Solution Structure of the CUUG Hairpin Loop: A Novel RNA Tetraloop Motif[†,‡]

Fiona M. Jucker and Arthur Pardi*

Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, Colorado 80309-0215

Received July 5, 1995; Revised Manuscript Received September 5, 1995[⊗]

ABSTRACT: The solution structure of a uniformly $^{13}C/^{15}N$-labeled CUUG RNA hairpin loop has been determined by multidimensional heteronuclear magnetic resonance spectroscopy in combination with distance geometry and restrained molecular dynamics calculations. The structure of this CUUG tetraloop represents a novel RNA loop motif where the first and last loop nucleotides form a standard Watson−Crick C-G base pair and the second loop nucleotide interacts directly with the closing base pair of the stem by folding into the minor groove. This structure helps explain why the closing base pair is phylogenetically conserved and indicates a six-nucleotide G(CUNG)C motif for the CUUG RNA tetraloop. Implications for the function of this CUUG tetraloop in ribosomal RNA and in RNA tertiary interactions are discussed.

RNA hairpins made up of helix−loop structures are among the most abundant structural elements in RNA. RNA tetraloops are a special class of hairpins that were originally identified because this loop size was found much more frequently in prokaryotic ribosomal RNA than other loop sizes (Uhlenbeck, 1990; Woese et al., 1990). Three families of tetraloops have been identified in rRNA,[1] the UUCG, GNRA, and CUUG hairpins (Woese et al., 1990), and phylogenetic and mutational analysis has shown that one tetraloop sequence frequently can be replaced by another tetraloop sequence (Woese et al., 1990; Selinger et al., 1993b). These frequently occurring tetraloops have a higher thermodynamic stability than other four-nucleotide loops with the same stem (Uhlenbeck, 1990; Antao et al., 1991; Antao & Tinoco, 1992). These tetraloops are thought to carry out a variety of functions including (i) stabilizing an adjacent stem by taking advantage of their high thermodynamic stability (Tuerk et al., 1988; Selinger et al., 1993b), (ii) providing recognition elements for tertiary interactions with both RNA and proteins (Michel & Westhof, 1990; Zwieb, 1991, 1992; Selinger et al., 1993a; Jaeger et al., 1994; Murphy & Cech, 1994), (iii) regulating the activity of a biological system by shifting the equilibrium between alternate structures (Nierhaus et al., 1992; Wool et al., 1992; Glueck et al., 1994), and (iv) providing an initiation point for the folding of complex RNAs (Uhlenbeck, 1990; Woese et al., 1990). Three-dimensional structures have been determined for the UUCG (Cheong et al., 1990; Varani et al., 1991) and the GNRA tetraloops (Heus & Pardi, 1991b; Orita et al., 1993; Szewczak et al., 1993; Pley et al., 1994) and revealed interesting features that help explain the unusual stability of these tetraloops. Here we used multidimensional heteronuclear NMR techniques and distance geometry and molecular dynamics calculations to determine the solution structure of the uniformly 99% $^{13}C/^{15}N$-labeled oligomer 5′-r(GGCGCUUGCGUC)-3′ (see Figure 1).

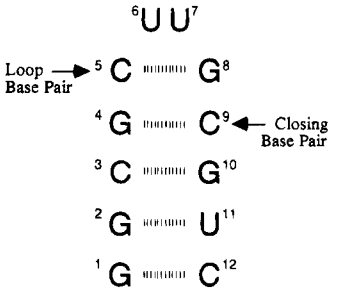

FIGURE 1: Sequence of the CUUG RNA hairpin shown in the secondary structure representation. The loop base pair and the closing base pair are indicated by arrows.

The CUUG RNA tetraloop is the preferred sequence for loop 83 in bacterial 16S rRNA (positions 83 to 86 in *Escherichia coli* 16S rRNA) (Woese et al., 1990). This particular hairpin in 16S rRNA shows a strong preference for tetraloops where only three tetraloop sequences occur frequently with the CUUG sequence found in 45% of all cases, followed by UUCG (36%) and GCAA (13%) (Woese et al., 1990). The closing base pair is also conserved in loop 83 with the CUUG loop closed by a G-C base pair, the UUCG loop closed by a C-G base pair, and the GCAA closed by an A-U base pair. Phylogenetic analysis thus strongly suggests that each tetraloop motif includes its closing base pair (Woese et al., 1990). The structure of the CUUG RNA tetraloop determined here reveals a new type of tetraloop motif and offers an explanation for the preference of a closing G-C base pair.

[†] This work was supported by NIH Grant AI33098 and a NIH Research Career Development Award (AI01051) to A.P. The VXR-500 NMR spectrometer was purchased with partial support from NIH Grant RR03283. We also thank the Colorado RNA Center and the W. M. Keck Foundation for their generous support of RNA science on the Boulder campus.

[‡] Coordinates of the final structures have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY 11973, file name 1RNG.

* To whom correspondence should be addressed.

[⊗] Abstract published in *Advance ACS Abstracts,* October 15, 1995.

[1] Abbreviations: NMP, nucleoside 5′-monophosphate; NTP, nucleoside 5′-triphosphate; NOE, nuclear Overhauser effect; RMSD, root mean square deviation; NMR, nuclear magnetic resonance spectroscopy; DQF-COSY, double quantum filtered correlation spectroscopy; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser effect spectroscopy; HETCOR, heteronuclear correlation spectroscopy; HMQC, heteronuclear multiple quantum spectroscopy; rRNA, ribosomal RNA.

0006-2960/95/0434-14416$09.00/0

MATERIALS AND METHODS

Sample Preparation. RNA oligonucleotides were prepared from DNA oligonucleotide templates by *in vitro* transcription with T7 RNA polymerase (Milligan, 1987). DNA oligonucleotides were obtained from DNA Express (Macromolecular Resources), ($5'$TAATACGACTCACTATAG$3'$) for the T7 promoter strand and ($5'$GACGCAAGCGCCTATAGTGAGTCGTATTA$3'$) for the template strand. T7 RNA polymerase was purified from *E. coli* strain BL21 containing the plasmid pAR1219 (Davanloo et al., 1984). ^{13}C- and ^{15}N-labeled nucleotides were prepared as previously described (Nikonowicz et al., 1992) by growing *E. coli* MY 285 on 99% [^{13}C]glucose (Isotec) and 99% [^{15}N]ammonium sulfate (Isotec), followed by enzymatic degradation of the ribosomal RNA to NMPs. After separation over a AG-1-X2 column (Biorad), the NMPs were enzymatically converted to NTPs and desalted over a Biorad Affigel 601 column (Nikonowicz et al., 1992). Transcription reaction mixtures (10−14 mL) were incubated at 37 °C for 4−8 h [40 mM Tris-HCl (pH 7.6), 35 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine, 0.1% Triton X-100 (v/v), 200 nM DNA template, and 0.1 mg/mL T7 RNA polymerase] with 4 mM of each NTP for the unlabeled RNA and 5 mM [$^{13}C/^{15}N$]GTP, 3.4 mM [$^{13}C/^{15}N$]-CTP, and 1.8 mM [$^{13}C/^{15}N$]UTP for the labeled RNA. The final product was purified by denaturing 20% polyacrylamide gel electrophoresis and electroeluted (Schleicher & Schuell) and dialyzed with a Centricon 3 microconcentrator (Amicon) into 2.5 mM sodium phosphate (pH 6.8) and 0.1 mM EDTA. The sequence of the final RNA product was confirmed by RNA sequencing (Donis-Keller et al., 1977; Donis-Keller, 1980). The lyophilized RNA was dissolved in 600 μL of 99.96% D_2O or 600 μL of 90% H_2O/10% D_2O. The final concentration for all NMR samples was 0.8 mM RNA.

NMR Spectroscopy. NMR experiments were performed on a 500 MHz Varian VXR-S spectrometer. For samples in 90% H_2O/10% D_2O, spectra were collected at 5 °C with a binomial 1331 sequence for solvent suppression (Hore, 1983). Unless noted, all spectra in D_2O were collected at 20 °C with presaturation of the residual HDO peak. Quadrature detection during the evolution periods in the multidimensional NMR experiments was achieved with the hypercomplex method (States, 1982) or the time proportional phase incrementation−States method (Marion, 1989a). Broadband ^{13}C or ^{15}N WALTZ decoupling was applied during the acquisition period for spectra collected on $^{13}C/^{15}N$-labeled RNA. The NMR data were processed on a Sun 4 or a Silicon Graphics computer with the program FELIX (Biosym).

Two-dimensional (2D) homonuclear ($^1H-^1H$) and heteronuclear ($^{31}P-^1H$) NMR experiments were collected on an unlabeled RNA sample. A 250 ms 2D NOESY experiment (Macura & Ernst, 1979) in D_2O was performed at 5 °C to compare the spectra to those collected at this temperature in H_2O. The 80 and 400 ms NOESY experiments used for obtaining resonance assignments and structural constraints were acquired at 20 °C. DQF-COSY experiments (Rance et al., 1983) were performed at 5 and 20 °C. In order to obtain *J* coupling constants, DQF-COSY spectra were collected for the sugar proton region with and without ^{31}P decoupling in t_1 and t_2. A TOCSY experiment with a 70 ms spin-lock period and a $^{31}P-^1H$ HETCOR experiment were acquired by standard procedures (Braunschweiler & Ernst, 1983; Sklenar et al., 1986). The detailed acquisition and processing parameters for the homonuclear 2D NMR experiments are given in Table 1 in the supporting information.

