NMR Structure of a Bacteriophage T4 RNA Hairpin Involved in Translational Repression†

Suman R. Mirmira and Ignacio Tinoco, Jr.*

Department of Chemistry, University of California, and Structural Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

Received February 21, 1996; Revised Manuscript Received April 8, 1996⊗

ABSTRACT: A high-resolution structure of a 16-nucleotide bacteriophage T4 RNA hairpin, 5′-GCCU-[AAUAACUC]GGGC (loop bases in square brackets), has been determined in solution by proton, phosphorus, and carbon (natural abundance) NMR spectroscopy. This RNA hairpin is known to play a crucial role in the translational repression of bacteriophage T4 DNA polymerase. Ultraviolet absorbance melting curves indicate that the structure formed is unimolecular. The NMR spectra indicate that a single conformation consistent with a hairpin structure is formed. Strong imino−imino NOEs confirm the formation of the G·U base pair at the stem−loop junction. There is no evidence that A_5 is protonated (at pH 6.0) and involved in an A^+·C pair. However, the NMR data indicate that the stem is extended beyond the G·U pair and that A-form stacking continues for three nucleotides on the 5′ side and one nucleotide on the 3′ side. Structure calculations using restraints obtained from NMR data give a precisely defined structure with an average root mean square deviation (RMSD) of approximately 1.2 Å for the entire molecule. The assignment of all the protons and most of the ^{31}P resonances in the loop yielded a large number of distance and torsion angle restraints for these nucleotides. These helped obtain a well-defined loop with an average RMSD of 1.1 Å for the loop nucleotides of 11 converged structures.

In bacteriophage T4, the protein product of gene *43* (gp43) is a DNA polymerase that is essential for replication of the phage genome (de Waard et al., 1965; Warner & Barnes, 1966). In addition to its enzymatic functions in DNA replication, T4 gp43 is known to regulate its own synthesis *in vivo* (Russel, 1973). This regulation occurs at the translation stage. *In vitro* translation of an N-terminal fragment of gp43 is specifically repressed by the addition of purified protein. The protein binds gene *43* mRNA at a site that overlaps the Shine−Dalgarno sequence, thereby preventing translational initiation (Andrake et al., 1988). Nitrocellulose binding assays in which purified gp43 was added to gel-purified RNA fragments were used to quantify the strength of the polymerase−mRNA interaction. Transcripts which extended from 38 to 64 nucleotides upstream of the AUG initiator codon had similar binding profiles, and the *in vitro* dissociation constant (K_d) for the gp43−RNA interaction was found to be about 1 nM (Tuerk et al., 1990). The minimum RNA sequence to which the DNA polymerase binds consists of 36 nucleotides. As shown in Figure 1a, this operator sequence includes a hairpin containing a five-base-pair stem and an eight-nucleotide loop.

This hairpin and, in particular, the loop sequence were found to be crucial for polymerase binding (Tuerk et al., 1990). Mutations that result in the disruption of the stem lead to a dramatic reduction in its affinity for the polymerase. The distantly related bacteriophage RB69 also has an RNA hairpin at the ribosome binding site of its DNA polymerase gene. Significantly, the loop sequence of the RB69 hairpin is identical to the one seen in T4. Randomization of the loop nucleotides in T4 results in a 50-fold decrease in polymerase binding, demonstrating a loop nucleotide bias in this operator region. These results demonstrate the importance of the hairpin as well as of this particular loop sequence.

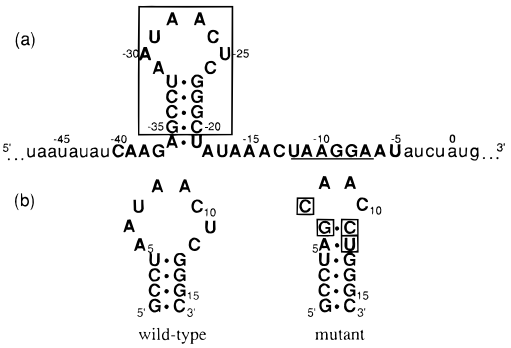

FIGURE 1: (a) The translational operator for gene *43* in bacteriophage T4. The minimum RNA sequence recognized by gp43 is shown in capitals. The Shine−Dalgarno sequence is underlined. The nucleotides are numbered from the A = 0 of the AUG initiation codon. For purposes of our NMR study, only the boxed region was synthesized. The first base pair (A·U) was not synthesized. Instead, the first nucleotide in the transcript was G(−35). This was done in order to get efficient transcription with T7 RNA polymerase. (b) Proposed secondary structures of the wild-type and mutant hairpin. The positions at which the sequence of the two hairpins differ have been boxed.

