Structure of an Unusually Stable RNA Hairpin†

Gabriele Varani, Chaejoon Cheong,‡ and Ignacio Tinoco, Jr.*
Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, California 94720
Received October 9, 1990; Revised Manuscript Received January 2, 1991

ABSTRACT: The structure of a very common RNA hairpin, 5'GGAC(UUCG)GUCC, has been determined in solution by NMR spectroscopy. The loop sequence, UUCG, occurs exceptionally often in ribosomal and other RNAs, and may serve as a nucleation site for RNA folding and as a protein recognition site. Reverse transcriptase cannot read through this loop, although it normally transcribes RNA secondary structure motifs. A hairpin with that loop displays unusually high thermodynamic stability; its stability decreases when conserved nucleotides are mutated. The three-dimensional structure for the hairpin was derived from interproton distances and scalar coupling constants determined by NMR using distance geometry, followed by restrained energy minimization. The structure was well-defined despite the conservative use of interproton distances, by constraining the backbone conformation by means of scalar coupling measurements. A mismatch G·U base pair, with *syn*-guanosine, closes the stem. This hairpin has a loop of only two nucleotides; both adopt $C_{2'}$-endo sugar pucker. A sharp turn in the phosphodiester backbone is stabilized by a specific cytosine–phosphate contact, probably a hydrogen bond, and by stacking of the cytosine nucleotide on the G·U base pair. The structural features of the loop can explain the unusual thermodynamic stability of this hairpin and its sensitivity to mutations of loop nucleotides.

RNA hairpin loops are very important to the biological function of RNAs. Several different proteins recognize RNA hairpins; among them are the R17 and QB coat proteins (Wu & Uhlenbeck, 1987; Witherell & Uhlenbeck, 1989), *Tat* and *Rev* from HIV (Heophy et al., 1990; Malin et al., 1990), and bacteriophage antiterminators (Lazinski et al., 1989). A stem–loop structure is also present just 5′ to transcription termination sites in *Escherichia coli*. Hairpins are one of the building blocks of the secondary structure of ribosomal and messenger RNAs (Noller, 1984). Stable hairpins may provide nucleation sites to direct the folding of RNAs with complex secondary and tertiary interactions. Not surprisingly, the size and sequence of hairpin loops are evolutionarily conserved. Most RNA hairpins in ribosomal RNAs and in *E. coli* transcription terminators have tetranucleotide loops; two families of sequences, -UNCG- and -GNRA-, account for ≈70% of all tetraloops in 16S rRNA (Woese et al., 1990).

Despite their biological functions, little is known about the structure of RNA hairpins at atomic resolution, with the exception of tRNA [see Saenger (1984) and references cited therein]. In tRNA, however, loops are much larger (seven to eight nucleotides) than in rRNA. At a more qualitative level, several DNA hairpins have been characterized by nuclear magnetic resonance (NMR)[1] spectroscopy (Hare & Reid, 1986; Blommers et al., 1989; Williamson & Boxer, 1989). Two RNA hairpins have also been investigated, including one with the same -UUCG- loop studied here (Sakata et al., 1990; Puglisi et al., 1990). Although biologically less frequent, DNA hairpins have been studied more extensively because of advantages in synthesis and spectral assignments. Very large amounts of essentially pure RNA can now be synthesized in vitro by using T7 RNA polymerase (Milligan et al., 1987). Spectral assignments for RNA sugar protons have been accomplished recently by using several correlated experiments to resolve overlapping cross-peaks in oligonucleotides up to about 30 nucleotides in length, or 10 000 molecular weight (Varani et al., 1989; Puglisi et al., 1990). Oligonucleotides of small size (10–15 nucleotides) can fold into hairpins, and conformational differences between the loop and the stem are expected to provide contrast in the patterns of interproton distances and scalar couplings. Thus, RNA hairpins are ideal candidates for structural studies because of their important biological function, their useful NMR properties, and our ignorance of their structure.

The exceptionally stable and very common -UUCG- RNA hairpin (Woese et al., 1990) has been investigated by NMR, and its three-dimensional structure was presented in a preliminary communication (Cheong et al., 1990); Sakata et al. (1990) have reported more qualitative results. That loop sequence, with a C·G base pair closing the stem, is found very often in ribosomal RNAs and in intercistronic regions of the bacteriophage T4 messages (Tuerk et al., 1988). Here we describe in detail the determination of the hairpin structure by applying distance geometry to NOE and scalar coupling data. The precision with which the structure is known is presented by comparing 10 different structures generated by distance geometry to improve the confidence in the structural conclusions. The similarity of the structures improves the confidence in the conclusions.

Materials and Methods

RNA Synthesis and Purification. Two oligoribonucleotides, 5′pppGGAC(UUCG)GUCC (referred to as UUCG in the text) and 5′pppGGAC(UUUG)GUCC (UUUG), were synthesized enzymatically using T7 RNA polymerase and synthetic DNA templates (Milligan et al., 1987). Synthesis, purification, and sequencing of the RNA oligonucleotides were

† This research was supported in part by National Institutes of Health Grant GM 10840 and by U.S. Department of Energy, Office of Energy Research, Office of Health and Environmental Research, Grant DE-FG03-86ER60406.

‡ Present address: Department of Chemistry, Yale University, P.O. Box 6666, New Haven, CT 06511.

[1] Abbreviations: PAGE, polyacrylamide gel electrophoresis: NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; 2QF-COSY, two-quantum-filtered correlated spectroscopy; TOCSY, totally correlated spectroscopy; TSP, 3-(trimethylsilyl)-1-propanesulfonate.

0006-2960/91/0430-3280$02.50/0

done as previously reported (Varani et al., 1989; Puglisi et al., 1990); yields were 1–2 OD's of purified RNA per milliliter of reaction. Full-length RNA products ran on sequencing gels as shorter fragments, probably because these very stable RNA hairpins are not fully denatured in the gel. Therefore, it was not possible to characterize enzymatically the transcription products at the 3′ end. NMR data show that the oligoribonucleotides have both the correct length and nucleotide sequence. Gel filtration chromatography was done on a Bio-Rad TSK-125 silica gel column preequilibrated with the buffer of the thermodynamic and NMR experiments to verify that only one unimolecular species was present.

Thermodynamics. Thermodynamic parameters for secondary structure formation were obtained by UV absorbance versus temperature profiles (Puglisi & Tinoco, 1989). Data were recorded at 260 nm on a thermoelectrically controlled Gilford spectrophotometer. The temperature was increased at the rate of 1 °C/min. Samples were dialyzed overnight against doubly distilled water, resuspended in degassed 10 mM phosphate buffer containing 0.01 mM EDTA (pH 6.5), and preheated before each measurement. Data were collected over a 100-fold range of nucleotide concentrations (≈0.01–≈1 mM) using cells of different path length. Thermodynamic parameters for hairpin formation were calculated by assuming a two-state model and employing standard techniques of extrapolation of upper and lower base lines. Standard entropies and enthalpies were obtained from the temperature dependence of the equilibrium constant for the single-strand to hairpin transition.

NMR Spectroscopy. NMR samples were dialyzed extensively against 10 mM sodium phosphate/0.01 mM EDTA (pH 6.5). Samples used in exchangeable proton experiments were prepared by adding a 10:1 mixture of H_2O and D_2O to dry RNA. For nonexchangeable proton experiments, approximately 3 mg of RNA was lyophilized several times with 99.8% D_2O and dissolved in 0.5 mL of 99.96% D_2O (Aldrich) to a final concentration of ≈1.5 mM in strand. Most experiments were run on a GE GN-500 spectrometer operating at 500 MHz for protons and at 202 MHz for phosphorus. Data were transferred to a μVAX and processed by using the program FTNMR (Hare Research, Inc.).

Exchangeable proton spectra were recorded with the 1331 solvent suppression scheme (Hore, 1983). The maximum of excitation was halfway between the imino and aromatic resonances with a sweep width of 10000 Hz. One-dimensional NOE experiments were done at different decoupler powers and preirradiation times (100–500 ms) to help distinguish true NOE's from spillover and spin-diffusion artifacts. NOE experiments were repeated at low pH (5.2) and temperature (0 °C) to reduce the rate of exchange with the solvent of exchangeable protons.