Two-dimensional ($^{13}C,^1H$) and ($^{15}N,^1H$) HMQC experiments were acquired in D_2O and H_2O, respectively (Mueller, 1979; Bax et al., 1983; Redfield, 1987). Both 2D ($^1H-^1H$) and 2D ($^1H-^{13}C$) and HCCH-COSY, HCCH-RELAY, and HCCH-TOCSY experiments were acquired as previously described (Pardi & Nikonowicz, 1992). A DIPSI-2 mixing scheme was applied during the 25 ms ^{13}C spin−lock period in the HCCH-TOCSY experiments, and the delays were optimized to obtain ω_2 methylene-filtered and ω_2 methylene-selected experiments (Pardi & Nikonowicz, 1992). The detailed acquisition and processing parameters for the heteronuclear 2D NMR experiments are given in Table 2 in the supporting information.

A 3D ($^1H,^{13}C,^1H$) NOESY-HMQC experiment optimized for the aromatic resonances in t_1 and sugar resonances in t_2 and t_3 was acquired with a 150 ms mixing time (Marion et al., 1989b; Nikonowicz & Pardi, 1993). A 3D ($^{13}C,^1H,^1H$) HMQC-NOESY experiment optimized for the sugar resonances in t_1 and t_2 and for the aromatic resonances in t_3 was acquired with a 400 ms mixing time. A 3D ($^1H,^{13}C,^1H$) NOESY-HMQC experiment optimized for the sugar resonances in all three dimensions was collected with a 400 ms mixing time. The detailed acquisition and processing parameters for the heteronuclear 3D NMR experiments are given in Table 3 in the supporting information.

Structure Calculations and Analysis. Distance geometry and simulated annealing were performed on a Silicon Graphics Indigo workstation using the program X-PLOR 3.1 (Brünger, 1992). The *in vacuo* molecular dynamics/simulated annealing calculations were performed with only slight modifications in the protocol provided with the program. The molecular mechanics potentials used were those of Nilsson and Karplus (1986), except that the default sugar bond angles were modified to match those of Saenger (1984). Quadratic harmonic potential functions were used for the bond length and angle energy terms, and the XPLOR "square well" potential function was used for both the NOE restraints and the torsion angle restraints with force constants of 50 kcal/(mol Å^2) and 10 kcal/(mol rad^2), respectively. A $1/R$ distance dependent dielectric was employed for the electrostatic function (Brünger, 1992).

NMR-derived distance and torsion angle constraints and nonexperimental A-form constraints were applied during the calculations as described below. To analyze the final structures, the program NEWHEL93 (obtained from R. E. Dickerson) was used in addition to the program X-PLOR (Brünger, 1992).

RESULTS

Sample Preparation. The sequence for the CUUG hairpin in Figure 1 can form a unimolecular hairpin or a bimolecular duplex, depending upon the temperature and the salt and oligomer concentration (Jucker Franzusoff, 1995). At the sample concentration of 0.8 mM RNA in the salt and buffer conditions employed here, this molecule formed only the hairpin conformation and no duplex could be detected in the NMR spectra.

Resonance Assignments. Resonance assignment of the exchangeable protons was made from 2D and 3D NOESY

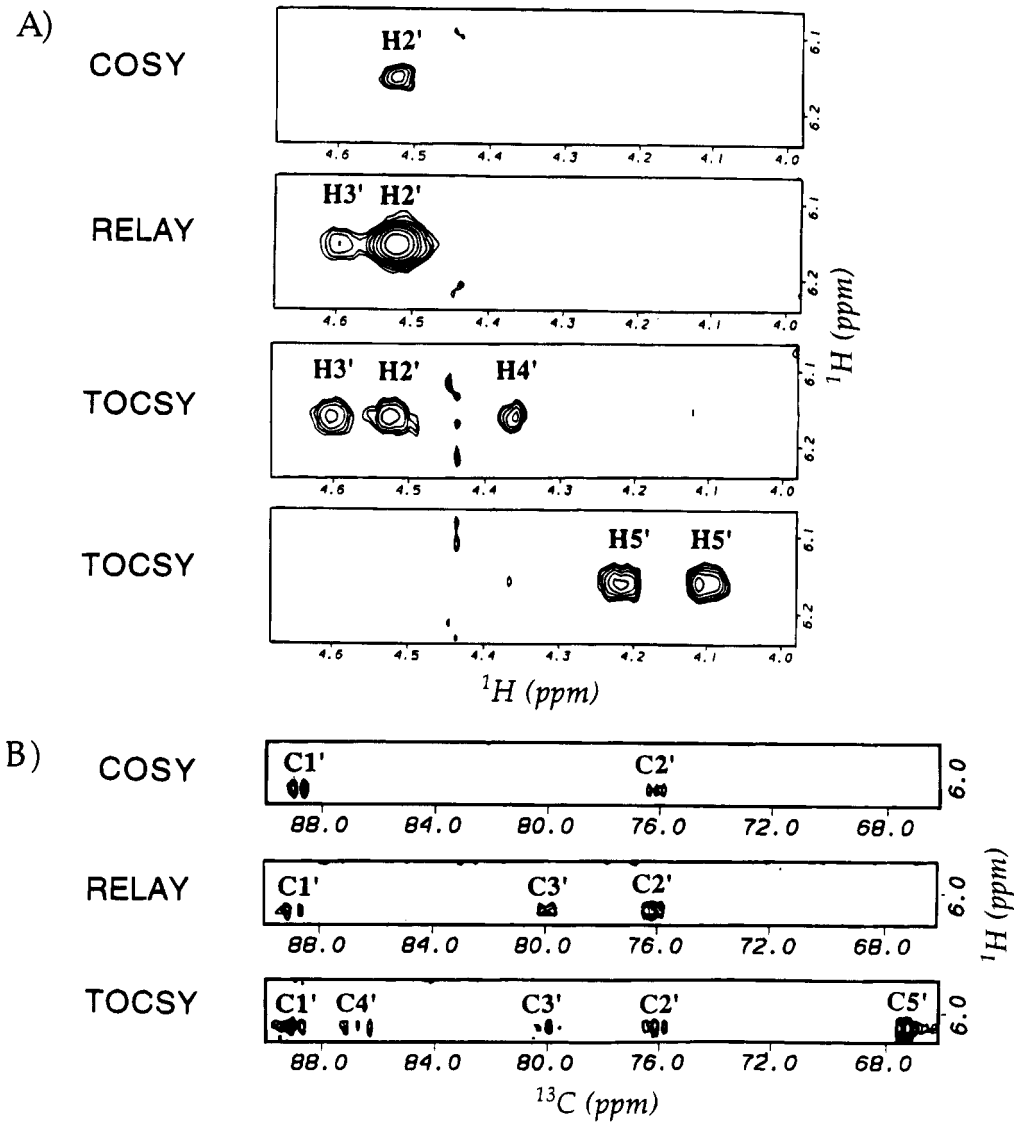

FIGURE 2: Part of the sugar region of the 2D HCCH COSY, 2D HCCH RELAY, and 2D HCCH TOCSY spectra used for assignment of the sugar proton and carbon resonances. (A) The H2′-H5′ to H1′ regions of the 2D (^{1}H,^{1}H) COSY, RELAY, ω_2 methylene-filtered TOCSY, and ω_2 methylene-selected TOCSY spectra are shown for nucleotide U6. The methylene-filtered TOCSY spectrum is plotted above the methylene-selected TOCSY spectrum below. (B) The C2′-C5′ to H1′ regions of the 2D (^{13}C,^{1}H) COSY, RELAY, and TOCSY spectra are shown for nucleotide U6.

experiments collected in H_2O by standard procedures (Wüthrich, 1986; Varani & Tinoco, 1991b; Nikonowicz & Pardi, 1993; Wijmenga et al., 1993). Imino protons in G-C base pairs were assigned on the basis of their connectivities to C H5-H6 cross-peaks. The imino protons in the G-U base pair were assigned on the basis of the strong G imino to U imino cross-peak, and the identity of each base was readily determined from the (^{1}H,^{15}N) HMQC spectrum collected in H_2O (data not shown) because the G and U imino nitrogens resonate at very different chemical shifts (Nikonowicz & Pardi, 1993). The two-loop U imino protons were not specifically assigned because no NOE cross-peaks were observed for those imino protons due to exchange with solvent H_2O; therefore, no structural information could be obtained for these protons.

Identification of the sugar spin system is an important step toward complete resonance assignment of RNAs (Patel et al., 1987; van de Ven & Hilbers, 1988; Varani & Tinoco, 1991b; Nikonowicz & Pardi, 1993; Wijmenga et al., 1993). Assignment of the sugar resonances was made from 2D HCCH experiments in a manner similar to that previously described (Nikonowicz & Pardi, 1993). These HCCH experiments allow unambiguous sugar resonance assignments because magnetization is transferred through the large one-bond coupling constants. Sugar proton resonance assignments were made from 2D (^{1}H,^{1}H) HCCH-COSY, 2D (^{1}H,^{1}H) HCCH-RELAY, and 2D (^{1}H,^{1}H) HCCH-TOCSY experiments by examination of the cross-peaks to each H1′ proton as illustrated in Figure 2A. Sugar carbon resonance assignments were made from 2D (^{13}C,^{1}H) HCCH-COSY, 2D (^{13}C,^{1}H) HCCH-RELAY, and 2D (^{13}C,^{1}H) HCCH-TOCSY experiments by examination of the cross-peaks to each H1′ as illustrated in Figure 2B.