Selection experiments in which the loop nucleotides of the T4 hairpin were randomized and selected on the basis of affinity for polymerase showed that two hairpins bind the polymerase with equal affinity (K_d = 4.8 nM) (Tuerk & Gold, 1990). One was the wild-type hairpin and the other

† This research was supported in part by grants from the National Institutes of Health (GM 10840) and the Department of Energy (DE-FG03-86ER60406), by instrumentation grants from the Department of Energy (DE-FG05-86ER75281), and the National Science Foundation (DMB 86-09305 and BBS 87-20134).

* To whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1996.

S0006-2960(96)00414-X CCC: $12.00

differed from the wild type at four positions. As shown in Figure 1b, the two selected hairpin sequences have different proposed secondary structures—differing in stem length and loop size. However, what is interesting is that both these hairpins bind the polymerase with equal affinity. In order to determine whether their solution structures are similar, and to determine what structure the hairpins present to the polymerase, we have studied the three-dimensional structures of both hairpins using nuclear magnetic resonance (NMR).[1]

In this paper, we present the results of our structural studies on the wild-type hairpin.

As shown in Figure 1a, the first base pair (A·U) in the stem of the hairpin was not synthesized. Instead, in order to get efficient transcription with T7 RNA polymerase, the first nucleotide in the transcript was G(−35). Proton, phosphorus, and carbon (natural abundance) NMR spectroscopy was used to characterize the structure of the 16-nucleotide hairpin. Restrained molecular dynamics calculations were used to determine structures consistent with the NMR data.

MATERIALS AND METHODS

RNA Synthesis and Purification. The RNA oligonucleotide, 5′-GCCU[AAUAACUC]GGGC-3′, was synthesized *in vitro* using T7 RNA polymerase from a synthetic single-stranded DNA template with a double-stranded promoter region (Milligan et al., 1987; Wyatt et al., 1991). Crude RNA was purified using denaturing 20% polyacrylamide gel electrophoresis. The yield of RNA was approximately 7−9 nmol of purified RNA/mL of reaction. The sequence of the RNA oligonucleotide was verified by partial digestion with base-specific RNases A, T1, U2, and CL3.

Thermodynamics. Ultraviolet absorbance melting curves were obtained and analyzed to get thermodynamic parameters (Puglisi & Tinoco, 1989). The data were recorded at 260 nm on a Gilford Model 250 UV−vis spectrophotometer with a thermoelectric cell holder. The heating rate during the experiment was 1 °C/min and the dwell time was 4 s. The molecularity of the system was determined from UV absorbance versus temperature profiles recorded over an approximately 80-fold concentration range (7.6−620 μM) of the RNA sample. The pH dependence of the melting temperature was determined by UV absorbance melting curves performed with RNA samples dissolved in buffers at pH 5.0, 6.0, and 7.0. All melting samples were dialyzed against water, lyophilized, and then dissolved in buffer. They were preheated to about 80 °C for 2 min, cooled rapidly in ice, and then degassed in a Savant Model SVC100H speed vac for 1 min. The samples used to measure UV melting curves as a function of RNA concentration were dissolved in 10 mM sodium phosphate and 10 μM EDTA, pH 6.0. The samples used for the pH dependence study were dissolved in 10 mM sodium phosphate and 10 μM EDTA buffers at pH 5.0, 6.0, and 7.0 (concentration of the RNA at all three pH values was similar and approximately 8 μM).

NMR Spectroscopy. RNA samples used to measure NMR spectra were dialyzed extensively (>72 h) against a sodium phosphate and EDTA buffer and finally dissolved in 10 mM sodium phosphate and 10 μM EDTA, pH 6.0. Samples used for the pH-dependent HMQC experiments were dialyzed extensively in buffer at the required pH and finally dissolved in 10 mM sodium phosphate and 10 μM EDTA, pH 4.5, 6.0, or 7.0. Samples used to measure exchangeable spectra were lyophilized and then suspended in 400 μL of 90% H_2O/10% D_2O, while those used to measure nonexchangeable spectra were lyophilized several times from 99.96% D_2O (Aldrich) and then suspended in 400 μL of 99.996% D_2O (Cambridge Laboratories). The RNA concentration was approximately 2 mM. The samples were annealed at 65 °C for 2 min and cooled in ice for at least 5 min prior to each experiment. All proton and carbon chemical shifts were referenced to TSP. Most NMR spectra, with the exception of the $^1H-^{31}P$ HETCOR (heteronuclear correlation spectroscopy), were recorded on either a GE GN-500 or a BRUKER AMX-600 spectrometer. The $^1H-^{31}P$ HETCOR was recorded on a BRUKER AMX-300. All data processing was done with FELIX v. 2.1 or 2.3 (Biosym Technologies, Inc.).