All 2D NMR spectra were recorded in the phase-sensitive mode using the TPPI method (Marion & Wüthrich, 1983) and preirradiation of the residual HDO peak. For all but the high-resolution experiments, 400–450 FID's of 2K complex data points were collected with 3200–4400-Hz sweep widths. Repetition delays were ≈1.8 s, and 80–96 scans were averaged for each FID, for an acquisition time of 20–24 h. Data were zero-filled to 1K real points in t_1 and apodized by using phase-shifted sine bells in both dimensions. The same data were routinely apodized with different phase shifts to improve resolution (15–30° phase shift) or sensitivity (70–90° phase shift).

NOESY spectra were recorded at different mixing times (60, 130, 200, 300, and 600 ms) as described by Macura and Ernst (1979). Zero-quantum contributions to cross-peaks between scalar-coupled protons were eliminated by incrementing the mixing time with t_1 (Macura et al., 1982). NOESY spectra in water were recorded at 250-ms mixing time at low pH (5.2) and temperature (0 °C). The jump-and-return sequence was used in different experiments on all pulses or only on the last pulse (Plateau & Gueron, 1982); the sweep width was 9615 Hz. A NOESY spectrum in water was also recorded at 600 MHz on a Bruker AMX-600 spectrometer with preirradiation of the solvent resonance.

Double-quantum spectra were recorded as proposed by Braunschwiler et al. (1983) with an excitation delay of 35 ms. Composite pulses were used for the mixing pulse and the π pulse in the middle of the excitation sequence (Muller et al., 1986). The sweep width (5000 Hz in both dimensions) led to folding-over of the spectrum in the two-quantum dimension (ω_1); this was tolerated to improve resolution. TOCSY experiments were acquired by using MLEV-17 for broad-band decoupling (Bax & Davis, 1985). A long mixing time (150 ms) maximized relayed magnetization transfer from the H1′ protons to the other sugar protons. High-resolution phosphorus-decoupled 2QF-COSY spectra were collected by using the standard pulse sequence with composite mixing pulses (Müller et al., 1986). Broad-band phosphorus decoupling was achieved by WALTZ; the decoupler power was carefully chosen by using a test sample to avoid heating, while still enabling good decoupling over the spectral range of interest. High resolution (≈1-Hz point in ω_2 with a 4464-Hz sweep width) was obtained by collecting a data set of 4K complex points and by zero-filling to 4K real points before Fourier transformation; 675 t_1 FID's of 80 acquisitions each were collected; zero-filling to 2K real points gave a digital resolution of ≈2-Hz point in ω_1. Spectra were apodized with 20°-shifted sine bells. Proton-detected heteronuclear ^1H–^{31}P COSY were recorded as proposed by Sklenar et al. (1986). The spectral widths were 1000 Hz in the ^{31}P dimension (ω_1) and 1500 Hz in the ^1H dimension (ω_2); 200 FID's of 2K complex points were recorded for a digital resolution (after zero-filling to 1K real points in t_1) of ≈1.5 Hz/point in ω_2 and ≈1 Hz/point in ω_1; 160 scans were averaged for each FID.

Interproton Distances and Scalar Coupling Measurements. Interproton distances between nonexchangeable protons were estimated from the buildup rates of cross-peak intensities in short mixing time NOESY's (60–200 ms). Cross-peak volume integration was performed with FTNMR. Apodization effects were found to be smaller than the experimental uncertainty in buildup rates by comparing data from spectra apodized with sine bells of different phase shift. Pyrimidine H5–H6 distances (2.41 Å) were used as reference. The buildup rates for the six well-isolated H5–H6 cross-peaks were the same, within experimental error, and the average value was used to reference all distances.

Since internal motion and spin diffusion can lead to errors in estimating interproton distances, the NOE information was used only semiquantitatively in defining the hairpin structure by means of distance geometry constraints, as usually done in protein studies (Wüthrich, 1986). NOE's between nonexchangeable protons were classified as strong, medium, weak, and, for some unambiguous cases, absent. The corresponding precisions on distance constraints were 1.8–3, 2–4, 2.5–5, and >4 Å, respectively. The intensities of cross-peaks involving exchangeable protons depend both on the interproton distance and on the rate of exchange with the solvent. The precision in distance constraints involving exchangeable protons was assumed to be 2–5 Å when NOE's were present at short

mixing time (100–150 ms), and >4 Å for some clearly absent NOE's. As the comparison of distance estimates from different controls described below suggests that the effects of internal motion and spin diffusion are not large, this is a conservative interpretation of the data. That is, any systematic error should be smaller than the uncertainty in the distance constraints used in structure determination.

To estimate the relevance of spin diffusion and internal motion, two isolated H5′-H5″ cross-peaks, one from the stem (A3) and one from the loop (C7), were used. The calculated H5′–H5″ distances (1.7 ± 0.2 Å) agree very well with that expected (1.78 Å), showing that spin-diffusion effects are small for H5′ and H5″ protons. Further controls were provided by intrasugar cross-peaks, since H1′–H2′, H2′–H3′ and H3′–H4′ distances are confined within narrow ranges by steric constraints. In all cases, distances calculated by using the H5–H6 distance as reference agree fairly well with those expected from the covalent structure. Therefore, spin diffusion is likely to be small for most protons at mixing times shorter than ≈100 ms, with the possible exception of some relatively distant protons. This is due to the small size of this molecule; for larger molecules, spin diffusion increases faster than sensitivity at increasing mixing time. These results also show that a single correlation time describes the motion of base, sugar, and backbone (H5′–H5″) protons, both in the stem and in the loop, within experimental uncertainty. However, it is not known whether the same correlation time also describes the relative motion of protons belonging to different moieties (i.e., base to sugar or base to backbone) or to different nucleotides.

^{1}H–^{1}H scalar couplings were measured in high-resolution 2QF-COSY spectra. Narrow line widths (≈1–1.5 Hz) and the simple ribose sugar spin system enabled accurate measurements of the coupling constants. Artifacts due to partial superposition of multiplet components (Wüthrich, 1986) were found to be negligible by comparing the same couplings obtained from cross-peaks with interchanged active and passive couplings. The uncertainty was estimated to be ≈1 Hz for well-resolved cross-peaks. For H4′–H5′, H4′–H5″, and H1′–H2′ protons, the cross-peaks were in most cases absent or very weak. An upper limit for the coupling was estimated to ≈2 Hz when the cross-peak was missing or very weak. Heteronuclear ^{1}H–^{31}P couplings were obtained from high-resolution heteronuclear COSY. Passive ^{1}H–^{1}H couplings lead to cancellation of mutiplet components, but the distance between the outer edges of the multiplet could be measured reliably for the stronger cross-peaks. The heteronuclear coupling was obtained by subtraction of all passive ^{1}H–^{1}H couplings, determined with good precision in homonuclear experiments, from the total multiplet width. In the case of some weak or clearly absent cross-peaks, only upper limits to the coupling (≈3 Hz) were estimated.

Structure Determination. Backbone torsion angle values consistent with measured scalar couplings were obtained as described under Results, typically with ±15–25° uncertainty. Torsion angle constraints were converted into distance constraints by using standard bond lengths and bond angles (Saenger, 1984). Interatomic distance constrains from the torsion angles and interproton distances based on NOE's were used in the distance geometry program DSPACE (Hare Research, Inc.). The total number of constraints was 415, approximately corresponding to the number of observed scalar couplings and NOE's; 125 intranucleotide interproton distances, 157 internucleotide (mostly sequential) interproton distances, 12 hydrogen bonds (11 from the stem and 1 from a G·U base pair), and 121 distances defining 86 backbone and sugar torsion angles were introduced. Refinement were performed by repeating cycles of simulated annealing. The structure was "heated", and minimization was performed while decreasing the kinetic energy of the system. Structures having a penalty function for violation of constraints at least 5 times larger than the best structures were considered to have not converged; only 10 of 35 trial structures converged. The remaining 25 structures are trapped in local minima; although it should be possible in principle to reach convergence for all structures, this would require an impractically long computer time. No deviation from experimental constraints larger than 0.006 Å was found in the converged structures, and all backbone angles were within the range defined by the scalar couplings.