The nonexchangeable base protons and their directly attached carbons were identified in HMQC experiments optimized for the aromatic resonances (Nikonowicz & Pardi, 1993). The pyrimidine H5 and H6 protons were identified in homonuclear 2D DQF-COSY experiments. The sequential resonance assignments were made from 3D (^{13}C,^{1}H,^{1}H) HMQC-NOESY and 3D (^{1}H,^{13}C,^{1}H) NOESY-HMQC ex-

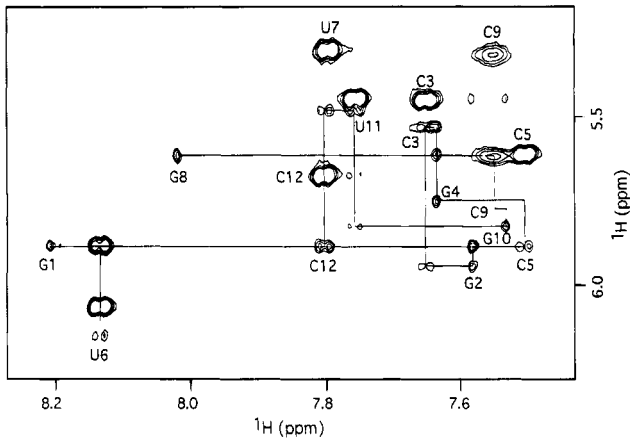

FIGURE 3: Sequential walk in the H1′ to aromatic region is displayed in the 400 ms 2D NOESY spectrum in 90% H_2O. The solid lines connect H1′ to H8/H6 NOE cross-peaks observed in both the 2D and the 3D NOESY experiments. The dashed lines indicate the sequential walk that was determined from the 3D NOESY experiments, where the H2′ and H3′ resonances were used to establish the sequential walk. The sequential walk between U7 and G8 is not illustrated in the figure because this H1′ to aromatic cross-peak falls outside the region displayed here.

periments by extending standard 2D homonuclear assignment methods (Wüthrich, 1986; Varani & Tinoco, 1991b; Wijmenga et al., 1993) into three dimensions as previously described (Nikonowicz & Pardi, 1993). The sequential assignments were confirmed in the homonuclear 2D NOESY experiment shown in Figure 3. The sequential NOEs are summarized in Figure 4, which shows that sequential aromatic proton to sugar proton NOEs are observed for most stem nucleotides, as expected for a helical conformation. For nucleotides G8 and C9, few sequential NOEs are observed due to the increased line width of the involved protons and the sequential NOE pattern is interrupted in the loop between U6 and U7 due to the unusual conformation for this part of the loop.

Phosphorus resonance assignments were made from the (^{31}P,^{1}H) HETCOR experiment (Sklenar et al., 1986) shown in Figure 5. The relatively good chemical shift dispersion and the small size of the RNA allowed for unambiguous cross-peak assignments. A summary of the ^{1}H, ^{13}C, ^{15}N, and ^{31}P chemical shifts is given in Table 1.

Obtaining Constraints for Structure Calculations. Distance and torsion angle constraints were used to define the structure of the CUUG RNA hairpin. Distance constraints were extracted from 80, 250, and 400 ms mixing time 2D (^{1}H,^{1}H) NOESY spectra, 150 and 400 ms mixing time 3D (^{1}H,^{13}C,^{1}H) NOESY-HMQC spectra, and 400 ms mixing time 3D (^{13}C,^{1}H,^{1}H) HMQC-NOESY spectra (Kumar et al., 1980; Fesik & Zuiderweg, 1988, 1990; Marion et al., 1989b; Zuiderweg et al., 1990). The H5 to H6 cross-peak of pyrimidines and the H5′ to H5″ cross-peaks of the sugars were used to calibrate the cross-peak size. A qualitative interpretation classified the cross-peak intensities into three categories, weak, medium, and strong, which corresponded to distance bounds of 3.0−5.0, 2.0−4.5, and 1.8−3.5 Å, respectively. The contribution of spin diffusion was evaluated from the 80 ms 2D (^{1}H,^{1}H) NOESY spectrum, with special care taken for cross-peaks involving the 2′ and 3′ sugar protons and the pyrimidine H6 and H5. Exchangeable protons had distance constraints of 1.8−4.5 Å. A list of distance constraints is given in Table 4 in the supporting information. The 14-base pair hydrogen bonds were defined as 14 distance constraints between the hydrogen and the hydrogen bond acceptor with distance bounds of 1.4−2.0 Å. Three distance constraints were used for each stem, and loop C-G base pair and two distance constraints were used for the G-U base pair. In addition, 23 negative distance constraints, derived from absent cross-peaks, were included for the loop nucleotides. These negative constraints were introduced after initial structure calculations resulted in structures that predicted large NOE cross-peaks involving the loop nucleotides. Negative constraints were only used if each of the involved protons showed an observable NOE to another proton.

To help restrain the torsion angles, J coupling constants were determined from analysis of the ^{31}P-coupled and -decoupled DQF-COSY (Wüthrich, 1986) using standard procedures (Altona & Sundaralingam, 1972; Altona, 1982; Varani & Tinoco, 1991b; Wijmenga et al., 1993). In addition, the multiplet structure of the cross-peaks in the (^{31}P,^{1}H) HETCOR spectrum (Sklenar et al., 1986) in Figure 5 was used to estimate the ^{1}H−^{31}P J coupling constants that were analyzed using Karplus equations (Haasnoot et al., 1980, 1981). The torsion angle β was restrained to 100−260° for nucleotides G2, C3, G4, and U11. The torsion angle γ was restrained to 35−65° for G2, C3, U11, and C12. The torsion angle ϵ was restricted to 165−245° for nucleotides G2, C3, G10, and U11. The torsion angle δ was used to define the sugar pucker and was derived from the H1′ to H2′ coupling constant (Saenger, 1984; Wüthrich, 1986; Varani & Tinoco, 1991b; Wijmenga et al., 1993). The sugar pucker was restricted to 3′ endo (δ restrained to 78−98°) for stem nucleotides G1−G4 and G10−C12 and to 2′ endo for the loop nucleotides C5−G8 (δ restrained to 110−190°). A number of carbon and proton resonances for nucleotides in the loop showed significant line broadening, indicating that these resonances were in intermediate exchange on the chemical shift time scale (Jardetzky & Roberts, 1981). The J coupling constants derived from the (^{31}P,^{1}H) HETCOR and the DQF-COSY spectra were generally not consistent with a single conformation for the loop nucleotides. Therefore, due to the dynamic nature of the loop, backbone torsion angle constraints derived from J coupling constants were only applied to the stem nucleotides. In addition, no backbone torsion angle restraints were applied to nucleotide C9 due to the increased line width observed for many C9 resonances. The only torsion angle constraints used in the four CUUG loop nucleotides were for the δ torsion angle which restricted the sugar pucker to 2′ endo.

In the initial distance geometry calculations, nonexperimental A-form constraints were applied to the stem nucleotides to help restrain the β, γ, δ, and ϵ torsion angles and the endocyclic sugar torsion angles for nucleotides G1−G4 and G10−C12. No nonexperimental A-form constraints were included for the nucleotides in the loop. These nonexperimental constraints were only used to generate starting structures and were not used in the subsequent molecular dynamics calculations, as described below.

Structure Determination. Distance geometry, molecular dynamics, and energy minimization were used to determine the loop structure of the CUUG RNA hairpin. The structure calculations were divided into two parts. In the first part, starting structures were built by distance geometry embedding followed by 100 steps of energy minimization, 6 ps of

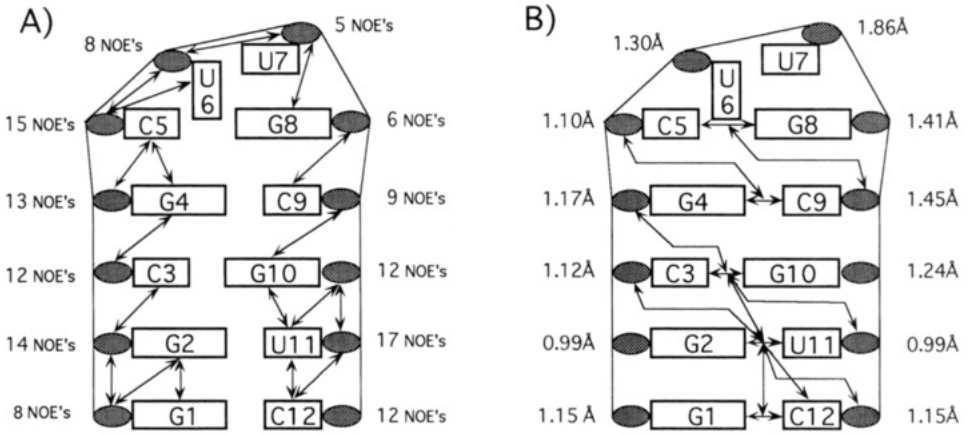

FIGURE 4: Sequential NOE constraints used to calculate the structure of the CUUG RNA hairpin indicated on the secondary structure representation. (A) NOE distance constraints involving nonexchangeable protons. The number of inter-nucleotide NOE constraints are shown for each nucleotide. (B) NOE distance constraints involving exchangeable protons. The RMSD is given for each nucleotide.