One-dimensional (1D) exchangeable proton spectra were recorded at different temperatures, from 7 to 40 °C, on the GN-500 using the 1-1 water suppression sequence (Plateau & Gueron, 1982). The maximum excitation was centered on the imino region. A total of 4096 real points was collected over a spectral width of 10 000 Hz. 1D nonexchangeable spectra were also recorded at different temperatures (on the GN-500) to determine the optimum temperature at which to record all other spectra. T_1 measurements using a nonselective 180° composite pulse and a relaxation delay of 20 s were recorded at 15 and 25 °C.

Exchangeable two-dimensional (2D) NOE experiments, i.e., NOESY spectra in H_2O, were recorded at 2 °C using the 1-1 water suppression pulse sequence (Plateau & Gueron, 1982) with a 65 μs delay between the two pulses. The water suppression was optimized empirically by adjusting the length of the second pulse to be slightly (≈0.1 μs) shorter than the first. A total of 80 scans was taken for each FID with a 2.0 s relaxation delay between scans. Approximately 360 FIDs of 2048 (2K) real points were collected. The mixing time used was 300 ms. The spectral width in both, the t_1 and t_2 dimensions, was 12 500 Hz. The acquired dimension was convoluted to reduce the intensity of the water peak before further processing. The spectra were zero-filled to 2K real points in the t_1 dimension and apodized in both dimensions with a skewed (skew factor = 0.7) sine bell function.

All 2D spectra were recorded in the phase-sensitive mode using TPPI (Marion & Wüthrich, 1983). All 2D nonexchangeable spectra were recorded at 25 °C. For these spectra, the HDO peak was attenuated by low-power presaturation during the relaxation delay. Typically, 1024 (1K) real points in t_2 and 400−450 points in t_1 (400−450 FIDs) were collected with spectral widths of 4000 or 5000 Hz for experiments collected at 500 or 600 MHz, respectively. Total acquisition times varied from 19 to 23 h.

NOESY spectra in D_2O with 100 and 400 ms mixing times were recorded on either the GN-500 (100 ms) or the AMX-600 (400 ms). These were used mainly for assignment purposes. Experiments with mixing times of 50, 100, and

[1] Abbreviations: UV, ultraviolet; EDTA, ethylenediaminetetraacetic acid; NMR, nuclear magnetic resonance; FID, free induction decay; 1D, one dimensional; 2D, two dimensional; ppm, parts per million; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlation spectroscopy; HMQC, heteronuclear multiple-quantum correlation spectroscopy; TOCSY, total correlation spectroscopy; rMD, restrained molecular dynamics; RMSD, root mean square deviation; TSP, 3-(trimethylsilyl)-1-propanesulfonate; TMP, trimethyl phosphate.

150 ms were collected sequentially on the AMX-600 and were used to estimate the interproton distances. A relaxation delay of 2.5 s was used for all the experiments. A total of 64 scans was collected for each FID, and 400−450 FIDs of 1K real points were collected for each experiment. Zero-quantum artifacts were shifted by incrementing the mixing time with t_1 (Macura et al., 1982). The spectra were zero-filled to 1K real points in t_1. They were apodized in both dimensions using a skewed sine bell function with a skew factor of 0.7 and a phase shift of 30° in the t_2 and 60° in the t_1 dimensions. The sequentially collected NOESY spectra with 50, 100, and 150 ms mixing times were apodized with a 60° phase-shifted sine bell function.

A high-resolution, phosphorus-decoupled, double-quantum-filtered COSY (DQF-COSY) spectrum was recorded at 600 MHz with a spectral width of 2000 Hz in both dimensions, 2K real points in t_2, and 750 points (FIDs) of 32 scans each in t_1. Phosphorus was decoupled using the GARP1 decoupling sequence (Shaka et al., 1985) during acquisition. The spectrum was zero-filled to 2K real points in t_2 and was apodized using a sine bell function (0° shift). The final resolution in t_2 was slightly better than 1 Hz/point. This spectrum was used to measure the proton−proton coupling constants.

A CLEAN-TOCSY (Griesinger et al., 1988) with compensated MLEV-17 sequence (Bax et al., 1985) for broadband decoupling was recorded on the AMX-600. A total of 512 FIDs was recorded with 48 scans and 1K real points for each FID. The mixing time was ≈75 ms and the relaxation delay was 2.5 s. The spectrum was zero-filled to 1K real points in t_1 and was apodized using a skewed sine bell function with a skew factor of 0.7 and a phase shift of 30° in the t_2 and 60° in the t_1 dimensions.