After distance geometry, the converged structures were subject to constrained energy minimization using the program AMBER (Kollman et al., 1981). Approximately 160 interproton distance and sugar and backbone torsion angle constraints were added from the distance geometry structures. The weight of the constraints was determined by trial and error to compromise between the search for low conformational energy and the need to alter as little as possible the structures generated by distance geometry.

Only four classes of distance constraints (1.8–3, 2–4, 2.5–5, and >4 Å) were used in the determination of the structure. However, to estimate the precision of the structure, the original data—the directly measured cross-relaxation rates from the NOE buildup curves and the scalar couplings—were used to calculate the rms deviation, D, between experimental and calculated observables:

$$D^2 = \frac{1}{N}\sum_{j=1}^{N}[I_j(\text{exptl}) - I_j(\text{calcd})]^2/f_j[I_j(\text{exptl})]^2$$

where I_j represents measured scalar couplings or cross-relaxation rates, f stands for the experimental uncertainty, and the sum is over the N observables for which $f \leq 1$ (approximately 160 NOE's between nonexchangeable protons and 90 scalar couplings). It is also useful to define a local D, where the sum is extended only over the NOE's or scalar couplings involving one nucleotide or a pair of neighboring nucleotides. Observables were calculated from the final coordinates by means of Karplus equations (scalar coupling) or by assuming a single correlation time (NOE's). The generalized Karplus equations are approximately as accurate as the experimental uncertainty (Altona, 1982). With NOE's, however, some care should be used when judging a structure by the rms deviation between observed and calculated cross-relaxation rates because of the assumption of the absence of internal motion.

Results

Thermodynamics. UV absorption melting curves at RNA concentrations between 0.01 and 1 mM were independent of RNA concentration, as expected for unimolecular transitions. Size-exclusion chromatography confirmed the presence of a single species in low salt. A second species, probably a duplex containing mismatched base pairs and melting approximately 35 °C lower than the hairpin, is found in 20 mM $MgCl_2$ plus 1 M NaCl at or above 0.5 mM RNA. Thermodynamic parameters of hairpin formation are reported in Table I.

A hairpin with a -UUCG- loop but different stem displayed an unusually high melting temperature (Tuerk et al., 1988). The data of Table I confirm that the hairpin's unusual stability is due to the loop. Hairpins with the same stem sequence, but with A_4, C_4, or U_4 loops, melt about 20 °C lower (Groebe & Uhlenbeck, 1988; V. Antao, personal communication). A significant decrease in melting temperature (7 °C) and in-

Table I: Thermodynamic Parameters for Hairpin Formation in 10 mM Sodium Phosphate/0.01 mM EDTA, pH 6.5 (pH 6.7 for the -UUUG- Hairpin)

	T_m (°C)	$\Delta G^\circ_{25\,°C}$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ [cal/(mol·K)]
-UUCG-	71.5 ± 0.3	−7.9 ± 0.1	−58.6 ● 0.6	−170 ± 2
-UUUG-	64.8 ± 0.5	−6.5 ± 0.2	−54.9 ± 0.2	−163 ± 1

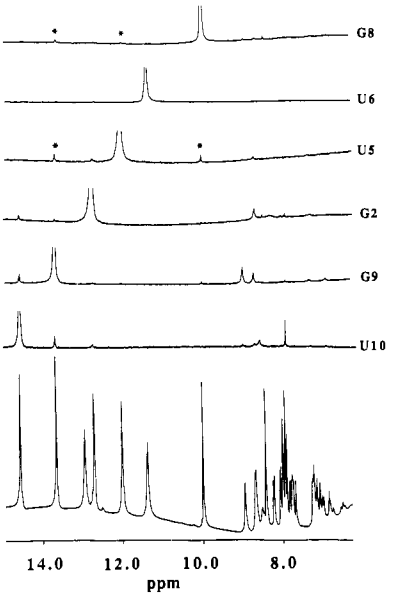

FIGURE 1: 1D NOE's at 0 °C and pH 5.2 for the UUCG hairpin. NOE's from U5 and G8 imino protons to each other, and to the neighboring (G9) imino proton, are marked by asterisks; these important NOE's were not observed at neutral pH. All spectra are referenced to TSP. No base-line correction was performed; 3-Hz line broadening was applied before Fourier transformation.

creased free energy of hairpin formation (+1.5 kcal/mol at 25 °C) occur upon changing -UUCG- to -UUUG-. These data parallel the results of Tuerk et al. (1988) and suggest that the presence of specific interactions in the loop stabilizes the hairpin structure.

Spectral Assignments. The imino proton spectra were assigned by standard methods using one- and two-dimensional NOE spectroscopy (Figure 1). At increasing temperature, the resonance at the end of the stem (G1) disappears first, followed by the loop resonances (U5 and U6); however, the G8 imino is as sharp as the stem resonances up to ≈50 °C. Amino protons from base-paired nucleotides were assigned from the strong NOE's to the corresponding G imino protons. The amino proton resonances from A3 and G9 broaden rapidly and disappear above ≈10 °C. NOESY cross-peaks from G imino to C H5 protons provided further support, once the nonexchangeable spectrum was assigned. NOE's due to the magnetization transfer pathway G imino → C amino′ → C amino″ → C H5—the second step is chemical exchange—were observed in NOESY spectra at long (250 ms) mixing time. It was not possible to assign the amino protons of C7 and G8; they are in fast or intermediate exchange. A few relatively sharp resonances at about 6.5–6.7 ppm show that at least two hydroxyl protons are protected from exchange with the solvent (Figure 1). NOESY cross-peaks to sugar protons observed at 600 MHz show that the slowly exchanging hydroxyl protons are from U5, U6, or C7, but conclusive sequence-specific assignments have not been possible so far.

The nonexchangeable proton spectrum for the UUCG hairpin is shown in Figure 2A; the ^{1}H-decoupled phosphorus spectrum is shown in Figure 2b. Spectral assignments for the stem resonances were obtained as previously reported (Varani et al., 1989; Puglisi et al., 1990). Sugar protons were unambiguously identified by means of correlated experiments (2QF-COSY, TOCSY, two-quantum). Most H1′–H2′ cross-peaks were missing or very weak, as expected for N-type sugar pucker (Altona, 1982), but H2′–H3′–H4′ connectivities were established in 2QF-COSY and two-quantum spectra. The unusually shifted G9 H1′ (4.38 ppm) could not be assigned by proton NMR due to overlap with G9 H2′ (also at 4.38 ppm); it was assigned by ^{13}C–^{1}H two-dimensional spectroscopy. The phosphorus spectrum was assigned by using the

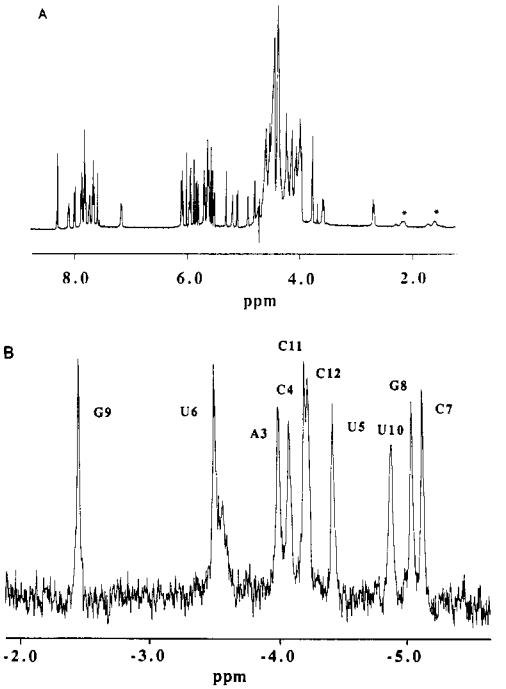

FIGURE 2: (A) Nonexchangeable proton spectrum and (B) proton-decoupled phosphorus spectrum for the UUCG hairpin in 10 mM sodium phosphate/0.01 mM EDTA (pH 6.5) at 26 °C. Proton chemical shifts are relative to TSP; the broad upfield resonances marked with asterisks are impurities. Phosphorus chemical shifts are referenced to an external TMP standard; the broad shoulder next to the U6 resonance is likely from G2. This spectrum was recorded at 161 MHz on a Bruker AMX-400 spectrometer; 1-Hz line broadening was applied before Fourier transformation.

