a GE GN-500 MHz spectrometer, a 600 MHz Bruker AMX spectrometer or a 300 MHz Bruker AMX spectrometer (for  $^{31}P$  experiments only). Typically 400–500 FIDs of 2048 complex points and 80 scans were collected with a relaxation delay of 2.0 s.

NOESY spectra (15) in  $D_2O$  were recorded with mixing times of 50, 100, 150, 200 and 400 ms at primarily 25°C, but also at 15, 20 and 30°C due to spectral overlap. NOESY spectra in  $H_2O$  were recorded with mixing times of 100, 200 and 300 ms at temperatures of 5, 10 and 15°C with a spectral width of 12 500 Hz using a 1-1 water suppression sequence to replace the last pulse (16). DQF-COSY spectra were collected using the pulse sequence with composite mixing pulses (17). For high resolution DQF-COSY broad band phosphorus decoupling was achieved by GARP1 (18). TOCSY experiments were acquired using a MLEV-17 for broad band decoupling (19), with mixing times of 60 and 80 ms.

A 3D homonuclear TOCSY-NOESY (20) was acquired with a spectral width of 3311 Hz in all three dimensions; eight scans of 512 points in  $t_3$ , 138 in  $t_2$  and 128 in  $t_1$  were acquired with a TOCSY mixing time of 65 ms and a NOESY mixing time of 150 ms.

Natural abundance  $^{13}C$ - $^1H$  HMQC (21) was acquired with a proton spectral width of 5000 Hz and a carbon spectral width of 7550 Hz with GARP1 carbon decoupling during acquisition (18) as previously described (10).

A proton-detected heteronuclear  $^{31}P$ - $^1H$  COSY (22) was acquired with a spectral width of 600 Hz in the  $^{31}P$  dimension and 1400 Hz in the  $^1H$  dimension. 138 FIDs of 4096 complex points and 144 scans were acquired.

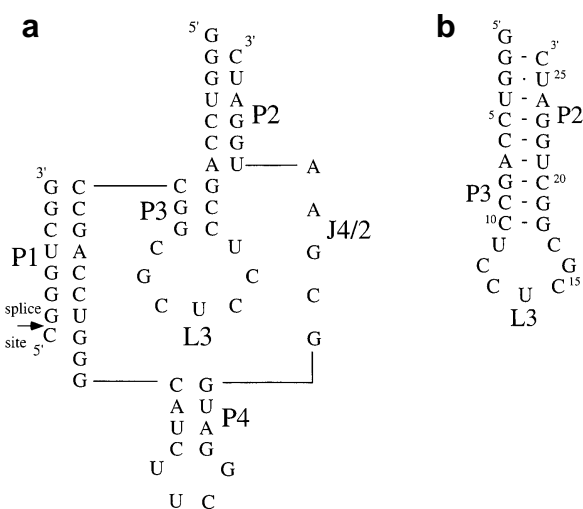
A  $^1H$ - $^{15}N$  HMQC with 1-1 solvent suppression was acquired on a uniformly labeled  $^{15}N$  sample in  $H_2O$  (23). A 2D  $^{15}N$ -correlated NOESY experiment was acquired on this sample in  $H_2O$  at 5°C as described (24). A 2D multiple bond  $^1H$ - $^{15}N$  heteronuclear single quantum coherence (25) was acquired in  $D_2O$  at 25°C.

### NOE distance constraints

The cross-peak intensities in NOESY experiments acquired at 50, 100, 150 and 200 ms were estimated and used to determine distance constraints as described previously (10). A total of 362 distance constraints were determined for the 25 nt stem-loop structure, of which 80 distance constraints were in the 7 nt loop and 49 of those 80 NOEs were internucleotide NOEs in the loop.

### Dihedral constraints

A 3'-endo conformation was assigned to the ribose units of nucleotides G2–U11 and C17–U25 based on the absence of  $H1'-H2'$  cross peaks in a DQF-COSY, indicating a weak  $^3J_{H1'-H2'}$ . C13, U14 and C15 were assigned 2'-endo conformations based on  $^3J_{H1'-H2'} > 8$  Hz. C12, G16 and C26 were assigned a broad range of C2'-endo to O4'-endo to C3'-endo based upon  $^3J_{H1'-H2'}$  between 2 and 8 Hz. The glycosidic torsion angle  $\chi$  was assigned as *anti* ( $110 \pm 70^\circ$ ) for all nucleotides except for C15, based upon



**Figure 1.** (a) The predicted secondary structure of a *trans*-acting ribozyme that was developed by Been *et al.* (5) based upon the genomic and antigenomic sequences of the hepatitis delta virus ribozyme. (b) The secondary structure of the L3 central hairpin loop of the hepatitis delta virus ribozyme with the P2 and P3 stems stacked on the loop.

the lack of a strong intranucleotide  $H1'-H8/6$  NOE. C15 was defined as *syn* ( $30 \pm 40^\circ$ ). During refinement, loose backbone torsion angles constraints ( $\pm 20^\circ$ ) for  $\beta$ ,  $\gamma$  and  $\epsilon$  were added to the torsion angle constraint file for G2–C10 and G18–C26; constraints were derived from standard values for A-form helices (26). Additionally,  $\alpha$  and  $\zeta$  were loosely constrained ( $\pm 30^\circ$ ) in the final minimization for the same nucleotides. No non-experimental dihedral constraint was used for any nucleotides in the 7 nt loop.