Natural abundance ^{1}H−^{13}C HMQC (Bax et al., 1983; Varani & Tinoco, 1991a) spectra were recorded at 600 MHz. The spectra were ^{13}C-decoupled using the GARP1 decoupling sequence (Shaka et al., 1985) during acquisition. The spectra were recorded on the AMX-600 and a relaxation delay of 2.5 s was used for all experiments. A total of 160 scans was collected for each FID, and 150−165 FIDs of 2K real points were collected for each experiment. The spectral widths were 5000 Hz (8.3 ppm) in t_2 (^{1}H) and 7550 Hz (50 ppm) in t_1 (^{13}C). The ^{13}C dimension was centered at 110 ppm, slightly downfield of the C5 resonances. The spectra were zero-filled to 1K real points in t_1 and apodized in both dimensions using a sine bell function with a 90° phase shift. In RNA, the carbons resonate over a frequency range of ≈100 ppm, from ≈60 ppm for the C5′s to ≈160 ppm for the adenine C2s. Our choice of carrier frequency and spectral width led to the folding in t_1 (the carbon dimension) of the C2′, C3′, C4′, and C5′ resonances as well as the aromatic (C6, C8, and C2) resonances. This gave us good resolution in the t_1 dimension with short acquisition times. The average acquisition time for these experiments was 21 h.

Proton-detected ^{1}H−^{31}P heteronuclear correlation spectroscopy (HETCOR) was recorded as previously proposed (Sklenar et al., 1986) on the AMX-300 spectrometer with an inverse probe. The spectral widths were 1200 Hz (4 ppm) in the proton dimension (t_2) and 300 Hz (≈2.5 ppm) in the phosphorus dimension (t_1). A total of 116 FIDs of 1K real points and 128 scans each were recorded. The spectrum was zero-filled to 1K real points in t_1 and apodized in both dimensions with a sine bell function phase shifted by 10°. The phosphorus frequency was referenced to an external standard of aqueous trimethyl phosphate (TMP). The resolution in t_2 was ≈1.2 Hz/point. This spectrum was used to measure the ^{1}H−^{31}P coupling constants.

Interproton Distances and Scalar Coupling Measurements. Interproton distances between nonexchangeable protons were derived from NOESY peak intensities at different mixing times (50, 100, and 150 ms). Cross-peak volume integration (of approximately 150 peaks) was done within FELIX (v. 2.3). The plot of cross-peak volume as a function of mixing time was linear for all eight pyrimidine H5−H6 cross peaks, indicating no spin diffusion even at 150 ms. The H5−H6 distance (2.41 Å) was used as a reference for determining the other distances. These distances were used to classify the cross peaks into four categories: strong (1.8−3.0 Å), medium (2.0−4.0 Å), weak (2.5−5.0 Å), and very weak (3.0−7.0 Å). Peaks that were not present in any of the above three NOESY spectra but were present in the 400 ms NOESY were classified "very weak". Exchangeable proton−proton NOEs were given looser constraints since relaxation could be due to chemical exchange with water as well as NOE cross relaxation. Inter-base-pair imino to imino cross peaks were assigned a distance range of 3.0−5.0 Å.

^{1}H−^{1}H scalar couplings were measured from a high-resolution DQF-COSY. For each pair of spins, the active and passive couplings measured in the t_2 dimension (≈1 Hz/point) were averaged. The uncertainty was estimated to be ±0.5 Hz for well-resolved cross peaks. Most of the H4′−H5′, H4′−H5″, and H1′−H2′ cross peaks were either very weak or absent. An upper limit of 2 Hz was estimated for weak or missing cross peaks. Heteronuclear ^{1}H−^{31}P coupling constants were measured from a HETCOR. The heteronuclear coupling, J_{HP}, was determined by subtracting the passive ^{1}H−^{1}H couplings from the total multiplet width. An upper limit of 3 Hz was estimated for weak or absent cross peaks. Torsion angles and the percentage of C3′-*endo* sugar conformer were determined as previously proposed (Altona, 1982; de Leeuw & Altona, 1982; Varani & Tinoco, 1991b).

Distance and Torsion Angle Constraints. In addition to 251 distance constraints derived from the NOESY data as described above, four distance constraints were used per Watson−Crick base pair (five for the G·U pair) which maintained hydrogen bonding between the bases and made the bases coplanar. These constraints were set to ±0.1 Å and were used only when imino resonances from the base pairs were seen.