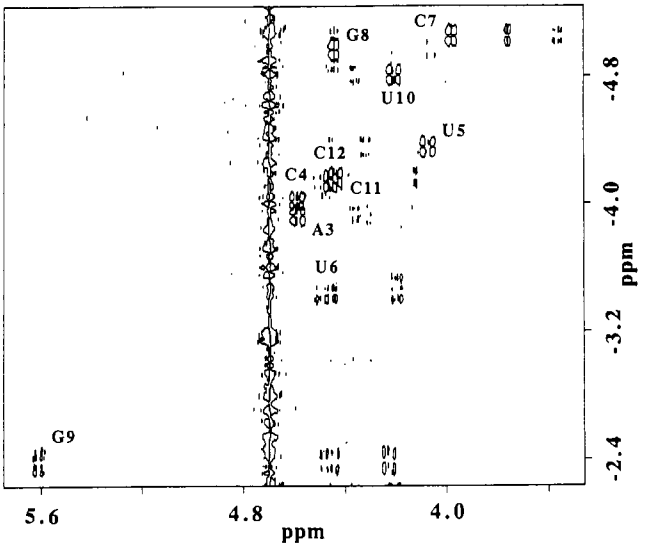

FIGURE 3: Heteronuclear COSY spectrum for the UUCG hairpin in 10 mM sodium phosphate/0.01 mM EDTA (pH 6.5) at 26 °C. Peaks corresponding to the internucleotide H3′(i − 1)–P(i) connectivities are labeled as the nucleotide 3′ to the phosphorus.

Table II: Chemical Shifts (ppm) of Proton (Relative to TSP) and Phosphorus (Relative to TMP) Resonances for the UUCG Hairpin at 26 °C in 10 mM Sodium Phosphate/0.01 mM EDTA[a]

	P	H8/6	H2/5	H1′	H2′	H3′	H4′	H5′/5″	imino/amino
5′ G1	na[b]	8.07	na	5.79	4.91	4.70	4.52	4.35/4.25	12.93
G2		7.57	na	5.87	4.61	4.58	4.53	4.45/4.28	12.68
A3	−4.11	7.86	7.79	5.99	4.59	4.58	4.48	4.58/4.12	8.50/6.81
C4	−3.95	7.14	5.18	5.30	4.36	4.07	4.38	4.37/3.97	8.67/7.23
U5	−4.40	7.72	5.62	5.62	**3.75**	4.49	4.32	4.45/4.04	11.99
U6	**−3.49**	7.98	5.82	6.06	4.64	3.98	4.44	4.19/4.01	11.36
C7	−5.09	7.67	6.09	5.93	4.06	4.44	**3.76**	**3.57/2.67**	
G8	−5.01	7.82	na	5.93	4.79	**5.61**	4.38	4.38/4.15	9.96
G9	**−2.44**	8.29	na	**4.38**	4.38	4.21	4.35	4.46/4.26	13.63
									8.92/6.81
U10	−4.85	7.80	5.10	5.54	4.52	4.46	4.36	4.48/3.99	14.53
C11	−4.19	7.87	5.67	5.57	4.23	4.44	4.39	4.51/4.03	8.64/7.18
C12	−4.21	7.64	5.52	5.70	3.95	4.13	4.11	4.46/3.98	8.39/7.18

[a] Unusual chemical shifts are in boldface. [b] Not applicable.

Table III: Chemical Shifts (ppm) of Proton and Phosphorus Resonances for the UUUG Hairpin at 26 °C in 10 mM Sodium Phosphate/0.01 mM EDTA[a]

	P	H8/6	H2/5	H1′	H2′	H3′	H4′	H5′/5″	imino/amino
U5	−3.84	7.62			4.02	4.36		4.33/4.01	
U6	−3.68		5.77	5.91	4.44	4.11	4.25	4.10/3.93	
U7	−4.59	7.50	5.75	5.80	4.18		3.94	**3.63/3.03**	11.36
G8	−4.31		na[b]			**5.49**			
G9	**−2.41**		na	**4.53**					
U10	−4.62								

[a] Only chemical shifts differing from those of the UUCG hairpin by more than 0.05 ppm are shown. Unusual chemical shifts are in boldface. [b] Not applicable.

Table IV: Scalar Couplings in Hertz for the UUCG Hairpin at 26 °C in 10 mM Sodium Phosphate[a]

	1′2′	2′3′	3′4′	4′5′/4′5″	3′P	P4′	P5′/P5″
5′ G1	3	5.5	9 ± 2	<3, 5		na[b]	na
G2	<2	4	9	<2.5, 4.5	8.5 ± 2		
A3	<2	5	10 ± 2	<3, <3	8.5 ± 2	<3	<5, ≈3
C4	<2	4.5	9 ± 1.5	<3, 3.5	9 ± 2	<3	<3, <3
U5	<2	4.5	10	<3, <3	7 ± 2	≈4	<3, <3
U6	**9**	5	**<2**	3.5, 3	≈8	≈4	6 ± 2, <3
C7	**8.5**	4.5	**<2**	<3, 3	8 ± 2	≈4	5 ± 2, <5
G8	<2	5	9		8 ± 2	<3	<3, <3
G9		4	10	<3, <3	8 ± 2	<3	**18 ± 3, 7.5 ± 3**
U10	<2	4.5	9.5	<3, <3	7 ± 2	≈4	<3, <3
C11	<2	4.5	10 ± 2	<3, <3	8 ± 2	≈3	≈3, <3
C12	2	5	9.5	4.5, <3		≈4	<5, <3

[a] Uncertainties are ±1 Hz, unless otherwise indicated. For the H5′ and H5″ protons, couplings to the downfield proton are reported first and couplings to the upfield proton second. Unusual couplings are in boldface. [b] Not applicable.

strong sequential H3′–P cross-peaks in heteronuclear COSY experiments (Figure 3) and the weaker intranucleotide phosphorus to H4′, H5′, and H5″ connectivities (Frey et al., 1985). Observation of strong H5′ to H5″ in 2QF-COSY spectra and weak heteronuclear H5′ and H5″ to ^{31}P cross-peaks led to unambiguous sequence-specific assignments of H5′ and H5″ protons. Further support came from some weak H4′ to H5′ and H5″cross-peaks in the 2QF-COSY. Stereospecific assignment of H5′ and H5″ resonances has not been possible so far.

Assignment of loop protons (U5–G8) required some caution, because structural assumptions are not possible. A-like connectivities between C4 and U5 enabled the assignment of U5, and G8 was identified from the very strong H8 to H1′ cross-peak, showing that the conformation around the glycosidic bond is syn. Sugar spin systems were identified in correlated experiments; a long mixing time TOCSY (150 ms) showed cross-peaks from H1′ to all other sugar protons. Both pyrimidines, U6 and C7, have large H1′ to H2′ scalar couplings, showing that the sugar pucker is S-type (Altona, 1982). Complete assignments were obtained from the sequential scalar connectivities H3′ → P → H4′, H5′, H5″, starting from U5 H3′ (Frey et al., 1985). The U5 H3′ (4.49)–P–U6 H4′ (4.44) connectivity (the two-peaks labeled U6 in Figure 3) establishes the assignments for the loop nucleotides. Assignments for the UUCG hairpin are reported in Table II. Resonances with significantly different chemical shifts for the UUUG hairpin are reported in Table III.