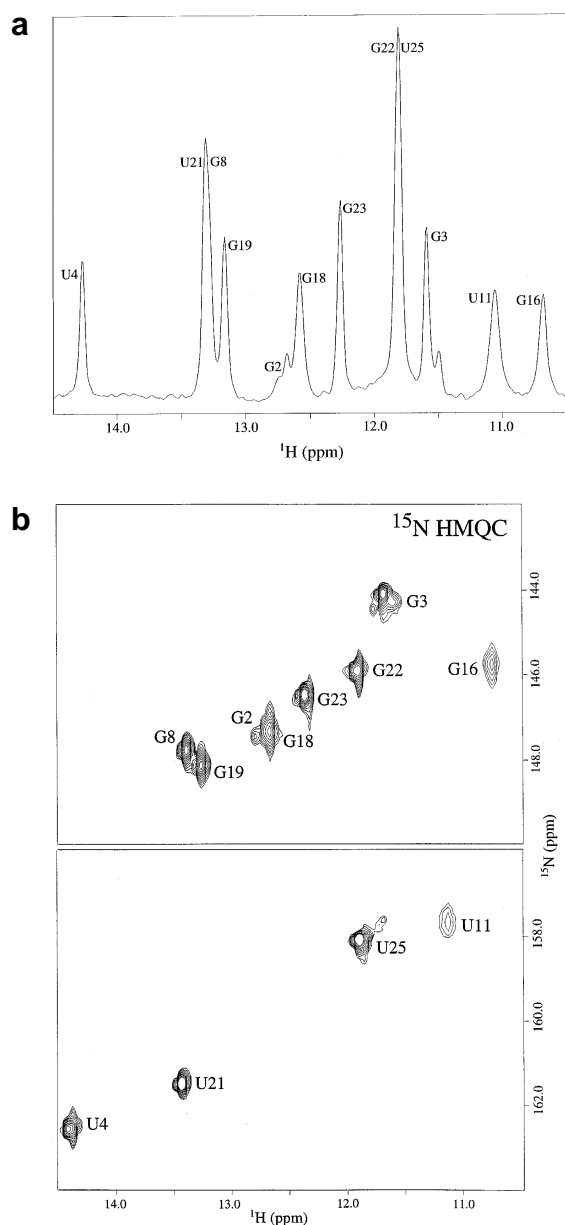
### Structure determination

Structure calculation on the stem-loop structure was accomplished by restrained molecular dynamics followed by energy minimization using the X-PLOR program (27) with a force field consisting of bond lengths, bond angles, improper angles, repulsive van der Waals potentials and the experimental distance and torsion angle constraints from the NMR data as described previously (10). The calculated structures were visualized with InsightII 95.0 (Biosym Technologies).

## RESULTS

### Exchangeable proton spectral assignment

The  $H_2O$  exchangeable proton NMR spectrum of the 26 nt stem-loop RNA molecule containing the P2 and P3 stems and L3 loop (Fig. 1b) of the *trans*-acting hepatitis delta virus ribozyme (Fig. 1a) was assigned with  $^1H$ -only experiments on an unlabeled NMR sample and with  $^{15}N$ -correlation experiments on an isotopically enriched  $^{15}N$  sample. The RNA molecule contains nine guanosine and five uridine residues; the 1D imino proton spectrum shown in Figure 2a contains 12 resonances of the possible 14. The nine stem base pairs can be easily assigned with a NOESY in  $H_2O$ . After assignment of the 10 imino protons in the 9 bp stem two observed imino protons remain. The 5'-terminal guanosine, which is unpaired, exchanges rapidly with the solvent and is not observed. From the imino nitrogen chemical shifts a  $^{15}N$ -HMQC spectrum (see Fig. 2b) shows that the resonance at 10.7 p.p.m. is definitely G16 and the one at 11.1 p.p.m. is a



**Figure 2.** (a) The region of the 1D  $^1\text{H}$  NMR spectrum in  $\text{H}_2\text{O}$  that contains the imino protons of guanosine and uridine of the L3 loop and P2 and P3 stems of the hepatitis delta virus ribozyme is shown. The peaks are labeled with the resonance assignments as determined by NOESY spectra in  $\text{H}_2\text{O}$ . The spectrum was acquired at  $5^\circ\text{C}$  with 4096 complex points, 80 scans using 1-1 solvent suppression on a 2 mM RNA sample in 10 mM sodium phosphate, 50 mM NaCl and 100  $\mu\text{M}$  EDTA at pH 6.0. (b) A  $^{15}\text{N}$ -HMQC is shown on a 1.5 mM fully  $^{15}\text{N}$ -labeled sample of the L3 loop with the P2 and P3 stems. The nitrogen shift of the imino nitrogen of guanine is up-field from that of uracil. Thus the resonance at 10.7 p.p.m. can be assigned as G16 and the resonance at 11.1 p.p.m. is assigned as U11. The data were acquired with 1-1 solvent suppression with the maximal excitation set to 12.5 p.p.m. for protons. The proton carrier was set to 4.8 p.p.m., the nitrogen carrier to 150 p.p.m. The proton spectral width was 12 500 Hz; the nitrogen sweep width was 1824 Hz. 128 FIDs of 1024 complex points and 16 scans were acquired.

uridine. Since there are two uridines in the loop, the resonance at 11.1 p.p.m. could be either U11 or U14. This resonance is assigned as U11 because of three NOEs observed, one to each amino proton on C17 and one to the H5 of C17. This assignment

was confirmed by a long range HSQC in  $\text{D}_2\text{O}$  (25) where a through-bond correlation is shown between H5 and N3.

Amino protons were assigned through observation of NOEs to imino and non-exchangeable protons using a NOESY in  $\text{H}_2\text{O}$  and a  $^{15}\text{N}$ -correlated NOESY (24). The amino protons on the terminal residue in the loop, C17, were also assigned by observation of internucleotide NOEs to the G18 imino proton and to the U11 imino proton and an intranucleotide NOE to the C17 H5 proton. Similarly, the amino protons of the other cytidine residues in the loop were assigned by observation of intranucleotide NOEs to the H5 proton. The amino protons of G16 were not observed.