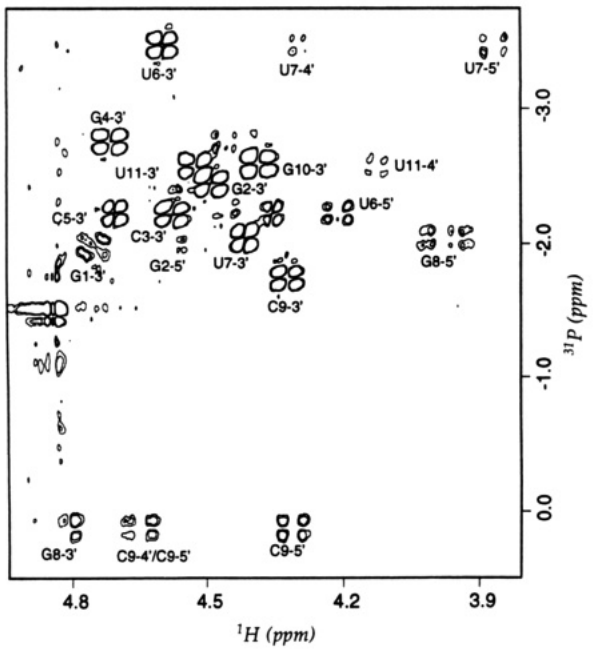

FIGURE 5: Contour plot of the 2D (^{31}P,^{1}H) HETCOR spectrum with assignments of individual cross-peaks.

restrained molecular dynamics at 2000 K, and then 3 ps of restrained molecular dynamics where the temperature was decreased from 2000 to 100 K. The X-PLOR repel function was used for these calculations, and the weighting of various energy terms was varied, similar to the protocol supplied with the X-PLOR program (Brünger, 1992). Nonexperimental A-form constraints for the stem, as well as all the NMR-derived constraints, were included in building starting structures. These nonexperimental A-form torsion angle constraints for the stem nucleotides help generate starting structures that had reasonable helical geometries for the stem nucleotides. Even though A-form backbone torsion angle constraints for the stem were used for calculating these starting structures, the stem still showed substantial structural variation. Although it is impossible to prove that this protocol sampled all conformational space consistent with our experimental data, we believe that this procedure represented a computationally efficient method for generating starting structures that sample conformation space consistent with the experimental NMR constraints. Approximately 50% of the structures created by this protocol had low energies and low NOE violations and thus could be used in the subsequent refinement procedure.

For the second part of the structure calculations, all nonexperimental A-form constraints were omitted for refinement of the starting structures that involved high-temperature and room-temperature molecular dynamics followed by energy minimization as outlined in Figure 6. The refinement protocol is a higher temperature version of the standard "gentle refine" protocol provided by X-PLOR (Brünger, 1992). Over 100 starting structures were refined, and all exhibited the same loop fold. The 13 low-energy structures were then used for the structural analysis. These structures fulfilled all NOE and dihedral angle constraints with no NOE violations >0.15 Å and no dihedral angle violations >1°. This level of violations is somewhat lower than normal and simply reflects the fact that conservative distance and torsion angle constraints were used in the calculations due to the observed dynamics in the molecule.

To analyze these hairpin structures, the program NEWHEL93 (R. E. Dickerson) was used in addition to X-PLOR. An average structure was generated by X-PLOR, followed by 500 steps of restrained energy minimization. RMSDs were calculated by the pairwise superimposition of nucleotides G1−U6 and G8−C12 onto the average structure, thus excluding the poorly defined nucleotide U7. The 13 final structures showed an RMSD to the average structure of 1.3 Å. This RMSD is in the range of what can be expected from NOE data that are interpreted with loose distance bounds (Metzler et al., 1990), especially given that many distance and torsion angle constraints for the loop nucleotides could not be defined due to the dynamics observed in the loop. A lower RMSD could be achieved by employing tighter distance constraints or assuming a single conformation, but given the dynamic behavior of the system under study, this could lead to inaccurate structures. In Figure 4, the number of NOEs and the RMSD are given for each nucleotide. Distance violations were calculated for all structures, and all possible inter-proton distances were analyzed for the average structure to ensure that no close proton−proton distances occurred that are inconsistent with the NMR data. The torsion angles of the 13 final structures were analyzed, and even though no torsion angle constraints were included for the loop nucleotides, all the final structures had both stem and loop torsion angles that were consistent with the experimental NMR spectra.

Table 1: Chemical Shifts of the CUUG RNA Hairpin[a]

residue		imino H1/H3	imino N1/N3	amino 2/4	aromatic 8/6	aromatic 5	sugar 1′	sugar 2′	sugar 3′	sugar 4′	sugar 5′/5″	^{31}P
G1	^{1}H	12.98			8.21		5.89	4.92	4.72	4.60	4.46/4.31	
	^{15}N/^{13}C		146.33		140.6		92.22	75.45	74.82	83.72	67.50	
G2	^{1}H	11.67		6.34	7.58		5.95	4.75	4.48	4.57	4.51/4.27	−2.11
	^{15}N/^{13}C		143.25		138.6		93.34	75.45	73.23	82.46	66.39	
C3	^{1}H			8.37/6.94	7.64	5.43	5.54	4.27	4.58	4.45	4.55/4.12	−1.58
	^{15}N/^{13}C				141.0	97.92	94.00	75.45	71.87	82.06	64.46	
G4	^{1}H	12.68		5.68	7.64		5.76	4.20	4.72	4.37	4.58/4.11	−2.34
	^{15}N/^{13}C		144.69		137.5		93.08	75.45	72.27	82.38	64.32	
C5	^{1}H			8.61/7.00	7.50	5.61	5.89	4.22	4.72	4.48	4.47/4.02	−2.89
	^{15}N/^{13}C				143.1	99.44	89.03	75.45	81.02	85.56	67.50	
U6	^{1}H	10.91/11.68[b]			8.14	6.06	6.16	4.52	4.60	4.37	4.22/4.10	−2.34
	^{15}N/^{13}C		156.18/157.84[b]		145.70	105.53	88.97	76.65	80.39	87.15	68.14	
U7	^{1}H	10.91/11.68[b]			7.80	5.32	4.92	3.96	4.43	4.31	4.12/3.88	−3.60
	^{15}N/^{13}C		156.18/157.84[b]		145.8	104.15	89.30	77.84	81.02	85.40	68.45	
G8	^{1}H	13.34			8.03		5.62	4.99	4.83	4.45	4.01/3.96	−2.16
	^{15}N/^{13}C		146.05		138.9		86.38	74.26	77.64	85.56	69.89	
C9	^{1}H			8.35/6.52	7.54	5.32	5.79	4.85	4.33	4.62	4.65/4.32	0.00
	^{15}N/^{13}C				143.1	97.80	93.94	75.38	74.82	83.81	69.97	
G10	^{1}H	13.49		8.07/5.93	7.53		5.83	4.69	4.40	4.54	4.48/4.18	−1.87
	^{15}N/^{13}C		146.64		137.2		92.98	75.45	73.23	82.22	66.47	
U11	^{1}H	12.18			7.73	5.43	5.49	4.11	4.53	4.40	4.54/4.06	−2.72
	^{15}N/^{13}C		157.30		141.7	104.08	94.01	75.85	72.11	82.54	64.00	
C12	^{1}H			8.58/6.84	7.80	5.68	5.89	4.02	4.25	4.13	4.51/4.04	−2.69
^{15}N/^{13}C					144.4	97.98	92.22	77.44	70.05	83.65	65.59	

[a] All chemical shifts are given in ppm. The imino and amino chemical shifts are given for 5 °C. ^{1}H chemical shifts were referenced to the internal H_2O peak at 5.0 ppm, and the ^{15}N shifts were referenced to an external standard of liquid NH_3 at 0.0 ppm. The aromatic, sugar and proton, and carbon chemical shifts are given for 20 °C. ^{1}H chemical shifts were referenced relative to the internal HDO peak at 4.85 ppm. The ^{13}C chemical shifts were referenced relative to external TSP. The ^{31}P chemical shifts are given for 20 °C and were referenced relative to the internal HPO_4^{2-} resonance at 0.0 ppm. The ^{31}P chemical shift is listed with its 3′ nucleotide $^{5'}$NpN$^{3'}$. [b] The two-loop U-imino protons resonate at 10.91 (^{1}H)/156.18 (^{15}N) ppm and 11.86 (^{1}H)/157.84 (^{15}N) ppm and were not assigned to individual imino protons.

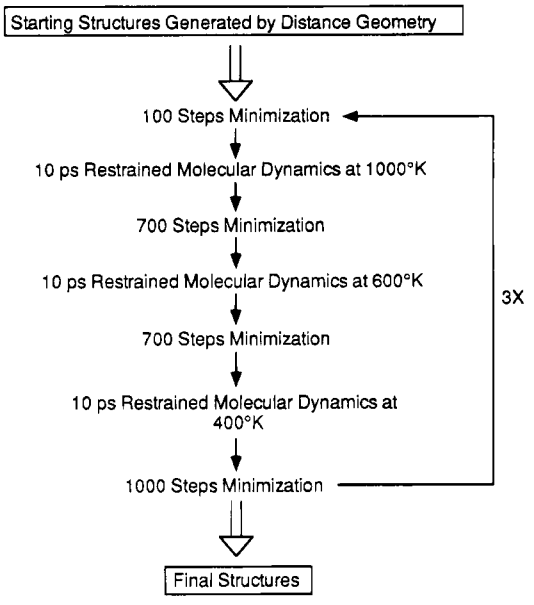

FIGURE 6: Flow chart of the structure refinement of the CUUG hairpin. The protocol is a high-temperature version of the "gentle refine" protocol provided by the program X-PLOR (Brünger, 1992).