Scalar connectivities are necessary for reliable assignments. The assumption that NOE's in this loop are sequential, as seen in helical regions, would lead to erroneous assignments (Sakata et al., 1990). In this unusual structure, the base protons of C7, not those of U6, are in close contact with the U5 sugar protons. All cross-peaks observed in correlated experiments have been assigned, and all expected cross-peaks were observed in correlated experiments, except for the superposition of two sets of resonances (A3 H2′ and H3′, G8 H4′ and H5′). This completeness adds confidence to the assignments of the H5′ and H5″ protons, rarely assigned before. The quality of the 2QF-COSY spectrum enabled observation of cross-peaks between resonances separated by as few as 0.03 ppm. Phosphorus decoupling simplifies the multiplet structure, improves the sensitivity, and helps prevent some spectral overlap. The chemical shift dispersion is also very favorable (Table II).

Backbone Conformation. The first step in the structure determination was to obtain the conformation of the backbone,

Table V: Torsion Angles (Degrees) for the UUCG Hairpin from Distance Geometry followed by Restrained Energy Minimization[a]

	α P–O5′	β O5′–C5′	γ C5′–C4′	δ C4′–C3′	ε C3′–O3′	ζ O3′–P	χ	P	ϕ_m
5′G1	na[b]	na	na	84 ± 5	−145 ± 11	−70 ± 5	−177 ± 3	3 ± 9	39.5 ± 1.3
G2	−79 ± 4	173 ± 7	66 ± 4	85 ± 2	−160 ± 3	−60 ± 4	−171 ± 2	15 ± 7	38.3 ± 1.5
A3	−85 ± 3	−171 ± 3	61 ± 3	83 ± 3	−160 ± 3	−65 ± 5	−156 ± 4	34 ± 12	37.2 ± 1.0
C4	−110 ± 10	−170 ± 5	60 ± 3	81 ± 2	−164 ± 8	−78 ± 12	−134 ± 8	37 ± 8	39.6 ± 1.2
U5	−109 ± 15	−176 ± 8	62 ± 3	85 ± 5	160 ± 6	−80 ± 4	−126 ± 14	36 ± 8	39.2 ± 1.6
U6	−180 ± 10	163 ± 7	66 ± 7	148 ± 7	−81 ± 5	−88 ± 8	−164 ± 14	178 ± 20	33.9 ± 1.5
C7	−97 ± 17	177 ± 11	58 ± 5	153 ± 7	−84 ± 16	98 ± 22	−156 ± 6	169 ± 5	35.8 ± 1.2
G8	86 ± 12	−172 ± 10	169 ± 10	85 ± 2	−171 ± 3	−29 ± 19	36 ± 4	13 ± 7	38.2 ± 1.2
					−112 ± 2	−78 ± 5			
G9	−166 ± 11	−116 ± 16	90 ± 12	84 ± 4	−166 ± 10	−68 ± 8	−148 ± 8	25 ± 15	39.1 ± 1.3
	−34 ± 5	122 ± 7							
U10	−90 ± 5	172 ± 4	69 ± 7	82 ± 2	−165 ± 2	−67 ± 3	−152 ± 7	16 ± 4	39.9 ± 0.6
C11	−84 ± 4	176 ± 4	69 ± 2	81 ± 2	−164 ± 4	−67 ± 3	−163 ± 2	20 ± 4	40.1 ± 0.6
C12	−84 ± 4	−174 ± 4	60 ± 2	80 ± 2	na	na	−159 ± 2	23 ± 4	38.4 ± 0.8
A-RNA	−68	178	54	84	−153	−71	−158	18	

[a] Upper and lower limits found in the 10 converged structures are shown by the uncertainties. Classical A-form values from fiber diffraction studies are also shown (Saenger, 1984). χ is the glycosidic angle, *P* the sugar pucker, and ϕ_m the amplitude of pucker; *P* = 18° correspond to 3′-endo, *P* = 162° to 2′-endo. Two sets of angles are shown for G8 and G9, corresponding to two equally acceptable families of structures (see text). [b] Not applicable.

as defined by six torsion angles for each nucleotide (Saenger, 1984). Direct information on the six backbone torsion angles was provided by scalar couplings and phosphorus chemical shifts. Observed scalar couplings are reported in Table IV for the UUCG hairpin. Very similar values were found for the UUUG hairpin; poorer chemical shift dispersion led to somewhat larger uncertainties.

Information on the torsion angles ζ ($O_{3'}$–P) and α (P–$O_{5'}$) came from phosphorus chemical shifts. Both α and ζ were assigned gauche (g) conformations (≈−100° to ≈+100°) whenever the phosphorus resonance was not shifted downfield (Gorenstein, 1984). No assumptions were made about two downfield phosphorus resonances (U5pU6 and G8pG9); several conformational effects can cause this shift (Giessner-Prettre et al., 1984).

The conformations around β ($O_{5'}$–$C_{5'}$) and γ ($C_{5'}$–$C_{4'}$) were obtained from scalar couplings involving H5′ and H5″ without stereospecific assignments. With the exception of G9, all P–H5′ and P–H5″ couplings are less than 3–4 Hz (Table IV). Small couplings to both H5′ and H5″ restrict β to trans (t) (Altona, 1982). The large P–H5′ (or P–H5″) coupling of G9 shows that β is not trans. However, only stereospecific assignment of the H5′ and H5″ resonances would allow the choice between g^+ and g^-. Small (≤2–3 Hz) H4′–H5′ and H4′–H5″ couplings restrict γ to g^+ in all but one case (G8) (Altona, 1982); spectral overlap prevents the determination of γ of G8. The percentages of trans conformer for β and g^+ conformer for γ were estimated by using Altona's rules (Altona, 1982). Small couplings to H5′ and H5″ protons indicated almost pure conformers in many cases, with minor (≤10%) contributions from other than the tg^+ conformations. Minor conformers in rapid conformational equilibrium, as much as 30%, were present only at the 5′ end (G1–G2) of the hairpin. This is likely responsible for the nonobservation of any cross-peak involving the G1pG2 phosphorus resonance. Small but observable (3–5 Hz) four-bond P–H4′ couplings confirmed that the conformation for βγ is mostly tg^+. More precise values of β and γ were obtained by comparing measured scalar couplings with those specified by generalized Karplus' equations (Altona, 1982; Lankhorst et al., 1984).

The angle δ is related to the pseudorotation phase angle *P* and the amplitude of pucker ϕ_m (De Leeuw et al., 1980). The percentage of N-type (3′-endo) conformers was estimated from either the H1′–H2′ or the H3′–H4′ couplings (Altona, 1982). With the exception of G1 and C12, the sugar pucker was pure (>95%) N or S. Flexibility at the end of helical regions is commonly observed in RNA and DNA oligonucleotides. The 5′ end (G1) was more flexible than the 3′ end (C12); the opposite has generally been observed (Rinkel & Altona, 1986; Varani et al., 1989; Puglisi et al., 1990). For N-type sugars, the relatively large H3′–H4′ couplings are consistent with amplitudes of pucker in the upper range of those observed in crystals of model compounds (37–40°) (Saenger, 1984). More precise values (±10°) for *P* and δ were obtained from H2′–H3′ and H3′–H4′ (H1′–H2′ for S-type sugars) couplings (De Leeuw & Altona, 1982).

The angle ε was obtained from sequential H3′–P couplings. The g^+ conformer is never populated due to steric hindrances, but g^- and trans conformers are allowed, and cannot be distinguished without ^{13}C–^{31}P couplings (Altona, 1982). Splittings of 8–10 Hz are consistent with both g^- and trans conformers for all nucleotides; both conformations were allowed in the structure determination. The small (3–4 Hz) sequential H2′–P four-bond coupling of C7 is consistent only with ε being g^- and 2′-endo pucker (Altona, 1982).

Structure Determination. All converged structures have a similar value for the distance geometry penalty function. Thus, all 10 structures are good solutions to the problem of generating a structure using NMR-derived constraints. The stems were very similar in all structures; the overall folding was the same. The major differences were the backbone between G8 and G9 and the glycosidic angle of C7 (see below). rms deviations betweens structures with similar backbone between G8 and G9 were ≤1 Å. In all cases, rms deviations between different structures were <2 Å. The converged structures were subject to restrained energy minimization using AMBER (Kollman et al., 1981) to improve van der Waals and electrostatic contacts. Approximately 160 constraints from the distance geometry structures were added, to ensure that only minimal changes would occur upon energy minimization. Low conformational energies (≈−540 kcal/mol) were obtained with only small changes in structure; energies were within 3% of each other for all structures. No attempt was made to further improve the conformational energy by molecular dynamics or other methods, because it would likely lead to major structural changes. Energy minimization was used only to search for low-energy members of the family of structures defined by the NMR-derived constraints using distance geometry.