### Non-exchangeable proton spectral assignment

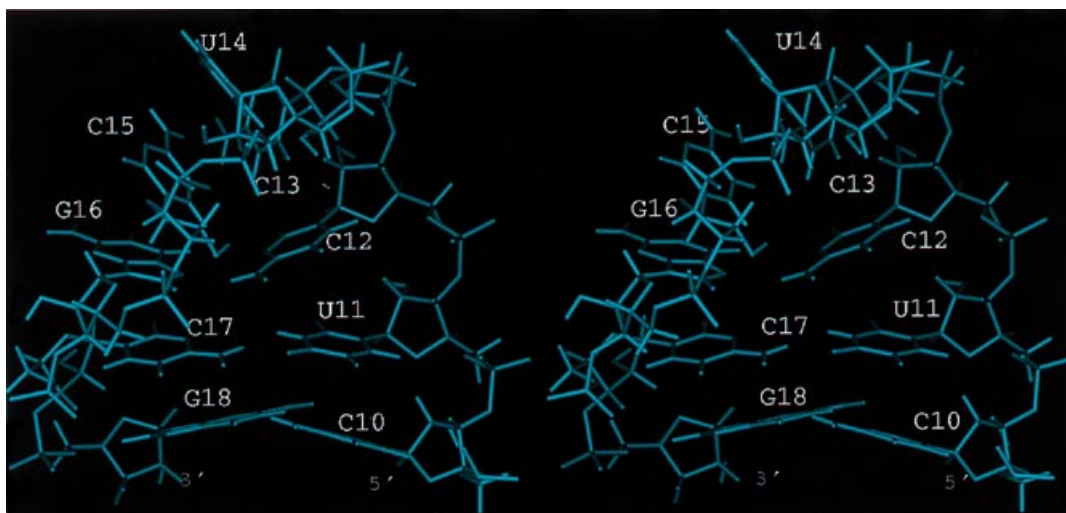
The  $\text{D}_2\text{O}$  non-exchangeable proton NMR spectrum of the 26 nt stem-loop RNA molecule (Fig. 1b) was assigned with a combination of  $^1\text{H}$ -only,  $^1\text{H}/^{13}\text{C}$ ,  $^1\text{H}/^{15}\text{N}$  and  $^1\text{H}/^{31}\text{P}$  experiments on either an unlabeled or uniformly  $^{15}\text{N}$ -labeled NMR sample. NOESY experiments were acquired at several temperatures to resolve spectral overlap. H2' and H3' protons were assigned through H2'/aromatic and H3'/aromatic NOE walks respectively. NOESY experiments at short mixing times (50 ms) were used for H2' assignments since the internucleotide H2'/aromatic and intranucleotide H1'/H2' NOEs are strong. A  $^1\text{H}/^{31}\text{P}$  heteronuclear COSY was acquired to help assign the H3' protons and, when possible, the H4' and H5' protons, but poor dispersion in the phosphorus dimension of the stem prevented walking from one nucleotide to the next. A 3D homonuclear TOCSY-NOESY was applied to further aid assignment, particularly pyrimidine H5 and H6 protons, H1' protons on C2'-endo riboses and H3', H4', H5' and H5''. In addition, a  $^{15}\text{N}$  long range HSQC was acquired on a uniformly  $^{15}\text{N}$ -labeled sample to confirm H1' assignments by correlation of the N1 of pyrimidines to the H1', H5 and H6 protons and by correlation of N9 of purines to the H1' and H8 protons. Furthermore, this experiment can correlate H2' protons of a C2'-endo ribose to the N1 of a pyrimidine or the N9 of a purine. The proton assignments of loop resonances are in agreement with the assignments recently published on this loop (28). Additional assignments are given in the supplementary material for  $^{15}\text{N}$ ,  $^{13}\text{C}$  and exchangeable  $^1\text{H}$ .

### Structure determination

Restrained molecular dynamics, with NMR distance constraints and dihedral constraints for sugar pucker and  $\chi$ , was used to obtain global folds from 50 starting random structures. The constraints used were: 201 inter-residue, 151 intra-residue, 37 hydrogen bond and 230 torsion angle. There were 32 structures within 12% of the lowest total energy; they were defined as converged. The first structure excluded was 22% higher than the lowest energy structure. These 32 structures were subjected to structural refinement with all distance and dihedral constraints and no electrostatics. The 10 final structures discussed were those with the lowest NOE energies, indicating that these structures best support the NMR data. These 10 structures had NOE energies that were 10% different from each other, but 40% lower than the remainder of the structures.

The r.m.s. deviations from ideal covalent geometry for the 10 structures were: 0.004  $\text{\AA}$  for bond lengths,  $0.86^\circ$  for bond angles and  $0.24^\circ$  for improper angles. None of the 10 structures had more than one NOE violation  $>0.2 \text{\AA}$  and there were no torsion angle violations  $>10^\circ$ . The 10 structures have an r.m.s. deviation





**Figure 3.** The lowest energy structure of the loop and closing C-G base pair is shown. Normal stacking is observed on the 5'-side of the loop through C12 and resumes on the 3'-side of the loop at C17.

from the average structure for the loop and the closing C-G base pair of  $1.15 \pm 0.28$  Å. The only purine in the loop, G16, is not well defined, as expected, since few NOE constraints were identified. The average structure was recalculated excluding G16. Again, the 10 lowest energy structures were compared with the average and the calculated r.m.s. deviation for the loop with the closing C-G base pair, excluding G16, was  $0.90 \pm 0.18$  Å. The lowest energy structure is shown in Figure 3; it is very similar to the average structure.