Structure of the CUUG RNA Tetraloop. The structure of the CUUG RNA hairpin is shown in Figure 7. All the glycosidic angles in the molecule are in the anti range. The four CUUG loop nucleotides adopt 2′ endo sugar puckers, whereas the rest of the nucleotides are in 3′ endo conformation. The first and last nucleotides of the tetraloop form a Watson−Crick C5-G8 base pair, and the stacking continues from the stem to the loop C5-G8 base pair. The first U in the loop, U6, is folded into the minor groove, interacting with the last stem G4-C9 base pair and the loop C5-G8 base pair. The second U in the loop, U7, is in the major groove and is stacked onto the loop C5-G8 base pair in some of the structures. The last base pair of the stem interacts with the U6 in the minor groove and thus forms a part of this structural motif.

DISCUSSION

Characteristic Chemical Shifts of Loop Resonances. All proton resonances were assigned as described in the previous section, except the two imino protons of the loop U's which showed no NOEs and therefore did not contain any structural information. The U7H1′ resonance is shifted significantly upfield from its standard chemical shift, to 4.92 ppm. Examination of the loop structure reveals that this H1′ is situated above the purine ring of G8, indicating that the ring current effect of G8 is causing the unusual upfield shift of the U7H1′ (Wüthrich, 1986).

The chemical shifts of all protonated carbons have been determined. The most unusual chemical shifts are the C3′ resonances of the CUUG residues in the loop that are shifted into the region characteristic for the 4′ carbons (see Table 1). These chemical shifts are consistent with the observation that the sugar carbon chemical shift may be correlated with sugar pucker (Varani & Tinoco, 1991a) because these residues have 2′ endo sugar puckers. A similar trend is observed for the loop CUUG C4′ carbons that are also shifted slightly downfield from the rest of the C4′ resonances. In contrast, the C1′ resonances of the CUUG loop nucleotides appear to be shifted upfield by at least 7 ppm. Although these trends are intriguing, additional data are needed before sugar carbon chemical shifts can be reliably correlated with the sugar pucker.

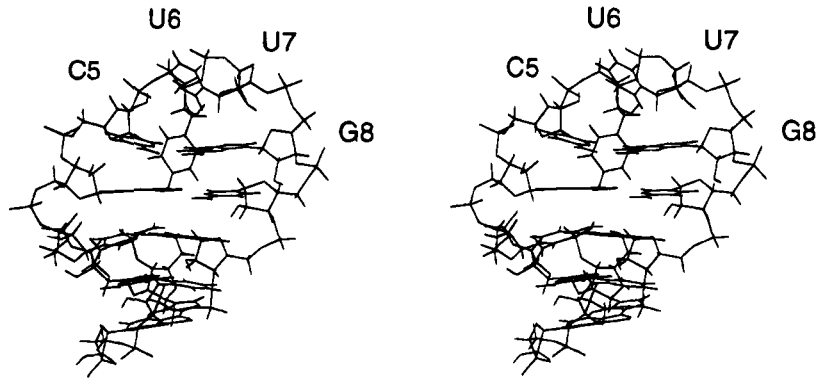

FIGURE 7: Stereoview of the structure of the CUUG RNA hairpin. The average structure is shown here. A Watson–Crick C-G base pair is formed in the loop, and the first loop uridine, U6, is folded into the minor groove.

Phosphorus chemical shifts have been proposed to be correlated with unusual backbone structures (Roongta et al., 1990). Two phosphorus resonances in the CUUG loop have nonstandard chemical shifts. The downfield resonance at 0.0 ppm was assigned to G8pC9, and the upfield shifted resonance was assigned to U6pU7 (see Figure 5). Although these phosphates are found in unusual conformations compared to A-form geometry, other loop phosphates are also found in unusual conformations and do not display unusual chemical shifts. Given the wide range of phosphorus chemical shifts observed in other RNAs (Legault & Pardi, 1994), especially tRNA (Salemink et al., 1979), we do not believe that ^{31}P chemical shift data can be directly used to generate structural information in RNA. Therefore, no structural constraints were derived from the ^{31}P chemical shifts.

Loop Dynamics. Chemical exchange processes with lifetimes of microseconds to milliseconds can lead to significant broadening of resonance lines (Jardetzky & Roberts, 1981). Such chemical exchange processes were observed in the CUUG loop where several of the protons and carbons showed broad resonance lines as compared to resonances in the stem. Inspection of a NOESY spectrum collected at 5 °C in H_2O revealed that the G8 imino proton exists in at least two conformations in slow to intermediate exchange (Jucker Franzusoff, 1995). A slightly different pattern of NOE cross-peaks was observed for these two resonances, indicating that G8 exists as two stable conformers under these conditions. No evidence was observed for any resonances being in the slow exchange regime at 20 °C. However, the increased line widths observed for G8H8 and for some resonances of U7 and C9 indicated that G8 and possibly other residues in the loop were undergoing motions in the intermediate to fast exchange time scale under the conditions used for acquiring the experimental data.

Due to evidence for dynamic behavior in the intermediate to fast chemical shift exchange regime for part of the CUUG loop, the NMR structural data were interpreted very conservatively (see Materials and Methods). U7 has a higher RMSD than other residues (see Table 2) and is poorly defined by the NOE data. The absence of NOE data combined with the significant number of resonances in intermediate exchange are consistent with extensive dynamics for this residue.

In contrast, U6 displayed sharp resonance lines, indicating that this nucleotide exists either as multiple conformers in fast exchange or as a single conformation. The NOEs observed involving U6 are all consistent with this base being bound in the minor groove, and it was assumed that U6 exists in a single conformation in solution.

Table 2: Proposed Hydrogen Bonding in the CUUG RNA Hairpin

hydrogen bond donor–acceptor[a]	hydrogen bond distance H–A (D–A)[b]	hydrogen bond angle donor–H···acceptor[c]	hydrogen bond angle H···O=C[d]
G8N2–G8N2H–U6O2	1.72 Å (2.67 Å)	156°	91°
G4N2–G4N2H–U6O4	2.13 Å (2.95 Å)	137°	88°
U6N3–U6H3–C9O2	1.92 Å (3.11 Å)	133°	162°

[a] Proposed hydrogen bonds between loop nucleotide U6 and functional groups in the minor groove. [b] The distance between the hydrogen (H) and the hydrogen bond acceptor (A) is given for the average structure. The distance between the hydrogen bond donor (D) and acceptor (A) is given in parentheses. [c] The hydrogen bond angle between the hydrogen bond donor atom and the hydrogen and between the acceptor atom and the hydrogen is given for the average structure. [d] The hydrogen bond angle between the hydrogen and the acceptor and the bond between the hydrogen bond acceptor and the directly attached carbon is given for the average structure.

Structure of the Loop C5-G8 Base Pair. C5 and G8 in the loop form a base pair and exhibit NOEs characteristic of a standard Watson–Crick C-G base pair. For example, the G8 imino proton shows NOEs to the C5 amino protons and, by spin diffusion, to C5-H5. The G8 imino proton resonates in the range typical for C-G base pairs, and the temperature dependence of this imino proton indicates that it is well-protected from exchange with solvent (data not shown). All these results are consistent with an imino proton in a C-G base pair. The NOE sequential walk observed in D_2O is also consistent with a C-G base pair and with continuation of the base stacking from the helix into the loop. The NOE pattern typical for helically stacked RNA is observed from the G4-C9 base pair of stem to the C5-G8 base pair; however, this pattern is interrupted between U6 and U7. In Figure 8A, the C5-G8 base pair is shown in an overlay of five structures onto the average structure. The C5-G8 base pair shows variable amounts of buckling. This variation may be a result of the limited structural constraints or, alternatively, may represent the dynamics of this base pair. The buckling of this base pair reduces the U6 phosphate to G8 phosphate distance, which is 12.7 Å in the average structure as compared to 17.5 Å in a standard C-G base pair. This buckling may be required to reduce the phosphate to phosphate distance so that it can be bridged by two nucleotides, thus allowing for the characteristic fold of the CUUG RNA hairpin loop.

Structure of the Loop Uridines. The two loop U's are the only nucleotides not involved in base pairs in this hairpin.