The range of backbone torsion angles found in the 10 final structures is presented in Table V. Since it is not known how well the conformational space is sampled, these are upper limits to the precision with which this structure is defined. Although

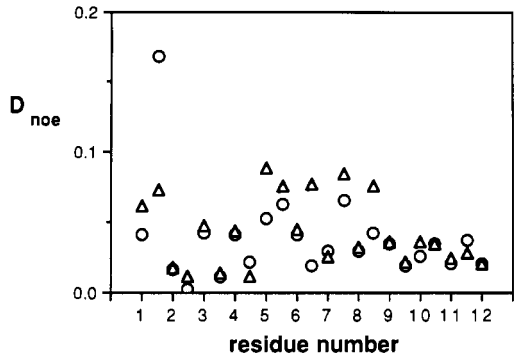

FIGURE 4: Locals rms deviations between observed and calculated NOE buildup rates for two representative structures. The rms deviations D_{NOE} for all intranucleotide NOE's for a given residue are shown above the residue number. Since most NOE's are sequential, rms deviations for NOE's between neighboring nucleotides are shown between the two corresponding residues.

most torsion angles changed slightly upon energy minimization, there was no significant change in the precision after energy minimization; the structural definition is not an artifact of energy minimization. In most cases, the torsion angles are distributed around a single average value. Between G8 and G9, however, two significantly different sets of values were found. The inability to stereospecifically assign H5′ and H5″ makes both g^+ and g^- conformers of β of G9 acceptable, thereby creating two distinct families of structures. This observation emphasizes the importance of the torsion angle constraints. Table V also reveals the role of the NOE's in defining poorly constrained backbone angles. The range of values for α and ζ is much smaller than that defined by the constraints. In the stem, all ϵ angles are trans although the g^- conformer was also allowed. Finally, α of U6 is always trans, as required by the strong NOE's between U5 and U6 sugar protons.

The rms deviation D between observed and calculated NMR observables provides an estimate of how well the original data are accounted for in the final structure. For all structures, D is between 0.75 and 0.80 after distance geometry; after energy minimization, D decreases to 0.65–0.75. This decrease is due in part to changes in the sugar conformation; the pucker amplitude ϕ_m is often larger than expected from the scalar couplings after distance geometry. The agreement between the data and the structure is very good if judged from the value of the rms deviation D_{NOE} between observed and calculated NOE's; $D_{NOE} \approx 0.05$ in all final structures. The conservative interpretation of NOE buildup rates into loose distance geometry constraints (corresponding to strong, medium, and weak NOE's) could have led to larger deviations. However, the small rms deviation indicates that the final structures are consistent with a tighter set of distance constraints, closer to those directly evaluated from the buildup rates. Scalar couplings are fitted more poorly, $D_J \approx 1.05$–1.2, probably because distance geometry is not designed to use torsion angle constraints. For some angles, large torsion angle changes correspond to only small changes in interatomic distances and distance geometry penalty. The values of the rms deviations suggest that all significant improvements in the quality of the structure will come from a better agreement between experimental and calculated scalar couplings.

Local rms deviations for NOE's provide a measure of how the deviation is distributed among different parts of the molecule. Intranucleotide NOE's and NOE's between neighboring nucleotides are plotted in Figure 4 versus the residue number. The distribution of rms deviations is fairly flat, showing that the data are fitted equally well for different nucleotides, with the exception of the 5′ end of the stem. This is due to fewer constraints and conformational flexibility. For one of the two structures shown in Figure 4, larger than average rms deviations are found in the loop. The poor local agreement between observed and calculated internucleotide NOE's suggests that one of two conformations for the glycosidic angle of C7 found by distance geometry is inconsistent with the internucleotide NOE's. The overall folding of the hairpin was not different for the two structures, but the poor local fit of NOE constraints led to an almost 2-fold increase in the distance geometry penalty.

Hairpin Structure

Stem. The precision with which oligonucleotide structures can be defined by NMR has been the subject of considerable debate. In previously published studies, little experimental information on the backbone was available. In the present work, the precision of nucleic acid structure determination was improved by defining the backbone conformation. This can be best appreciated in the stem, where the helical conformation is similar to that of previously investigated structures. Inspection of Table V reveals that the backbone is well-defined. The angle δ is very precisely defined; many interproton couplings and NOE's define the sugar conformation (and hence δ). The angles β, γ, and ϵ are also well-defined, although less precisely than δ. The angles next to the phosphorus, α and ζ, are more poorly defined because of the lack of precise experimental information. The values of the backbone torsion angles, as well as the precision with which the torsion angles are defined, depend on the sequence (Table V). The pyrimidine-rich side of the stem (U10–C12) is better defined than the purine-rich side. This is due in part to the presence of larger conformational flexibility on the 5′ side of the stem, but also to some extra NOE's provided by the H5 proton. Differences between dinucleotide steps are often larger than the uncertainty between different structures, particularly for the angles for which detailed information is available (β, γ, δ, ϵ). Nevertheless, the differences observed at individual dinucleotide steps are comparable to those observed before and after energy minimization, and might therefore be due to the AMBER force field. It is not yet clear from the present results whether the resolution is sufficient to distinguish sequence-specific structural variations in double-helical RNA that might be important in biological function.

It has also been a matter of debate whether NMR can distinguish A-form from B-form helices (Aboul-Ela et al., 1988; Fairall et al., 1989). The pattern of internucleotide NOE's is typical of A-form helices, and very different from that of B–form oligonucleotides. In order to reach a quantitative conclusion, helical parameters were calculated with the program MARLA (Babcock, 1990). In the stem, helical parameters are close to those found in fiber diffraction studies of A-form helices, and very different from those expected for B-form. The displacements, the most sensitive measure of A- and B-form stacking, are large and negative, very close to the values observed in A-form fiber diffraction studies ($\approx$–4 Å); in the B form, the displacement is close to 0. The twist angles are $\approx$33°, and the base pair inclination are between 8° and 10°. Both sugar puckers and glycosidic angles are A-like, and the minor groove width is $\approx$10 Å. Thus, the double-helical stem is A form, and NMR can distinguish A- from B-form double helices in solution.

Loop. Several unusual interactions are found in the loop; they presumably contribute to the unusual thermodynamic stability of the hairpin. The precision with which the structure