#### Structural features of the pyrimidine-rich L3 loop from the hepatitis delta virus ribozyme

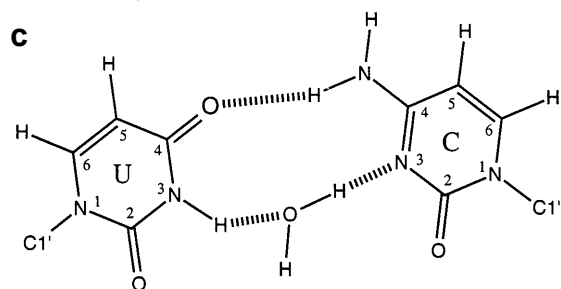
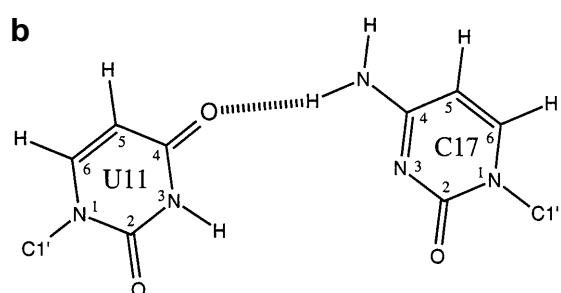
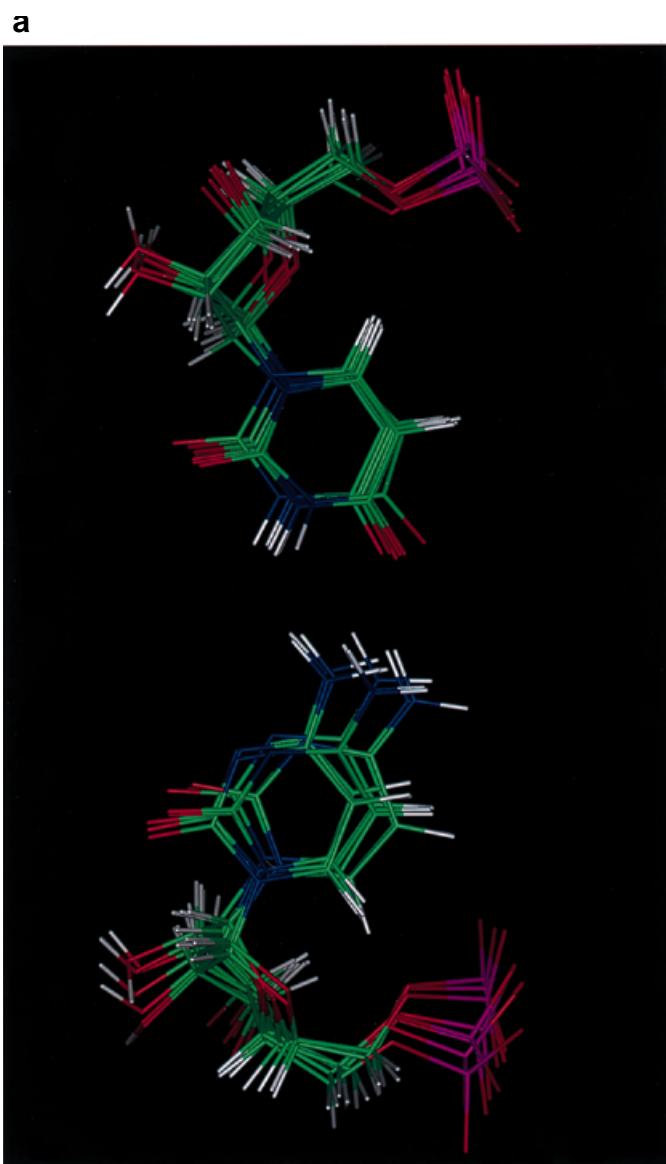
Two imino proton resonances are observed in the 1D spectrum (Fig. 2a) that do not arise from stem imino protons. These loop protons are thus protected from exchange with solvent, and are possibly, but not necessarily, hydrogen bonded. The assignment of the resonance at 11.1 p.p.m. in Figure 2a to U11 and the assignment of the amino protons on C17 in combination with NOE data suggest that C17 is stacked into the helix with the U11 imino proton close to the C17 amino proton. In the non-exchangeable spectra the normal aromatic/H1' NOE walk and the aromatic/H2' NOE walk continue from C10 to U11, while on the 3'-side of the loop the normal NOE walks start again at C17–G18. These data suggest a base pair between U11 and C17. The U11 imino proton resonance is broader than those arising from stem imino protons and its resonance further broadens at higher temperatures or higher pH. This indicates that the U11 imino proton is exchanging faster with solvent than the stem imino protons. The faster exchange with solvent suggests that either a base pair is present between U11 and C17 that is more dynamic than a stem base pair or a base pair is formed between U11 and C17 that does not involve hydrogen bonding to the U11 imino proton. Structural refinement suggests the latter, as seen in Figure 4a; the 10 lowest energy structures are presented with just U11 and C17 superimposed. The r.m.s. deviation from the average for just these two nucleotides is  $0.56 \pm 0.18$ . While C17 is not as well defined as U11, the amino

proton on C17 is always within hydrogen bonding distance to the O4 of U11. However, the distance from the U11 imino proton to the C17 O4 is always  $>3$  Å. Although the U11 imino proton is not hydrogen bonded in the structure, it is stacked within the loop and thus is shielded from the solvent. That produces a moderately broad imino proton in 1D spectra.

C12 faces across the helix as shown in Figure 3. The r.m.s. deviation for the 3 nt ensemble of U11, C12 and C17 is  $0.61 \pm 0.18$  Å. The superposition of structures demonstrates that the exocyclic amino proton of C12 is within hydrogen bonding distance to the N3 on C17 in some structures, suggesting a possible hydrogen bond. The average distance between the amino proton on C12 and the N3 on C17 is  $2.95 \pm 1.05$  Å, with a minimum distance of 2.13 Å and a maximum distance of 4.65 Å. However, the C12 base is not co-planar with C17 like U11, indicating that C12 and C17 could not form a good hydrogen bond. The position of C12 relative to C17 is primarily dependent upon the location of G16, which is not well defined. This uncertainty prevents a definitive hydrogen bonding scheme for C12 to C17.

C10 and G18 form a Watson–Crick base pair as expected. However, the base pair is buckled, as shown in Figure 3. The buckling results from the non-Watson–Crick base pair that is stacked on top of it. The U–C pair creates some distortion in the backbone, causing the C–G pair not to be planar. The imino proton on G18, as shown in Figure 2a, is not as sharp as the other stem imino protons, indicating a faster exchange rate with solvent. The faster exchange rate probably results from buckling of the base pair and from the lack of Watson–Crick base pairs on both sides of the C–G, as the U–C base pair with only one hydrogen bond is not as stable.

The turn in the backbone of the loop occurs between C13, U14 and C15, as shown in Figure 3. The r.m.s. deviation of 10 structures for these 3 nt alone is  $0.50 \pm 0.11$  Å. The bases on all 3 nt are turned into the loop, as there are many NOE contacts to the aromatic groups of all 3 nt. No obvious hydrogen bonding interactions are observed in the structures.