Torsion angle constraints were used to keep the amino groups in the plane of the base. Nucleotides with weak or absent H1′−H2′ couplings were constrained as C3′-*endo* by constraining four of the endocyclic torsion angles in the sugar ring. Nucleotides with larger H1′−H2′ couplings were determined to be intermediate between C3′-*endo* and C2′-*endo*. Standard A-form values (Saenger, 1984) were used for the glycosidic torsion angle, χ, and the backbone torsion angles (α, about P−O5′; β, about O5′−C5′; γ, about C5′−C4′; ϵ, about C3′−O3′; and ζ, about O3′−P). These constraints were used only for regions of the hairpin where the H1′−H2′ couplings as well as the NOE connectivity patterns indicated A-form geometry, i.e., for the stem region as well as for the few nucleotides which stack in A-form geometry on the stem. Ranges used for the torsion angles

in the stem (C_2-U_4 and $G_{13}-C_{16}$) were ±50° for χ, ±20° for β, γ, and ϵ, and ±30° for α and ζ. For some loop nucleotides, sufficient information from scalar couplings was available to determine some of the backbone torsion angles. In these cases, and for G_1 (which is not as well defined by NMR data as the rest of the stem), backbone torsion angle constraints were used, but with broader ranges: ±40° for β and ϵ, ±30° for γ, and ±70° for α and ζ.

Structure Determination. The molecular dynamics program X-PLOR (Brünger, 1990) was used to generate three-dimensional structures consistent with the constraints derived from NMR data. The force field for the calculation included bond lengths, bond angles, improper angles (used to maintain chirality and base planarity), the repulsive term of van der Waals potential, and the distance and torsion angle constraints from the NMR data. The bond length and angle force constraints were set to 1000 kcal mol^{-1} $Å^{-2}$ and 500 kcal mol^{-1} rad^{-2}, respectively. The NOE force constant was set to 50 kcal mol^{-1} $Å^{-2}$, and the torsion angle force constant was varied from 5 and 50 kcal mol^{-1} rad^{-2} during the calculations. Structure calculations were done in two stages: the global fold using a modified simulated annealing protocol followed by a refinement protocol (Wimberly, 1992). All NOE-derived distance constraints were used at both stages of the calculation. However, only those torsion angle constraints needed to keep the amino groups in the plane of the base and maintain the correct sugar pucker were used for the global fold. Twenty starting structures with randomized α, β, γ, ϵ, and ζ torsion angles were generated within X-PLOR and then subjected to the global fold. This stage of the structure calculation began with 500 cycles of initial energy minimization, followed by restrained molecular dynamics (rMD) at 1000 K and subsequent rMD while cooling to 300 K. The calculation was concluded with 1000 cycles of final energy minimization. These structures were then checked for NOE violations as well as for close contacts between different regions of the molecule for which no NOEs were seen. Converged structures generated in the annealing protocol were then subjected to the refinement protocol. This stage of structure calculation involved 500 cycles of initial energy minimization, rMD at 1000 K with backbone torsion angle constraints introduced in two steps (first, χ, β, γ, ϵ, and then, α, ζ), rMD while cooling to 300 K, and 1000 cycles of subsequent energy minimization. The refined structures underwent further rMD at 300 K and finally 3000 cycles of energy minimization with Lennard-Jones potentials and electrostatic energies added. This final minimization step reduced the total energies of the refined structures but did not significantly alter the conformation of the molecule. The conformation of the loop was virtually unchanged by the final minimization step during which the electrostatic terms were added. The calculated structures were viewed using INSIGHT II (Biosym Technologies, Inc.).

RESULTS

Thermodynamics. UV absorbance melting curves of the wild-type RNA hairpin showed no change (within experimental error) in the melting temperature (T_m) over ≈80-fold concentration range, as expected for a unimolecular system. Fitting the melting curve data to a two-state, unimolecular model yielded $\Delta H° = -35 \pm 2$ kcal/mol, $\Delta S° = -109 \pm 6$ cal/(K mol), and $T_m = 48 \pm 1$ °C.

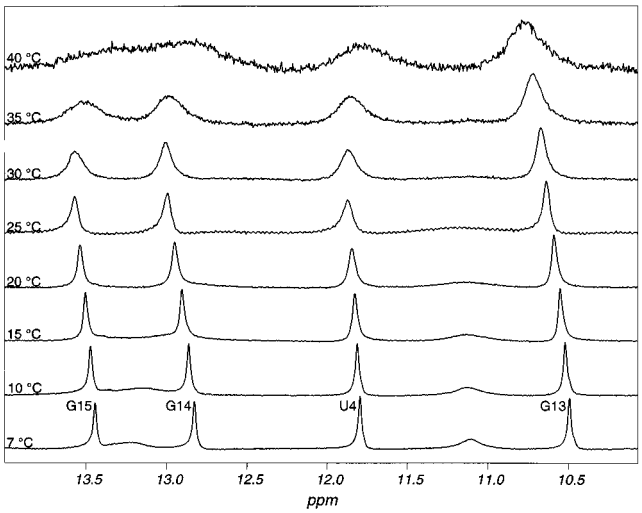

FIGURE 2: Temperature dependence of the imino spectrum of the 16-nucleotide wild-type T4 RNA hairpin in 90% H_2O/10% D_2O recorded from 7 to 40 °C on a GN-500 spectrometer using the 1-1 water suppression pulse sequence. All the spectra are referenced to the H_2O chemical shift at 7 °C and hence the impression that the imino peaks move downfield as the temperature increases.