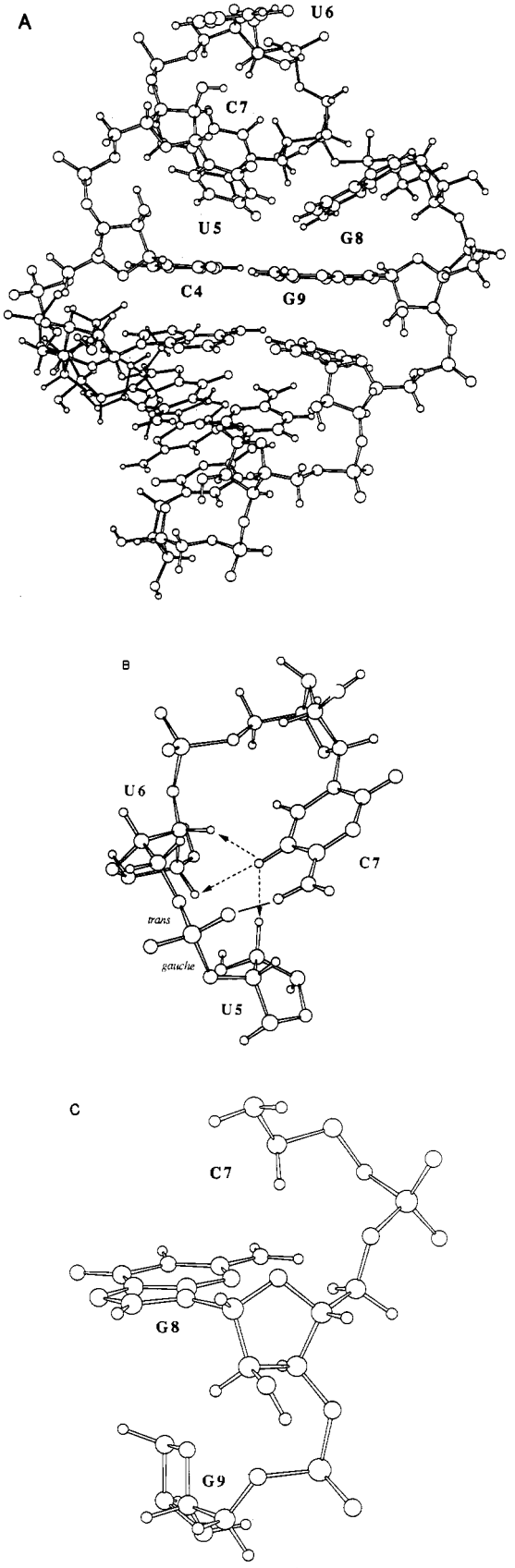

FIGURE 5: View of the hairpin; 1 of the 10 structures generated by distance geometry and restrained energy minimization was chosen for presentation. (A) Overall view from the major groove side. (B) Detail of the loop nucleotides with the C7 amino to phosphate oxygen contact; base protons of U5 and U6 are not shown for clarity. Strong NOE's from the C7 H5 proton to U5 and U6 sugar protons observed in NOESY experiments are indicated with arrows. (C) Unusual stacking around the *syn*-guanosine G8; base protons of C7 and G9 are not shown for clarity.

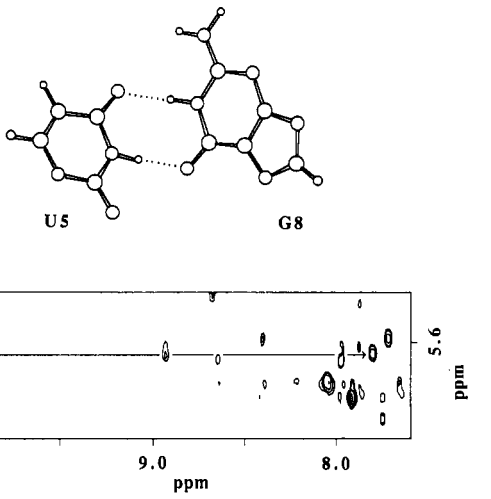

FIGURE 6: NOESY spectrum in water (90:10 H_2O/D_2O) at 250-ms mixing time for the UUCG hairpin in 10 mM sodium phosphate/0.01 mM EDTA (pH 5.2) at 0 °C. The spectrum was acquired by using the jump-and-return sequence for solvent suppression on the last pulse. Only the region corresponding to cross-peaks between amino and aromatic protons (plus the G8 imino proton at 9.96 ppm) and H5 and sugar H1′ protons is shown. The G8 imino to U5 H5 and the intranucleotide U5 H5–H6 cross-peaks are joined by an arrow. The G8 imino to U5 H5 NOE is a good marker for the reverse-wobble base pair shown in the upper part of the figure.

is defined enables more confident conclusions about the loop.

(1) Stacking is continued into the loop from both strands of the stem. As observed in other RNA loops (Puglisi et al., 1990; Varani et al., 1989), the first loop nucleotide (U5) stacks on the 5′ strand of the stem, conserving A-like geometry. This hairpin also shows stacking on the other side of the loop, despite the syn G8, as revealed by NOE's between G8 and G9 protons. In order to accommodate the syn base, the backbone between G8 and G9 undergoes some rearrangement. β of G9 is either g^+ or g^-, as shown by the large couplings between the phosphorus and either H5′ or H5″. Two alternate structures are found by distance geometry: ϵ, ζ, and α are either all gauche when β is g^+, or both ϵ and α are trans when β is g^-, ^{13}C–^{31}P couplings would resolve the ambiguity by constraining ϵ. The rest of the structure and its overall folding are not affected by these two possibilities.

(2) Both U5 and G8 imino protons have NOE's to amino and imino protons of the C4·G9 base pair (Figure 1). This supports a G·U base pair; both imino protons would be close to the exchangeable protons of the C·G base pair at the top of the stem. The NOE from the G8 imino to U5 H5 strongly supports a reverse-wobble base pair with syn G (Figure 6); no NOE was observed from U5 imino to G8 H8, showing that the guanine N7 is not close to the U5 imino. However, the G imino is not downfield-shifted, and the NOE between U5 and G8 imino protons is not as strong as expected (Figure 1). Nevertheless, it is not known what chemical shifts to expect for a reverse-wobble pair, and the high rate of exchange with the solvent of the U5 imino proton could be responsible for the weak U5 imino to G8 imino NOE. A weak NOE between G and U imino protons of a G·U pair has been reported previously (Puglisi et al., 1990). The evidence is inconclusive as to whether the G·U base pair has one or two hydrogen bonds, although the observation of the G8 imino to U5 H5 NOE strongly supports the reverse-wobble pair with two hydrogen bonds.

Only the U imino to G carbonyl hydrogen bond was introduced as a constraint in distance geometry. Some of the structures showed a G·U base pair with a single hydrogen bond

and a close G amino to phosphate oxygen contact instead of a second hydrogen bond. After energy minimization, a reverse-wobble base pair with two hydrogen bonds is always found, but a close G amino to phosphate contact remains (Figure 5C). This contact could explain the decrease in stability observed when the G is mutated to inosine (Sakata et al., 1990). Inspection of the helical parameters shows that the G·U base pair is considerably buckled (50°) (Figure 5A) and underwound (twist is 15°). The buckle helps bridge the two sides of the stem by bringing them closer (Blommers et al., 1989).

(3) This hairpin has a loop of only two nucleotides; both U6 and C7 have 2′-endo sugar pucker. This extends the backbone by about 2 Å (Saenger, 1984) and helps bridge the gap between opposite strands of the stem. A 2′-endo sugar pucker has been observed in our laboratory for other RNA hairpins with small loops (Puglisi et al., 1990; P. Davis, unpublished results). We think this is very common in RNA hairpins with small loops.

(4) The U5pU6 phosphorus resonates downfield from the stem resonances. This suggests a conformational change in the phosphodiester backbone. Very unusual NOE's (U5 H2′ to U6 H2′ and H3′, U5 H4′ to U6 H5) are consistent only with α being trans, as found in all converged structures despite the absence of any specific constraint. The backbone conformation is in this case well-defined by the NOE's. The unusual g⁻t conformation for $\zeta\alpha$ turns the backbone sharply toward the opposite strand of the stem, thereby helping to bridge the gap with only two nucleotides.

(5) NOE's from C7 aromatic protons to U5 and U6 sugar protons reveal unusual interactions between C7 and the two preceding nucleotides (Figure 5B). Strong NOE's from C7 H5 to both U5 and U6 H2′ and H3′ protons require the C7 amino group to be in close contact with the U5pU6 phosphate. Upon energy minimization, a hydrogen bond is formed between a C7 amino proton and a phosphate oxygen. Since the rate of rotation for the C7 amino group is too fast to detect two separate resonances for the amino protons, there is no direct experimental support for the hydrogen bond, but the close amino–phosphate contact is required by the NOE's. The favorable amino–phosphate contact stabilizes the turn in the backbone. A similar motif was observed in tRNA, where a U imino to phosphate hydrogen bond stabilizes the π turn in the anticodon loop (Saenger, 1984).