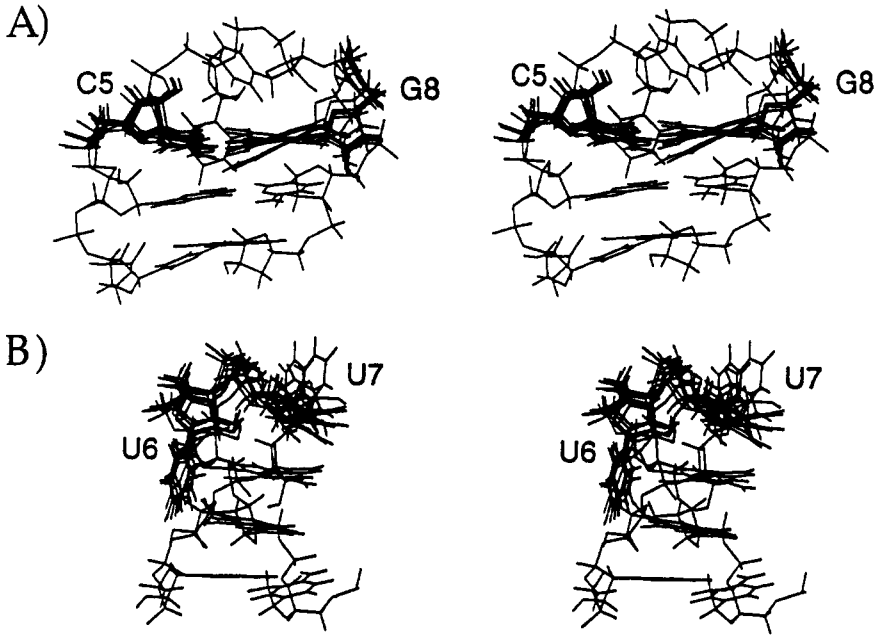

FIGURE 8: Stereoview of the superposition of five CUUG loop structures. For clarity, residues C3−G10 are only shown for one structure. (A) The loop C5-G8 base pair is displayed for the five superimposed structures, and (B) the loop nucleotides U6 and U7 are displayed for the five superimposed structures.

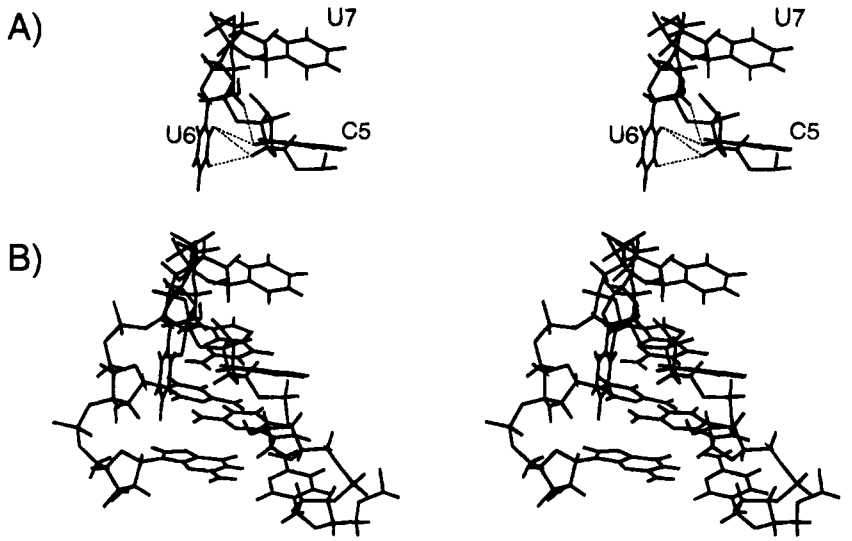

FIGURE 9: NOE-derived distance constraints that define U6 in the minor groove shown in a stereoview of the CUUG RNA hairpin. In part A, only nucleotides C5, U6, and U7 are displayed and dashed lines indicate the NOEs that define the position of U6 in the minor groove, the NOE between U6-H6 and C5-H1′, the NOE between U6-H6 and C5-H4′, the NOE between U6-H5 and C5-H4, and the NOE between U6-H2′ and C5-H1′. In part B, the entire loop and nucleotides C3-G10 are displayed in the same orientation as in part A.

Figure 8B shows U6 and U7 in an overlay of five structures onto the average structure. U6 is well-defined by NOE data and is folded into the minor groove. Figure 9 shows the NOEs between U6 and the sugar moiety of C5 that position U6 into the minor groove. U6-H6 gives a strong NOE to C5-H1′ and a weak NOE to C5-H4′, whereas U6-H5 shows no NOE to C5-H1′ but has a strong NOE to C5-H4′. In addition, a strong NOE is observed between U6-H2′ and C5-H1′. In the minor groove, U6 is making close contacts with the loop C5-G8 base pair and the stem G4-C9 base pair, allowing for van der Waals or hydrogen-bonding interactions. In Figure 10, the potentially stabilizing hydrogen-bonding or electrostatic interactions between U6 and the functional groups in the minor groove are schematically drawn, and Table 2 lists the hydrogen-bonding distances and angles involving U6 for the average structure. The U6-O2 can potentially hydrogen bond with the G8 amino proton of the loop C5-G8 base pair; the U6-O4 is within hydrogen-bonding distance of the G4 amino proton, and the U6 imino proton is in proximity of the C9-O2. Thus, the closing G4-C9 base pair is directly involved in formation of the CUUG tetraloop motif, and the structure indicates that a six-nucleotide sequence G(CUUG)C is required for this tetraloop motif.

The second loop uridine, U7, is found in the major groove but is not well-defined by the NMR data. The U7-H1′ is positioned over the base of G8, which is consistent with the upfield shift of the H1′ resonance to 4.92 ppm, due to the ring current of G8. If U7 were stacked onto G8, typical NOEs between G8-H8 and the U7 sugar would be expected. A small NOE is observed between G8-H8 and U7-H2′; however, many loop protons have broad resonances, and therefore medium to long inter-proton distances will not give

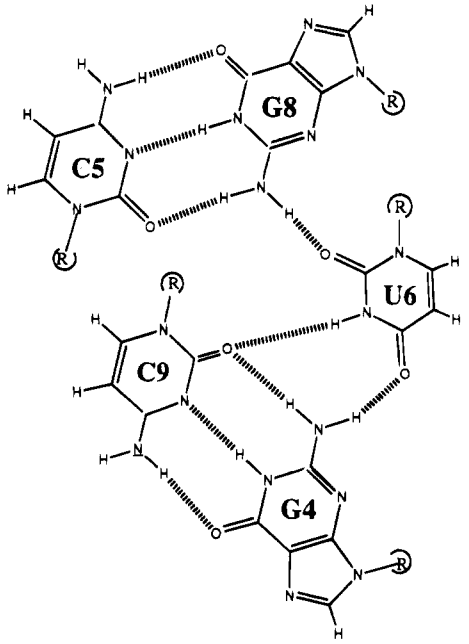

FIGURE 10: Schematic representation of the potential hydrogen-bonding interactions in the minor groove, between U6 and the loop C5-G8 base pair (between U6-O2 and G8 amino group, and between U6-O4 and G4 amino and between U6) and between U6 and the closing G4-C9 base pair (between U6-H3 and C9-O2). The figure is not drawn to scale, and is not meant to imply that all the bases are in the plane.

rise to observable NOE cross-peaks. One NOE was observed in H_2O between the G8 imino proton and U7-H5, indicating that U7 is stacked on G8 at least part of the time. However, since this one NOE distance constraint would force U7 to stack onto G8, it was not used in the structure calculations; therefore, U7 is relatively disordered in the final set of structures. Since U7 does not interact through hydrogen bonding with other parts of the loop, the structure suggests that this nucleotide can be replaced by other bases and a G(CUNG)C sequence motif can be proposed for the CUUG RNA tetraloop.

Structure of the Stem and the Closing G4-C9 Base Pair. NOEs characteristic of A-form helices are observed in the stem. For example, sequential H1′ to imino proton connectivities (Heus & Pardi, 1991a) are present for all the stem base pairs, and sequential imino proton to imino proton NOEs are observed at long mixing times (400 ms) (Wüthrich, 1986; Varani & Tinoco, 1991b). Since the second base pair in the stem is a G-U wobble pair and the closing G4-C9 base pair is part of the loop motif, interacting with U6, it is not surprising that some deviations from A-form are observed in the stem. The minor groove is slightly narrower than standard A-form (C1′-C1′ $\approx$ 10.1 Å) and widens at the G4-C9 base pair to ~10.9 Å, to accommodate the loop U6. This widening of the minor groove is achieved by moving C9 toward the major groove, thus altering the stacking pattern on the 3′ side of the helix.

Sugar Pucker and Glycosidic Angles in the CUUG RNA Hairpin. The torsion angles of the CUUG RNA hairpin were analyzed from the 13 final structures, and the results are given in Table 5 in the supporting information. All the stem nucleotides and the closing G4-C9 base pair had 3′ endo sugar pucker, whereas the four CUUG loop nucleotides had H1′-H2′ J coupling constants consistent with 2′ endo sugar conformations (Wüthrich, 1986; Varani & Tinoco, 1991b). In some of the solution structures, U7 exhibited a trend toward 3′ endo sugar pucker. Unfortunately, the increased line width of U7-H1′ and a resonance frequency for this proton close to water make accurate measurement of this H1′-H2′ J coupling constant difficult, and the relative populations of the two conformers could not be estimated (Altona & Sundaralingam, 1972). Therefore, the NMR spectra are consistent with U7 being either in 2′ endo conformation or in exchange between the two puckering modes.

The glycosidic angle is anti for all stem and loop nucleotides. U7 showed the largest variations in the glycosidic angle, consistent with the high RMSD of this nucleotide, and values in the high anti range were found. However, all U7 glycosidic angles fall within the expected range for 2′ endo sugar conformation (Saenger, 1984). The glycosidic angle for G8 is found in the high anti range, at the high end of what is usually observed for base-paired 2′ endo purines (Saenger, 1984). The G8 amino group is interacting with the U6 carbonyl O2 group, and it has been suggested that such interactions might shift the glycosidic angle to the high anti range (Olson, 1973; Yathindra & Sundaralingam, 1973; Lespinasse, 1976).