### Conformation of C15

The fifth nucleotide in the loop, C15, has a strong intranucleotide NOE from the H1' proton on the ribose to the H6 proton on the base, as shown in the 50 ms mixing time NOESY presented in Figure 5, with the aromatic/H1' region shown in the bottom panel and the sugar proton to aromatic proton region shown in the top panel. The NOE highlighted with a box in the bottom panel is the intranucleotide NOE of C15 from the H1' at 5.70 p.p.m. to the H6 at 7.65 p.p.m. The other intense peaks in this spectrum result from either intranucleotide H5/H6 NOEs on the pyrimidine bases, which have a fixed distance of 2.4 Å, or internucleotide NOEs from adenosine H2 protons, which are the sharpest peaks in the spectrum.

In the top panel the NOE highlighted in the box is the intranucleotide NOE on C15 from the H2' proton at 4.27 p.p.m. to the H6 proton at 7.65 p.p.m. This NOE might be expected to be one of the most intense NOEs in the spectrum because C15 has a C2'-endo sugar pucker, causing the H2' proton to be ~2.0 Å from the H6 proton on the pyrimidine base. However, this NOE is not as intense as others. The H2'/H6 distance is dependent upon two primary torsion angles: the sugar pucker, which was established to be C2'-endo in a DQF-COSY, and the glycosidic bond.

The torsion angle about the glycosidic bond is normally *anti*. A strong intranucleotide H1'/H6 NOE suggests that the torsion angle is either high-*anti* or *syn*. An estimation of the torsion angle can be deduced from a ratio of NOE volumes of three intranucleotide NOEs: H1'/H6, H2'/H6 and H3'/H6. Although spectral overlap prevented a good quantitative evaluation of H3'/H6, the other two cross-peaks are well resolved. The results are shown in Table 1. The volumes of the two NOEs are compared with values for the intranucleotide H5/H6 on C15. The volume of the H1'/H6 NOE for C15 is consistently larger than that of the H2'/H6 NOE, as expected for a *syn* conformation. In contrast, for C13, which also has a C2'-endo conformation, the H1'/H6 NOE is always smaller than H2'/H6, consistent with an *anti* conformation.

The calculated distances from H1' to H6, H2' to H6 and H3' to H6 are shown in Figure 6 as a function of the glycosidic torsion angle for a cytidine with a C2'-endo sugar pucker. A value of ~20° for the glycosidic torsion angle is most consistent with the NOE volume data presented, since the H1'/H6 is greater than the H2'/H6 NOE but weaker than H5/H6. In addition, the H3'/H6 NOE is observed only at longer mixing times, indicating that H3' is not as close to H6 as either H1' or H2'.

Furthermore, <sup>13</sup>C chemical shift data concur with the NOESY data. C15 has the most down-field shifted C6-aromatic carbon resonance at 144 p.p.m., consistent with an unusual conformation. While no data exist on the C6 chemical shift of *syn* cytidine nucleotides in RNA, several *syn* purines have been identified in RNA structures determined by NMR, such as the UUCG tetraloop and the ATP aptamer. The C8 resonance of the *syn* purine is the most down-field of all C8 resonances in the RNA oligonucleotides (29,30).

**Figure 4.** (a) The 10 lowest energy structures are presented with just U11 and C17 superimposed. The r.m.s. deviation from the average for just these 2 nt is  $0.56 \pm 0.18$ . (b) The proposed U-C base pair is shown schematically, demonstrating a base pair with a single hydrogen bond from the cytosine exocyclic amino proton of C17 to the O4 of U11. (c) The hydrogen bonding scheme as determined by Holbrook *et al.* for a U-C base pair in the crystal structure of a 12 nt duplex which includes a central U-C mismatch. A water molecule observed in the crystal structure is shown to bridge between the imino proton of uracil to the N3 of cytosine.

**Table 1.** Volume of intranucleotide NOEs of the H6 proton relative to the H5/H6 cross-peak<sup>a</sup>

Mixing time (ms)	Temperature (°C)	Field strength (MHz)	Residue	Sugar pucker	H1'/H6	H2'/H6	H3'/H6	H5/H6
50	25	500	C15	C2'-endo	0.30	0.18	n.o.	0.62
			C13	C2'-endo	n.o.	0.45	n.o.	0.43
100	25	500	C15	C2'-endo	0.91	0.89	0.54 <sup>b</sup>	0.93
			C13	C2'-endo	0.06	0.88	0.05 <sup>b</sup>	0.68
150	15	500	C15	C2'-endo	1.14	0.86	0.41 <sup>b</sup>	0.88
			C13	C2'-endo	0.04	0.50	0.14 <sup>b</sup>	0.55
150	25	500	C15	C2'-endo	1.34	1.09	0.55 <sup>b</sup>	1.00
			C13	C2'-endo	0.13	1.15	0.05	1.00
150	32	600	C15	C2'-endo	1.05	0.82	0.16 <sup>b</sup>	1.00
			C13	C2'-endo	0.22	1.92 <sup>b</sup>	0.99 <sup>b</sup>	1.00

<sup>a</sup>Volume of cross-peaks in NOESY experiments relative to volume of H5/H6 cross-peak at 150 ms mixing time measured on either a 500 or 600 MHz spectrometer.

<sup>b</sup>NOE cross-peak overlapped with other peaks; volume estimated.  
n.o., not observed.