(6) The backbone between C7 and G8 is not well-defined because of the few NOE's observed, and because γ of G8 is not constrained. Despite the apparent precision of Table V, one must be aware of possible insufficient sampling of the conformational space by distance geometry. However, very unusual chemical shifts support the proposed structure. H4′, H5′, and H5″ protons from C7, and H1′ of G9, are shifted upfield by as much as 2 ppm, whereas H3′ of G8 is about 1 ppm downfield from other H3′ protons (Table II). For a syn base and 3′-endo sugar pucker, the H3′ proton is in the plane of the base and deshielded by the ring current. In the three-dimensional structure, the C7 base stacks on U5, while the H4′, H5′, and H5″ protons stack directly above the purine ring; the H1′ proton of G9 is underneath the G8 base (Figure 5C). These anomalous chemical shifts can be attributed to the ring current of the syn base.

UUUG Hairpin. Comparison of the thermodynamics of hairpin formation shows that the UUUG hairpin is considerably less stable than the UUCG hairpin ($\Delta\Delta G°$ = +1.5 kcal/mol at 25 °C). Imino resonances from the UUUG hairpin melt at lower temperature than those from the UUCG hairpin, probably because of the lower thermodynamic stability. However, the proton and phosphorus chemical shifts between the two hairpins are very similar (Tables II and III), and scalar couplings are the same within experimental error. The observed NOE's are also the same at the level of precision used in the structure determination (cross-peaks classified as strong, medium, and weak). Thus, the structures of the two hairpins are very similar. Examination of the structure of the UUCG hairpin suggests two explanations for the reduced stability of the mutant. First, differences in stacking between C and U should diminish the UUUG hairpin stability. Second, the favorable C7 amino–phosphate oxygen contact (Figure 5B) is lost when a carbonyl of U7 substitutes for the amino of C7. The energetics of these effects—changes in stacking and loss of a hydrogen bond—are of the same magnitude as the free energy difference at 25 °C (1.5 kcal/mol) between the two hairpins.

Conclusions

The structure of a very common and unusually stable RNA hairpin (Tuerk et al., 1988) has been determined in solution by NMR and distance geometry. Nearly complete assignments have been achieved by correlated experiments and NOESY. Scalar coupling constants have been used to define the backbone conformation and the sugar pucker; NOE's provided estimates of interproton distances. More than 400 constraints—interproton distances from NOESY experiments and torsion angles from scalar couplings—were introduced into distance geometry; the ratio between experimental constraints and degrees of freedom was ≈5:1. The information from scalar couplings enabled us to generate well-defined structures despite the conservative use of interproton distances (Wüthrich, 1986). Restrained energy minimization (AMBER, Kollman et al., 1980) was done on the structures defined by distance geometry to improve van der Waals and electrostatic contacts. Constraints from the distance geometry structures prevented major changes upon minimization, while still allowing very low conformational energy. The structure was uncertain only where the lack of stereospecific assignments of H5′ and H5″ prevented the choice between two different backbone conformations.

Several specific interactions in the loop are expected to contribute to the hairpin thermodynamic stability. All bases but U6 stack on other bases, and the C7 phosphate and sugar stack on the G8 base. A G·U base pair with a *syn*-guanosine closes the stem, leaving only two nucleotides to close the gap between opposite strands of the stem. A specific base–phosphate contact, probably a hydrogen bond between the C7 amino and a U5pU6 phosphate oxygen, stabilizes a sharp turn in the phosphodiester backbone. The turn and the 2′-endo puckering of the two loop nucleotides enable the bridging of the stem with only two nucleotides. The amino–phosphate contact, and differences in stacking between U and C, can explain why a mutant with C7 changed into U7 is less stable than the UUCG hairpin, despite the similar structure. The proposed structure agrees with the results of chemical mapping of hairpins with identical sequence in 16S rRNA (Moazed et al., 1986; Cheong et al., 1990). This compact structure can explain the poor reactivity of this hairpin to single-strand-specific nucleases, and the inability of reverse transcriptase to read through this loop. Only U6 lacks specific contacts with other nucleotides; this nucleotide, not others, is often mutated in ribosomal RNAs (Woese et al., 1990).

These results show that the structure of nucleic acids can be determined in solution by NMR, if the backbone is defined by means of scalar coupling measurements. The added experimental information improves the precision with which the

structure is known and helps ensure that the folding is correct.

ACKNOWLEDGMENTS

It is a pleasure to thank Mr. David Koh for synthesizing DNA templates, Ms. Barbara Dengler for managing the laboratory, Mr. Bill Ross for help with AMBER, and Mr. Brian Wimberly for helpful discussions. Prof. W. Olson and Dr. M. Babcock, Rutgers University, kindly analyzed the coordinates of our structures in terms of helical parameters.

REFERENCES

Aboul-Ela, F., Varani, G., Walker, G. T., & Tinoco, I., Jr. (1988) *Nucleic Acids Res. 16*, 3559–3572.
Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas 101*, 413–433.
Babcock, M. (1990) Ph.D. Thesis, Rutgers University.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
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*, 7491–7498.
Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.
Cheong, C., Varani, G., & Tinoco, I., Jr. (1990) *Nature 346*, 680–682.
De Leeuw, F. A. A. M., & Altona, C. (1982) *J. Chem. Soc., Perkins Trans. 2 2*, 375–384.
De Leeuw, F. A. A. M., Haasnoot, C. A. G., & Altona, C. (1980) *Isr. J. Chem. 20*, 108–126.
Fairall, L., Martin, S., & Rhodes, D. (1989) *EMBO J. 8*, 1809–1817.
Frey, M. H., Leupin, W., Sorensen, O. W., Denny, W. A., Ernst, R. R., & Wüthrich, K. (1985) *Biopolymers 24*, 2371–2380.
Giessner-Prettre, C., Pullman, B., Ribas Prado, F., Cheng, D. M., Iuorno, V., & Ts'o, P. O. P. (1984) *Biopolymers 23*, 377–388.
Gorenstein, D. G. (1984) *Phosphorous-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) Academic Press, New York.
Groebe, D. R., & Uhlenbeck, O. C. (1988) *Nucleic Acids Res. 16*, 11725–11735.
Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.
Heaphy, S., Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Karn, J., Lowe, A. D., Singh, M., & Skinner, M. A. (1990) *Cell 60*, 685–693.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.
Kollman, P., Weiner, P., & Dearing, A. (1981) *Biopolymers 20*, 2583–2611.
Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C., & Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387–1405.
Lazinski, D., Grzadienska, E., & Das, A. (1989) *Cell 59*, 207–218.
Macura, S., & Ernst, R. R. (1979) *Mol. Phys. 41*, 95–117.
Macura, S., Wüthrich, K., & Ernst, R. R. (1982) *J. Magn. Reson. 46*, 269–282.
Malin, M. H., Tiley, C. S., Mc Carn, D. F., Rusche, J. R., Hauber, J., & Cullen, B. R. (1990) *Cell 60*, 675–683.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.
Moazed, D., Stern, S., & Noller, H. F. (1986) *J. Mol. Biol. 187*, 399–416.
Müller, N., Ernst, R. R., & Wüthrich, K. (1986) *J. Am. Chem. Soc. 108*, 6482–6492.
Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119–162.
Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310–7311.
Puglisi, J. D., & Tinoco, I., Jr. (1989) *Methods Enzymol. 180*, 304–325.
Puglisi, J. D., Wyatt, J. R., & Tinoco, I., Jr. (1990) *Biochemistry 29*, 4215–4226.
Rinkel, L. J., & Altona, C. (1986) *J. Biomol. Struct. Dyn. 4*, 621–649.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Sakata, T., Hiroaki, H., Oda, Y., Tanaka, T., Ikehara, M., & Uesugi, S. (1990) *Nucleic Acids Res. 18*, 3831–3839.
Sklenar, V., Miyoshiro, H., Zon, G., & Bax, A. (1986) *FEBS Lett. 208*, 94–98.
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*, 1364–1368.
Varani, G., Wimberly, B., & Tinoco, I., Jr. (1989) *Biochemistry 28*, 7760–7772.
Williamson, J. R., & Boxer, S. G. (1989) *Biochemistry 28*, 2819–2831, 2831–2836.
Witherell, G. W., & Uhlenbeck, O. C. (1989) *Biochemistry 28*, 71–76.
Woese, C. C., Winker, S., & Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467–8471.
Wu, H. N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221–8227.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.