To test for the formation of a protonated A^+·C pair (AH^+_5·C_{12}), the pH dependence of T_m was measured. Decrease in pH would stabilize an AH^+·C pair and lead to an increase in T_m. In the case of a 5′[AUUUCUGAC]3′ loop, a decrease in pH stabilized the AH^+·C pair at the end of the loop and allowed the formation of two more base pairs (an A·U and a G·U), thus leading to an increase in T_m (Puglisi et al., 1990). However, the T_m of a truncated TAR hairpin, 5′[CUGGGA]3′, did not change over a pH range of 4.5−7.5, implying that no AH^+·C pair forms (Jaeger & Tinoco, 1993). The wild-type T4 hairpin also showed no pH dependence of its T_m over a pH range of 5.0−7.0, indicating that the A_5H^+·C_{12} pair is not formed over this pH range.

NMR Analysis. (A) Assignment and Analysis of the Exchangeable Proton Spectra. The temperature dependence of the imino spectrum of the 16-nucleotide wild-type T4 RNA hairpin is shown in Figure 2. Four sharp imino resonances and two broad peaks are seen in the spectrum. The broad resonances at ≈13.2 and 11.1 ppm are the first to disappear as the temperature is increased. This disappearance, and their chemical shift positions, suggests that they belong to the G_1 and non-hydrogen-bonded loop imino protons, respectively. The sharp resonances seen at 11.79 and 10.47 ppm are characteristic of the imino protons of the U (U_4) and G (G_{13}), respectively, involved in a G·U pair. From the spectra at higher temperatures, it is seen that the G·U pair does not disappear or broaden until the other sharp imino resonances broaden. This indicates that the stem does not fray apart at the G·U pair. It is evident that there must be some bases which stack over this G·U pair in order to prevent its imino protons from rapid solvent exchange as the temperature increases. The remaining two resonances were assigned to G_{14} (12.78 ppm) and G_{15} (13.42 ppm). This accounts for all the imino resonances in the molecule. The presence of the U_4, G_{13}, G_{14}, and G_{15} imino resonances supports the formation of a hairpin structure with at least three base pairs. These assignments were confirmed from a 2D NOESY experiment of the hairpin in 90% H_2O/10% D_2O recorded at 2 °C. A strong NOE was seen between the imino protons of U_4 and G_{13}. Imino−imino cross peaks

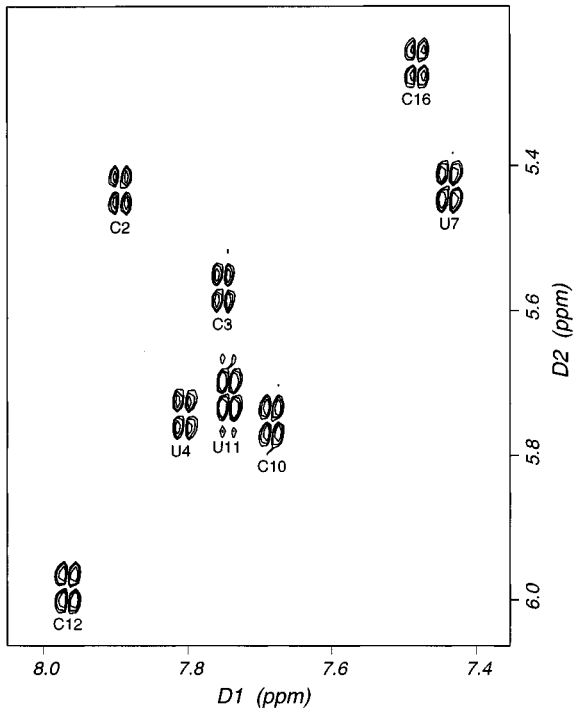

FIGURE 3: H6 (horizontal axis) to H5 (vertical axis) region of a DQF-COSY spectrum for the wild-type RNA hairpin in D_2O recorded at 25 °C on a GN-500 spectrometer. This spectrum was recorded witha spectral width of 4000 Hz. A total of 64 scans were collected for each of the 372 FIDs of 1K real points. Only eight peaks are seen. This is consistent with a single conformation for the hairpin.