## DISCUSSION

### A U-C pair

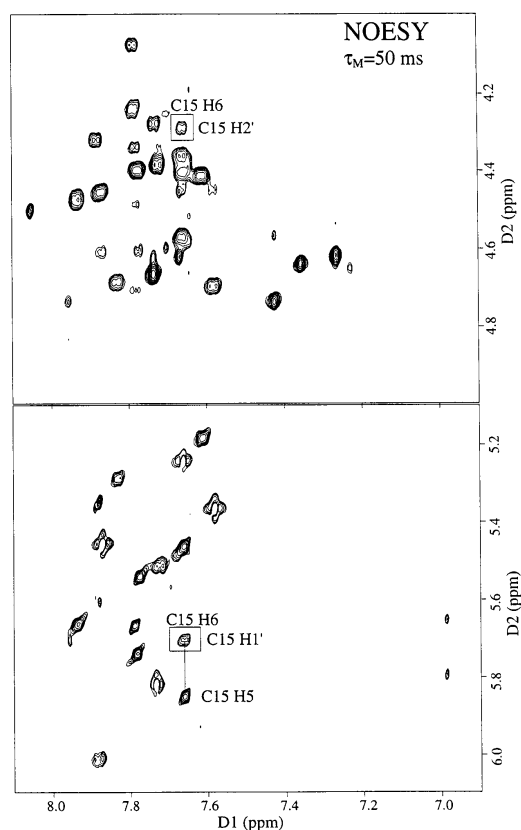
The structure of the central hairpin loop of the ribozyme from the antigenomic sequence of the hepatitis delta virus has been determined by two groups: Kolk *et al.* (28) and this work. A comparison of the two structures shows several distinct differences. First, both structures have a U-C base pair between the first and last nucleotides in the loop that involves the exocyclic amino proton on C17, but the two structures do not agree on the hydrogen bonding. In this study the exocyclic amino proton on C17 is hydrogen bonded to the O4 on U11, as shown in the superimposition of 10 structures in Figure 4a and schematically in Figure 4b; in the Kolk *et al.* study the exocyclic amino proton on C17 is hydrogen bonded to the O2 on U11. The different geometry of this base pair is a result of a different orientation of U11 to the preceding nucleotide, C10. U11 is rotated toward the major groove in the previous study of this hairpin, which causes the O2 to be positioned near the amino proton on C17, while in this study the U11 is stacked normally into the helix.

The proposed U-C pair in our structure is similar to that observed in the crystal structure of the 12 nt dimer with sequence 5'-GGAC(UUCG)GUCC-3', which contains two central mismatches: a G-U mismatch and a U-C mismatch (31). The structure of the base pair observed by Holbrook *et al.* (31) is shown schematically in Figure 4c. In both the Holbrook *et al.* crystal structure and in this study the exocyclic amino proton of cytidine is hydrogen bonded to the O4 of uridine. A second hydrogen bond could be formed between the imino proton of uridine and the N3 of cytidine, but this would introduce radial distortion to the helix. In the Holbrook *et al.* crystal structure bridging water molecules were observed between the N3 atoms of both uridine and cytidine as well as the phosphate backbone and the bases. In NMR bound water molecules are very rarely observed due to exchange, however, a water molecule could fit into the structure stabilizing the base pair formed with a single hydrogen bond between the two bases.

### The loop conformation

The structure of the central hairpin loop of the hepatitis delta virus ribozyme presented here is quite different from that of Kolk *et al.* (28). Although we both agree on a U-C pair between the first and last nucleotides in the loop, the middle 5 nt are in different positions in the two structures. The most likely explanation for the difference in the structures is the difference in solution conditions. The NMR experiments for this structure were done at a pH of 6.2 with 100 mM NaCl, whereas those of Kolk *et al.* were done at a higher pH (6.6) and no additional salt beyond 10 mM sodium phosphate buffer (28). The lack of sodium could cause a major difference in structure. We found that lowering the salt concentration led to multiple species with overlapped imino spectra, so no high resolution structure determination was attempted. While there are no data that illustrate the importance of electrostatic interactions in the structure or function of this loop, without sufficient counter ions electrostatic repulsion between the phosphate groups will dominate. This fact is particularly significant for a hairpin loop where an abrupt turn must be made. It is in this region of the structure that the two structures disagree the most.

A second possibility for the discrepancy in the two structures is loop dynamics. NOEs build up and are detected on a millisecond time scale. Dynamics of the loop occurring faster than that will not be detected by NMR, as NMR will see an average conformation. Without many hydrogen bonds within the loop the loop might be expected to be fairly dynamic, despite the convergence of the family of structures presented here and in the previous work (28). The NMR resonances of residues with fast dynamics would be expected to be sharper and hence intensities of cross-peaks in a DQF-COSY would be more intense. Alternatively, NMR resonances of residues that are dynamic on the NMR time scale of chemical exchange, particularly milliseconds, would be broader than those of other residues and hence they would have weaker cross-peaks in a DQF-COSY. In this study intensities of DQF-COSY cross-peaks were quite similar between all 15 pyrimidines in the 26 nt molecule, indicating that no dynamics are detectable for the loop under the solution conditions used in this study, with the exception of the one purine in the loop, G16,

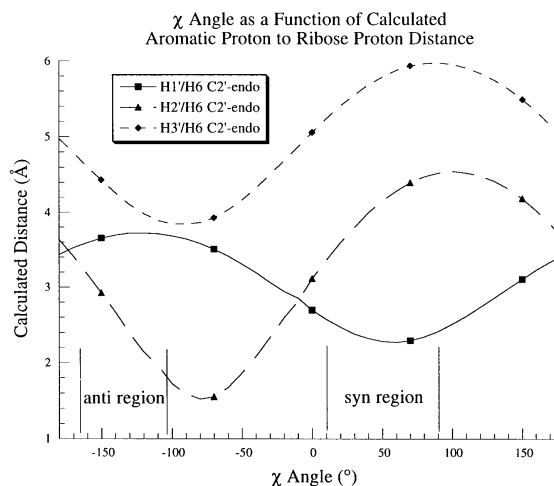


**Figure 5.** A NOESY spectrum acquired at 25 °C with a mixing time of 50 ms on a 2 mM sample of the L3 central hairpin loop with the P2 and P3 stems is shown. Highlighted are the intranucleotide cross-peaks to the H6 aromatic proton of C15 from H1' and H2' on the ribose and H5 on the base. The stronger NOE from the H1' to the H6 than from the H2' to the H6 indicates a *syn* nucleotide. The spectrum was acquired in D<sub>2</sub>O with 2048 complex points, 80 scans and 512  $t_1$  measurements.

as discussed previously. However, in the study of Kolk *et al.* dynamics are detected on the millisecond time scale for C15, the *syn* cytidine, when the spectra were obtained with a 750 MHz spectrometer. With this spectrometer they observed broadening of the NMR resonances for C15 that sharpened with a 400 MHz spectrometer. This is an indication of a dynamics not observed in the study presented here, where all NMR work was done on 500 and 600 MHz spectrometers. The putative dynamics could involve the *syn* glycosidic torsion angle of the cytidine to the *anti* conformation. In both previous studies identifying a *syn* cytidine (discussed below) conformational flexibility involving the glycosidic torsion angle was observed as solution conditions changed (32,33). Thus the low sodium concentration in the Kolk *et al.* study could have caused the flexibility in the loop, particularly around C15, that they observed on a 750 MHz spectrometer; this flexibility could, in turn, lead to distinct differences in the two structures presented for the central hairpin loop of the hepatitis delta virus ribozyme.