Backbone Torsion Angles in the CUUG RNA Hairpin. For A-form helices, the standard backbone conformation is gauche$^-$, trans, gauche$^+$, gauche$^+$, trans, and gauche$^-$ for α, β, γ, δ, ϵ, and ζ, respectively (Saenger, 1984). For the 5′ stem nucleotides, including G4 from the closing base pair, the torsion angles fall within the standard A-form range (see Table 5 in supporting information), but C9 and G10 on the 3′ strand show slight deviations from standard A-form. The γ torsion angles for these nucleotides are trans, but the two terminal nucleotides, U11 and C12, again show standard A-form torsion angles. The altered γ torsion angle permits the base of C9 to move toward the major groove, thus widening the minor groove to accommodate U6.

In order for the loop nucleotides to bridge the double helix, the backbone must deviate from standard A-form geometry. C5α is the first torsion angle in the sequence that diverges from typical A-form values, thus indicating that the backbone is changing direction here. There is presently no NMR method for directly measuring α and ζ torsion angles in nucleic acids; instead, these torsion angles are only indirectly restricted by steric, NOE, or other torsion angle restraints (Wüthrich, 1986). The phosphate at G4ζ/C5α is found in gauche$^-$/trans conformation, and γ falls into the trans range, which gives the only sterically allowed conformation for this ζ/α range that permits a helical conformation (Sasisekharan & Pattabiraman, 1978; Saenger, 1984). The C5ζ/U6α torsion angles are found in gauche$^-$/trans with γ in gauche$^+$ conformation. This conformation permits the characteristic loop structure with the base of U6 folded into the minor groove. The torsion angles of C5 and U6 are among the best defined in this hairpin, but high standard deviations are observed for the U7 torsion angles. Thus, U7 is not well-defined by the NMR data, and multiple nucleotide conformations were obtained in the calculations. The nucleotide G8 adopts several conformations in the calculated structures, with various degrees of base pair buckling for C5-G8. The G8pC9 phosphate is upfield shifted and appears as a broad resonance in the NMR spectra, also indicating that this part of the molecule is undergoing conformational exchange.

Torsion angle deviations are found in a subset of the structures that do not affect the overall fold of the hairpin. Such concerted motions of torsion angles have been previously described (Shakked & Kennard, 1983; Kitamura et al., 1984; Saenger, 1984). In the CUUG RNA hairpin, deviations in ζ torsion angles are compensated by changes in the α, β, and/or ϵ torsion angles and deviations in α torsion angles are accompanied by changes in β, γ, and/or ζ torsion angles. A correlation is observed between γ torsion angles and the α torsion angle. In helical regions, γ is found in gauche$^+$ conformation, but for nucleotide G4, 7 of the 13 final structures have γ torsion angles in the gauche$^-$ range. All of these structures have gauche$^+$ G4α torsion angles, instead of the gauche$^-$ α torsion angle found in standard RNA helices. Such correlations between α and γ torsion angles have been observed for 3′ endo oligonucleotides (Shakked & Kennard, 1983). It therefore not only appears that ζ/α are closely correlated but also that the torsion angles α/γ are correlated (Kitamura et al., 1984; Saenger, 1984).

Comparison to Other RNA Tetraloops. RNA tetraloops have been previously classified into three categories, the UNCG, GNRA, and CUUG families (Woese et al., 1990). NMR solution structures have been determined for the UUCG loop (Cheong et al., 1990; Varani et al., 1991), and NMR (Heus & Pardi, 1991b; Orita et al., 1993; Szewczak et al., 1993) and X-ray structures (Pley et al., 1994) have been determined for the GNRA tetraloops. These RNA tetraloops differ from other four-nucleotide loops because of their unusual thermal stability relative to hairpins with UUUU and AAAA loops (Uhlenbeck, 1990; Antao et al., 1991; Antao & Tinoco, 1992). Both the UUCG and the GNRA hairpins gain at least part of their stability by forming a base pair between the first and the last nucleotide in the loop (Cheong et al., 1990; Heus & Pardi, 1991b), and the results here show that this is also true for the CUUG tetraloop. The initial NMR structure of the UUCG loop had a reverse wobble U-G base pair with the G nucleotide in syn conformation (Cheong et al., 1990); however, this structure has been recently revised where this U-G base pair does not form a reverse wobble but instead is stabilized by both base−base and base−sugar hydrogen bonds (Allain & Varani, 1995). In the GNRA loops, a non-Watson−Crick G-A base pair is formed [type X in Saenger (1994)] (Heus & Pardi, 1991b; Orita et al., 1993), and in the CUUG loop, a Watson−Crick C-G base pair is formed.

In all three tetraloop structures, the loop base pair reduces the phosphate to phosphate distance by ~5 Å, compared to the distance found for the closing base pair. The phosphate to phosphate distance in A-form helices is ~17.5 Å, which is too large a gap for two nucleotides to close (Saenger, 1984). In the CUUG and UUCG tetraloops, the loop base pair is buckled, and in the GNRA tetraloop, a G-A base pair is formed that reduces the phosphate to phosphate distance, making it possible for all these loops to be closed by only two additional nucleotides. Thus, each of these tetraloops finds a different solution for forming an extra base pair in the loop while still being able bridge the phosphate backbone with two nucleotides.

Implications for Tertiary Interactions. Although extensive base stacking is a characteristic feature of both the UUCG and GNRA loops, a very different loop architecture is found in the CUUG loop. Here, the second nucleotide is folded into the minor groove and interacts with functional groups in the minor groove. In the GNRA and UUCG loops, the second loop nucleotide is freely accessible on top of the loop and the third nucleotide stacks in the loop and forms specific hydrogen bonds across the loop. In the CUUG RNA hairpin, the third nucleotide is on top of the loop in the major groove and no hydrogen-bonding interactions are formed across the loop. Therefore, the third nucleotide in the CUUG box is similar to the second nucleotide in the GNRA and UUCG loop. It has been proposed that this unstructured nucleotide (N) in the GNRA and UNCG hairpins is important for RNA−protein interactions (Heus & Pardi, 1991b; Monzingo & Robertus, 1992). A similar role could be ascribed to the third nucleotide in the CUNG loop. Thus, U7 may serve as a recognition element in tertiary interactions involving the CUUG RNA hairpin loop.

A novel tertiary interaction mode can be proposed from the structure of the CUUG RNA hairpin. The second nucleotide of the CUUG tetraloop is folded into the minor groove and is interacting with the functional groups in the minor groove. This uridine offers an example of how nucleotides can probe and recognize the minor groove, and it may be that similar interactions will be found in RNA−RNA tertiary interactions.

Phylogeny and Structure of the CUUG RNA Hairpin. The CUUG RNA hairpin loop is a naturally occurring motif that is found in loop 83 in bacterial 16S ribosomal RNA (positions 83−86 in *E. coli* 16S rRNA) (Woese et al., 1990). Besides the dominant CUUG loop sequence, the UUCG and the GCAA sequences are also found at this position in 16S rRNAs (Woese et al., 1990). Thus, the specific architecture of the CUUG RNA hairpin loop, with its U in the minor groove, is not required for the function of this particular hairpin in 16S rRNA. Possibly, the ability of tetraloops to stabilize adjacent stems may be the selected feature at this position, or the propensity of these sequences to form hairpins may be necessary to facilitate the correct folding of the nascent RNA chain. It is curious, however, that, from all the GNRA members, only the GCAA sequence occurs at this position and therefore other constraints must exist on the sequence of this hairpin loop. Noller and co-workers applied chemical and enzymatic probing experiments to *E. coli* 16S rRNA which has the CUUG hairpin in position 83 (Moazed, 1986). Our NMR solution structure predicts that the loop C-G base pair and the first loop U would be protected, whereas the second loop U should be fully accessible to the chemical probes. The chemical probing results show that the loop C-G base pair is indeed protected, with partial protection of the first loop U. Surprisingly, the second loop U is well-protected in the naked RNA and becomes partially protected in the 30S subunit. It has also been proposed that this hairpin is not involved in any tertiary interactions, on the basis of the variable length of the underlying helix (Woese et al., 1990). A possible interpretation of the NMR and chemical probing data is that this hairpin forms some tertiary interaction involving the second U of the CUUG loop. This interpretation would explain why only some tetraloop sequences are tolerated in this hairpin. Each tetraloop has one nucleotide that is protruding into the solution, ideal for engaging in tertiary interactions. In the GCAA and UUCG loops, it is the second nucleotide, and for the CUUG loops, it is the third nucleotide. In these three

sequences, the protruding nucleotide is a pyrimidine and the flexibility of these nucleotides could allow for specific interactions with a common partner.

Phylogenetic comparison of ribosomal RNA reveals that tetraloops often have a conserved closing stem base pair (Woese et al., 1990). The CUUG RNA hairpin shows a strong preference to be closed by a G-C base pair. When a CUUG tetraloop is replaced by the UUCG loop, the closing base pair is switched to C-G, the preferred closing base pair for UUCG loops. The phylogeny therefore suggests that tetraloops are six-nucleotide motifs. The importance of the closing base pair has also been observed in thermodynamic studies on the UUCG hairpin (Antao et al., 1991), where switching the closing base pair reduced the melting temperature by 10 °C. The solution structure determined here indicates a six-nucleotide motif G(CUNG)C for the CUUG RNA hairpin. The C9 carbonyl oxygen and the G4 amino group are stabilized by hydrogen-bonding or electrostatic interactions with the uridine in the minor groove. Thus, the structure of the CUUG RNA hairpin offers an explanation for why the closing base pair is conserved in this hairpin.