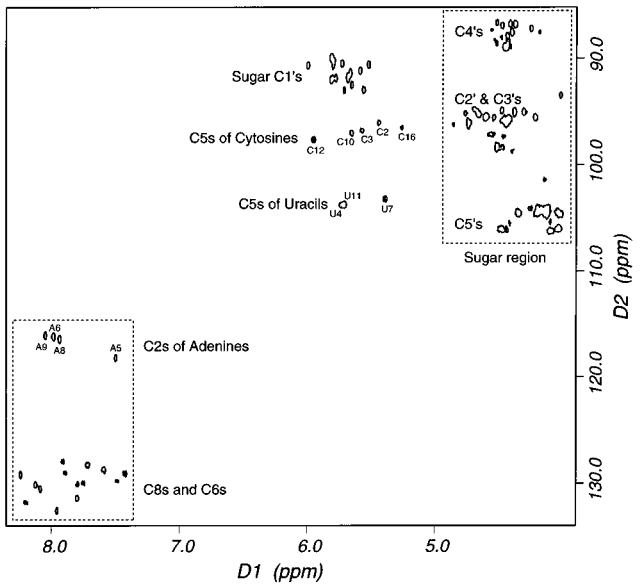

FIGURE 4: Natural abundance $^1H-^{13}C$ HMQC spectrum for the hairpin in D_2O. This spectrum was recorded at 25 °C on an AMX-600 spectrometer. The ^1H dimension is on the horizontal axis, and the ^{13}C dimension is on the vertical axis. The C2′, C3′, C4′, and C5′ resonances as well as the aromatic (C6, C8, and C2) resonances were folded in. Only the uracil and cytosine C5s and the sugar C1′s were not folded. The boxes drawn with dotted lines enclose the resonances that have folded in. The C8s and C6s resonate at ≈140 ppm, the C2s resonate at ≈153 ppm, and the sugar carbons (except C1′) resonate from ≈63 to ≈83 ppm.

seen between U_4 and G_{14}, G_{13} and G_{14}, and G_{14} and G_{15} confirmed the imino proton assignments. Assignments of the aminos in the stem were made from the same exchangeable NOESY experiment using the assignment pathways typically seen in an A-form stem (Heus & Pardi, 1991).

(B) Assignment and Analysis of the Nonexchangeable Spectra. Assignment of the nonexchangeable proton resonances proved to be crucial for the determination of the three-dimensional structure of the molecule. Assignment of ≈75% of the stem and all of the loop protons as well as most of the loop ^{31}P resonances enabled us to get a large number of distance and torsion angle constraints which were used to calculate the structure. The nonexchangeable proton resonances were assigned by standard methods (Varani & Tinoco, 1991b).

Figure 3 shows the H5−H6 region of a DQF-COSY spectrum of the wild-type hairpin acquired in D_2O. The number of H5−H6 cross peaks observed in the spectrum is consistent with a single conformation of the hairpin which contains eight pyrimidines. There is not much variation in the intensities of the cross peaks seen. This indicates the absence of flexibility or local dynamics in the different regions of the molecule.

The adenine H2 protons typically have larger T_1 values than other aromatic resonances (Wüthrich, 1986). Nonselective inversion of protons gives approximate values of T_1s (Derome, 1987) which are useful for AH2 assignment. Resonances at 7.54, 7.93, 7.98, and 8.05 ppm were found to have larger T_1 values than the other aromatic resonances. These resonances were assigned to the four AH2s in the molecule and were confirmed on the basis of the chemical shifts of their directly bonded carbon atoms (AC2s) using a natural abundance HMQC experiment. Adenine C2s resonate ≈10−15 ppm downfield from the C8 and C6 resonances. As seen in Figure 4, this allows for the clear separation of the adenine H2−C2 cross peaks from the H6−C6 or H8−C8 cross peaks. This spectrum also helps distinguish between the uracil H5s and cytosine H5s as the uracil C5s resonate ≈5 ppm downfield from the cytosine C5s. Information from the DQF-COSY, which assigned all the H5−H6 cross peaks, characteristic chemical shifts of the $^{5'}G_1$ and $^{3'}C_{16}$ nucleotides (Varani & Tinoco, 1991b), and the assignments of the adenine H2s, cytosine H6s, and uracil H6s were all used to simplify the aromatic to anomeric proton assignment for the wild-type hairpin. Figure 5 shows the H8/H6/H2−H5/H1′ region of a 400 ms NOESY spectrum. The H8/H6−H1′ connectivities are continuous from G_1 to U_7 and from A_8 to C_{16}. No cross peak was seen between U_7H1′ and A_8H8 in the 400 ms NOESY. Standard A-form connectivities were seen from G_1 to U_7 and from C_{12} to C_{16}.