### A *syn* cytidine

The conformation of the glycosidic torsion angle of C15 was extensively analyzed by NOESY experiments and carbon and nitrogen correlation experiments to prove that the average conformation of the glycosidic torsion angle of C15 is *syn*. The



**Figure 6.** A plot of the calculated intranucleotide distance from the H6 to the H1' (solid line with squares), H2' (dashed line with triangles) and H3' (dashed line with diamonds) as a function of  $\chi$  is shown. The distances were measured in InsightII (Biosym) for a model cytidine nucleotide with a C2'-*endo* sugar pucker and plotted as a function of  $\chi$ . The  $\chi$  angle is normally *anti* (110–170°) for cytidine, however, the data presented in Table 1 demonstrate that the H2' proton must be further from the H6 proton than the H1' proton, which indicate that the  $\chi$  angle is  $\sim$ 20°, in the range for a *syn* cytidine.

NOE volumes shown in Table 1 confirm that the H1' and the H6 of C15 are closer than the H2' and H6 on the same residue and, through comparison with calculated base–ribose intranucleotide distances shown in Figure 6 as a function of  $\chi$  for cytidine, a glycosidic torsion angle of  $\sim$ 20° is determined. This defines C15 to be a highly unusual *syn* cytidine. In nucleotides, particularly pyrimidines, the *anti* conformation is thermodynamically favored over the *syn* conformation. Steric hindrance from the carbonyl O2 group of a pyrimidine to its H3' as well as O4' makes the *syn* conformation less favored, however, the energetic cost is less than  $\sim$ 1 kcal/mol for ribonucleotides with a C2'-*endo* sugar pucker in calculated conformational energy maps (34). In spite of the small calculated energetic cost for a *syn* conformation, pyrimidines are found to have  $\chi$  angles in the *anti* region.

*Syn* guanosine residues have been observed many times in DNA and RNA, however, *syn* pyrimidines are rare but not unprecedented. A *syn* cytidine was observed by X-ray crystallography of a Z-DNA hexanucleotide duplex with sequences d(m<sup>5</sup>CGCCm<sup>5</sup>CG) and its complement d(m<sup>5</sup>CGGGm<sup>5</sup>CG) (32) and by NMR of a circular decamer with sequence d(CGCTTGCGTT) (33). A *syn* thymidine has been observed by X-ray crystallography of a Z-DNA duplex (35). Pyrimidines are usually *anti* as a result of steric hindrance (36), therefore, higher van der Waals energies should result for a structure with a *syn* pyrimidine. However, calculated van der Waals energies for the Z-DNA hexanucleotide with a *syn* cytidine are actually smaller when compared with a similar hexanucleotide Z-DNA molecule with sequence d(m<sup>5</sup>CGm<sup>5</sup>CGm<sup>5</sup>CG) with all 5-methylcytidines in the *anti* conformation (32). While there is steric hindrance in the crystal structure between the O2 carbonyl oxygen and the hydrogen at the H3' position, the authors suggest that the *syn* cytidine causes a buckling that displaces the base toward the major groove (32). The *syn* nucleotide is both buckled and unstacked in the hexanucleotide structure, which in turn disrupts the nucleotides on either side of the *syn* nucleotide. Thus a larger surface area of the base is exposed to water, which



increases the hydrophobicity of the Z-DNA helix. In the crystal structure this result was observed as a disruption of the spine of water molecules along the minor groove surface of the DNA (32).

In this study of the central hairpin loop from the hepatitis delta virus ribozyme the fifth nucleotide in the loop, C15, is *syn*. The *syn* nucleotide has a C2'-endo sugar pucker as in DNA. In addition, the nearest neighbors do not have the normal RNA C3'-endo sugar pucker; on the 5'-side of C15 U14 has a C2'-endo sugar pucker and on the 3'-side G16 is dynamic and was modeled with a mixture of C3'-endo, C2'-endo and O4'-endo. Similar to the Z-DNA hexanucleotide duplex, there is distortion on either side of the *syn* cytidine, likely resulting from a combination of the *syn* nucleotide and the C2'-endo sugar pucker. The loop has a significant reversal in the direction of the backbone on the 5'-side of the *syn* cytidine, while on the 3'-side G16 is disordered, with the base primarily out in solution. With C15 in the *syn* conformation the hydrophobic surface of the base, C5 and C6, is turned toward the center of the loop, while the amino protons and N3 groups are turned out into solution. A hydrophobic center in the turn of the loop is probably critical for the loop. The *syn* C15 makes the hydrogen bond donors and acceptors of the middle three nucleotides, C13, U14 and C15, easily accessible to the solvent. Furthermore, the disruption on the 3'-side of C15 forces G16 out in solution and prevents formation of a Watson-Crick base pair with C12. As discussed earlier, G16 is critical for cleavage, probably involving a tertiary interaction with a nucleotide near the cleavage site in the whole ribozyme.

### Structural implications for the ribozyme

The structure of the central hairpin loop of the hepatitis delta virus ribozyme may not be the same as in the active ribozyme, since tertiary interactions with residues of the cleavage site are absent. However, some specific findings in this structure have implications for the whole ribozyme. The first nucleotide, U11, is critical for cleavage in the ribozyme (8) and may be part of the catalytic site. However, formation of the loop U-C base pair as seen in this structure could be critical. This base, along with the *syn* cytidine, prevents C12 and G16 from forming a Watson-Crick base pair. The purine in the loop, G16, is required for efficient cleavage in the ribozyme (8). In this structure it is not well ordered; it is out of the loop and available for interactions with another part of the ribozyme. While C12 does not appear to be involved in any specific contacts in the loop structure, it is also critical for cleavage (8). One possibility is that it is involved in a triple base pair in the loop along with U11 and C17 that is suggested in the loop structure, but not proven. Alternatively, C12 is not involved in any contacts within the loop and it is able to form tertiary interactions in the ribozyme. The middle three loop nucleotides seem to be less important in the ribozyme, but still affect cleavage rates (8). Most likely they only function to reverse the direction of the loop, in which case pyrimidines are preferable to purines due to steric hindrance. Assigning the NMR spectra and determining the structure of the L3 loop of the hepatitis delta virus ribozyme will help in structure determination of the whole ribozyme and will allow us to see the effects of tertiary interactions on secondary structure.