Comparison of the CUUG RNA Hairpin Structure to DNA Hairpins. Thermodynamic studies on RNA and DNA hairpins revealed that the UUCG RNA tetraloop has a much higher melting temperature than its DNA counterpart (TTCG) (Antao et al., 1991; Antao & Tinoco, 1992). In contrast, the melting temperatures are similar for the r-CUUG and the d-CTTG hairpins, and both hairpins show higher thermodynamic stability compared to the UUUU or TTTT hairpins with the same stem nucleotides (Blommers, 1989; Antao & Tinoco, 1992). NMR investigations showed that in the d-^{a}CTAG hairpin loop (^{a}C = arabinofuranosylcytosine) a Watson−Crick ^{a}C-G base pair is formed and the loop T is folded into the minor groove (Pieters et al., 1990). The loop A is stacked onto the ^{a}C, and the loop has a structure very similar to the RNA CUUG hairpin. NMR and model building studies suggested a very similar structure for the d-CTTG hairpin loop (Blommers, 1989; Blommers et al., 1991). Therefore, for the CUUG/CTTG tetraloop sequence, both DNA and RNA can adopt a similar loop structure, and these results are also consistent with the G(CUNG)C loop motif proposed for the CUUG RNA hairpin.

CONCLUSIONS

The CUUG RNA hairpin displays a novel RNA loop structure. A Watson−Crick base pair is formed between the first and last loop nucleotide, and the second nucleotide (U6) is folded into the minor groove. Both carbonyl oxygens and the imino proton of this U6 are interacting with minor groove substituents of the loop C5-G8 base pair and the closing G4-C9 base pair in the stem. Like the GNRA and UUCG RNA tetraloops, the CUUG RNA hairpin is thermodynamically stabilized by the intraloop C-G base pair. The NMR solution structure of this hairpin explains why the closing base pair is conserved along with the CUUG loop sequence, therefore indicating a six-nucleotide G(CUNG)C motif for the CUUG RNA tetraloop. This CUUG structure provides an example of how nucleotides can bind specifically in the minor groove, and it will be interesting to see if this mode of interaction is found in other inter- or intramolecular RNA−RNA recognition motifs.

ACKNOWLEDGMENT

We thank Arthur Zaug for help in the RNA sequencing, Kathleen Morden for help with structure analysis, and Robin Gutell for help with phylogenetic comparisons and useful discussions.

SUPPORTING INFORMATION AVAILABLE

Five tables giving the detailed acquisition and processing parameters for the homonuclear 2D NMR experiments, the heteronuclear 2D NMR experiments, and the heteronuclear 3D NMR experiments, a list of distance constraints for exchangeable protons, and an analysis of the torsion angles of the CUUG RNA hairpin from the 13 final structures (5 pages). Ordering information is given on any current masthead page.

REFERENCES

Allain, F. H. T., & Varani, G. (1995) *J. Mol. Biol. 250*, 333.
Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas 101*, 413.
Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205.
Antao, V. P., & Tinoco, I. J. (1992) *Nucleic Acids Res. 20*, 819.
Antao, V. P., Lai, S. Y., & Tinoco, I., Jr. (1991) *Nucleic Acids Res. 19*, 5901.
Bax, A., Griffey, R. G., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301.
Blommers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1989) *Biochemistry 28*, 7491.
Blommers, M. J. J., van de Ven, F. J., van der Marel, G. A., van Boom, J., & Hilbers, C. W. (1991) *Eur. J. Biochem. 201*, 33.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521.
Brünger, A. T. (1992) *X-PLOR Version 3.1 Manual*, Yale University, New Haven, CT.
Cheong, C., Varani, G., & Tinoco, I., Jr. (1990) *Nature 346*, 680.
Davanloo, P., Rosenberg, A. H., Dunn, J. J., & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035.
Donis-Keller, H. (1980) *Nucleic Acids Res. 8*, 3133.
Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977) *Nucelic Acids Res. 4*, 2527.
Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588.
Fesik, S. W., & Zuiderweg, E. R. P. (1990) *Q. Rev. Biophys. 23*, 97.
Glueck, A., Endo, Y., & Wool, I. G. (1994) *Nucleic Acids Res. 22*, 321.
Haasnoot, C., de Leeuw, F. A. A. M., & Altona, C. (1980) *Tetrahedron 36*, 2783.
Haasnoot, C., de Leeuw, F. A. A. M., de Leeuw, H. P. M., & Altona, C. (1981) *Org. Magn. Reson. 15*, 43.
Heus, H. A., & Pardi, A. (1991a) *J. Am. Chem. Soc. 113*, 4360.
Heus, H. A., & Pardi, A. (1991b) *Science 253*, 191.
Hore, P. J. (1983) *J. Magn. Reson. 54*, 539.
Jaeger, L., Michel, F., & Westhof, E. (1994) *J. Mol. Biol. 236*, 1271.
Jardetzky, O., & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, Academic Press, Inc., Orlando, FL.
Jucker Franzusoff, F. M. (1995) Ph.D. Thesis, University of Colorado at Boulder, Boulder, CO.
Kitamura, K., Mizuno, H., Amisaki, T., Tomita, K., & Baba, Y. (1984) *Biopolymers 23*, 1169.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1.
Legault, P., & Pardi, A. (1994) *J. Magn. Reson., Ser. B 103*, 82.
Lespinasse, J. N., Broch, H., Cornillon, R., & Vasilescu, D. (1976) *J. Theor. Biol. 57*, 225.
Macura, S., & Ernst, R. R. (1979) *J. Magn. Reson. 46*, 269.
Marion, D., Ikure, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393.
Marion, D., Kay, L. E., Sparks, W. W., Torchia, D. A., & Bax, A. (1989b) *J. Am. Chem. Soc. 111*, 1515.

Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 581.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. D. (1987) *Nucleic Acids Res. 15*, 8783.
Moazed, D., Stern, S., & Noller, H. (1986) *J. Mol. Biol. 187*, 399.
Monzingo, A. F., & Robertus, J. D. (1992) *J. Mol. Biol. 227*, 1136.
Mueller, L. (1979) *J. Am. Chem. Soc. 101*, 4481.
Murphy, F. L., & Cech, T. R. (1994) *J. Mol. Biol. 236*, 49.
Nierhaus, K. H., Schilling-Bartetzko, S., & Twardowski, T. (1992) *Biochimie 74*, 403.
Nikonowicz, E. P., & Pardi, A. (1993) *J. Mol. Biol. 232*, 1141.
Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M., & Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507.
Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591.
Olson, W. K. (1973) *Biopolymers 12*, 1787.
Orita, M., Nishikawa, F., Shimayama, T., Taira, K., Endo, Y., & Nishikawa, S. (1993) *Nucleic Acids Res. 21*, 5670.
Pardi, A., & Nikonowicz, E. P. (1992) *J. Am. Chem. Soc. 114*, 9202.
Patel, D. J., Shapiro, L., & Hare, D. (1987) *Q. Rev. Biophys. 20*, 35.
Pieters, J. M., de Vroom, E., van der Marel, G. A., van Boom, J. H., Koning, T. M., Kaptein, R., & Altona, C. (1990) *Biochemistry 29*, 788.
Pley, H., Flaherty, K. M., & McKay, D. B. (1994) *Nature 372*, 111.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479.
Redfield, A. G. (1987) *Chem. Phys. Lett. 96*, 537.
Roongta, V. A., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer Verlag, New York.
Salemink, P. J. M., Swarthof, T., & Hilbers, C. W. (1979) *Biochemistry 18*, 3477.
Sasisekharan, V., & Pattabiraman, N. (1978) *Nature 275*, 159.
Selinger, D., Brennwald, P., Liao, X., & Wise, J. A. (1993a) *Mol. Cell. Biol. 13*, 1353.
Selinger, D., Liao, X., & Wise, J. A. (1993b) *Proc. Natl. Acad. Sci. U.S.A. 90*, 5409.
Shakked, Z., & Kennard, O. (1983) *The A-form of DNA*, John Wiley & Sons, New York.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286.
Szewczak, A. A., Moore, P. B., Chang, Y. L., & Wool, I. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9581.
Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N., & Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364.
Uhlenbeck, O. C. (1990) *Nature 346*, 613.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1.
Varani, G., & Tinoco, I., Jr. (1991a) *J. Am. Chem. Soc. 113*, 9349.
Varani, G., & Tinoco, I., Jr. (1991b) *Q. Rev. Biophys. 24*, 479.
Varani, G., Cheong, C., & Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280.
Wijmenga, S. S., Mooren, M. M. W., & Hilbers, C. W. (1993) in *NMR of Macromolecules* (Roberts, G. C. K., Ed.) pp 217, Oxford University Press, New York.
Woese, C. R., Winker, S., & Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A 87*, 8467.
Wool, I. G., Glueck, A., & Endo, Y. (1992) *Trends Biochem. Sci. 17*, 266.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
Yathindra, N., & Sundaralingam, M. (1973) *Biopolymers 12*, 2075.
Zuiderweg, E. R. P., McIntosh, L. P., Dahlquist, F. W., & Fesik, S. W. (1990) *J. Magn. Reson. 86*, 210.
Zwieb, C. (1991) *Nucleic Acids Res. 19*, 2955.
Zwieb, C. (1992) *Nucleic Acids Res. 20*, 4397.

BI9515174