### ACKNOWLEDGEMENTS

We thank Ms Barbara Dengler for general assistance, Mr David Koh for synthesizing DNA templates and Dr Jeffrey Pelton for assistance with NMR experiments. Dr Stephen Lynch was

supported by a post-doctoral fellowship from the National Institutes of Health (1 F32 GM16949-01). This research was supported in part by National Institutes of Health grant GM 10840, by Department of Energy grant DE-FG03-86ER60406 and through instrumentation grants from the Department of Energy (DE-FG05-86ER75281) and from the National Science Foundation (DMB 86-09305).

See supplementary material available in NAR Online.

### REFERENCES

- Sharmeen,L., Kuo,M.y.-P., Dinter-Gottlieb,G. and Taylor,J. (1988) *J. Virol.*, **62**, 4439–4444.
- Wu,H.N., Lin,Y.J., Lin,F.P., Makino,S., Chang,M.F. and Lai,M.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 1831–1835.
- Perrotta,A.T. and Been,M.D. (1991) *Nature*, **350**, 434–436.
- Rosenstein,S.P. and Been,M.D. (1991) *Nucleic Acids Res.*, **19**, 5409–5416.
- Been,M.D., Perrotta,A.T. and Rosenstein,S.P. (1992) *Biochemistry*, **31**, 11843–11852.
- Perrotta,A.T. and Been,M.D. (1993) *Nucleic Acids Res.*, **21**, 3959–3965.
- Thill,G., Vasseur,M. and Tanner,N.K. (1993) *Biochemistry*, **32**, 4254–4262.
- Perrotta,A.T. and Been,M.D. (1996) *Nucleic Acids Res.*, **24**, 1314–1321.
- Puttaraju,M., Perrotta,A.T. and Been,M.D. (1993) *Nucleic Acids Res.*, **21**, 4253–4258.
- Cai,Z. and Tinoco,I.,Jr (1996) *Biochemistry*, **35**, 6026–6036.
- Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.
- Batey,R.T., Battiste,J.L. and Williamson,J.R. (1995) *Methods Enzymol.*, **261**, 300–322.
- Nikonowicz,E.P., Sirt,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) *Nucleic Acids Res.*, **20**, 4507–4513.
- Haynie,S.L. and Whitesides,G.M. (1990) *Appl. Biochem. Biotechnol.*, **23**, 205–220.
- Macura,S., Wüthrich,K. and Ernst,R.R. (1982) *J. Magn. Resonance*, **46**, 269–282.
- Sklenar,V., Tschudin,R. and Bax,A. (1987) *J. Magn. Resonance*, **75**, 352–357.
- Muller,N., Ernst,R.R. and Wüthrich,K. (1986) *J. Am. Chem. Soc.*, **108**, 6482–6492.
- Shaka,A.J., Barker,P.B. and Freeman,R. (1985) *J. Magn. Resonance*, **64**, 547–552.
- Bax,A., Davis,D.G., Gregory,R.J., Cahill,P.B., Thurlow,D.L. and Zimmermann,R.A. (1985) *J. Magn. Resonance*, **65**, 295–307.
- Wijmenga,S.S., Heus,H.A., Werten,B., van der Marel,G.A., van Boom,J.H. and Hilbers,C.W. (1994) *J. Magn. Resonance*, **103B**, 134–141.
- Bax,A., Griffey,R.H. and Hawkins,B.L. (1983) *J. Magn. Resonance*, **55**, 301–315.
- Sklenar,V., Miyashiro,H., Zon,G., Miles,H.T. and Bax,A. (1986) *FEBS Lett.*, **208**, 94–98.
- Bax,A., Ikura,M., Kay,L.E., Torchia,D.A. and Tschudin,R. (1990) *J. Magn. Resonance*, **86**, 304–318.
- Mueller,L., Legault,P. and Pardi,A. (1995) *J. Am. Chem. Soc.*, **117**, 11043–11048.
- Sklenar,V., Peterson,R.D., Rejante,M.R. and Feigon,J. (1994) *J. Biomol. NMR*, **4**, 117–122.
- Saenger,W. (1984) In Cantor, C.R. (ed.) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, NY, pp. 230–232.
- Brünger,A.T. (1992) *X-PLOR Version 3.0*. Yale University Press, New Haven, CT.
- Kolk,M.H., Heus,H.A. and Hilbers,C.W. (1997) *EMBO J.*, **16**, 3685–3692.
- Varani,G. and Tinoco,I.,Jr (1991) *J. Am. Chem. Soc.*, **113**, 9349–9354.
- Dieckmann,T. and Feigon,J. (1997) *J. Biol. NMR*, **9**, 259–272.
- Holbrook,S.R., Cheong,C., Tinoco,I.,Jr and Kim,S.H. (1991) *Nature*, **353**, 579–581.
- Schroth,G.P., Kagawa,T.F. and Ho,P.S. (1993) *Biochemistry*, **32**, 13381–13392.
- Ippel,J.H., Lanzotti,V., Galeone,A., Mayol,L., van den Boogaart,J.E., Pikkemaat,J.A. and Altona,C. (1995) *J. Biomol. NMR*, **6**, 403–422.
- Saran,A., Perahia,D. and Pullman,B. (1973) *Theor. Chim. Acta (Berl.)*, **30**, 31–44.
- Wang,A.H.-J., Gessner,R.V., van der Marel,G.A., van Boom,J.H. and Rich,A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3611–3615.
- Haschemeyer,A.E.V. and Rich,A. (1967) *J. Mol. Biol.*, **27**, 369–384.