It is important to have alternate assignment pathways to confirm assignments and resolve ambiguities in the H8/H6−H1′ connectivities. The H8/H6(i)−H2′(i)−H8/H6(i+1) connectivities can be very useful for both, confirming assignments as well as providing information about A-like stacking. In A-form RNA, the H8/H6(i)−H2′(i) NOE is weak, whereas the H2′(i)−H8/H6(i+1) peak is characteristically strong. For an A-form helix with C3′-*endo* sugar pucker, $J_{1'2'}$ is typically <2 Hz (Altona, 1982) (smaller than the line width), and hence H1′−H2′ are not often seen in the DQF-COSY spectrum. However, the short, sugar pucker-independent H1′−H2′ distance (2.5−2.8 Å) (Wüthrich, 1986) produces the strongest NOEs to the H1′ protons in a short mixing time NOESY. In the case of the wild-type hairpin, a few H2′ resonances could be assigned from the H1′−H2′ cross peaks of the DQF-COSY. A combination of this spectrum and

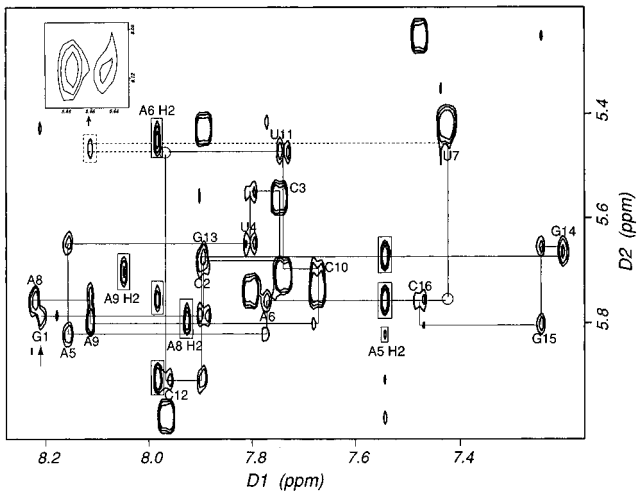

FIGURE 5: Fingerprint (aromatic to anomeric) region of a 400 ms NOESY acquired at 25 °C on an AMX-600 spectrometer. The H8/H6(i)−H1′(i)−H8/H6(i+1) walk is shown in solid lines. The nucleotides have been labeled near the cross peak that corresponds to the intranucleotide H8/H6−H1′ NOEs. The large, unlabeled peaks seen along the pyrimidine H6 resonances are the intranucleotide H5−H6 cross peaks. The arrow indicates the start of the walk at the G_1H8−G_1H1′ cross peak. The positions of cross peaks seen at lower contour levels than shown here are indicated with an open circle. Dotted lines mark the unusual, long-range connectivities (A_9-H8 to U_7H1′ and A_9H8 to U_{11}H1′). The cross peaks corresponding to these unusual connectivites are enclosed in a dotted box. In the spectrum shown, these peaks appear as one elongated peak. However, on the other side of the diagonal, (i.e., with the H1′ resonances along D1 and the aromatic resonances in D2), two peaks are clearly visible, as shown in the inset. The AH2−H1′ (both sequential and cross strand) have been boxed by solid lines.

the H1′−H2′ region of a 100 ms NOESY was used to assign all the H2′ resonances. It was then possible to establish independent assignment pathways using the H8/H6−H2′ connectivities. The H8/H6−H2′ connectivities were continuous from G_1 to C_{16} except for U_{11} and C_{12} where one cannot be certain as the H2′ resonances are overlapped. Standard A-like connectivities were seen from G_1 to U_7 and from C_{12} to C_{16}. In most cases, the H2′(i)−H1′(i+1) and H2′(i)−H5(i+1) connectivities provided further corroboration for the assignments.

Assignment of the remaining sugar protons is crucial for structural information. A high-resolution, ^{31}P-decoupled DQF-COSY was used to assign the different sugar spin systems. The quality of the spectrum enabled observation of resonances separated by as little as 0.04 ppm. The phosphorus decoupling simplifies the multiplet and reduces spectral overlap. The sugar spin systems were identified by the characteristic multiplet structures of the H2′−H3′, H3′−H4′, H4′−H5′/5″, and H5′−H5″ cross peaks (Varani & Tinoco, 1991b). A TOCSY spectrum showed cross peaks from the H1′ proton to the H2′, H3′, and H4′ protons for those nucleotides which had $J_{1'2'} > 2$ Hz. This confirmed the assignments obtained from the DQF-COSY.

Sequence-specific carbon assignments for the base and C1′ resonances were obtained from the natural abundance ^{1}H−^{13}C HMQC spectrum shown in Figure 4 using the available proton assignments. All four adenine C2s and 14 of the aromatic resonances were unambiguously assigned. Two of the aromatic resonances (corresponding to the C8s of G_{14} and G_{15}) were not seen. We believe that this is because they resonate at or very close to 135 ppm and the foldover at this frequency resulted in nulling these peaks. This is supported by evidence (Mirmira & Tinoco, following paper in this issue) of guanine C8 carbons that resonate at ≈134.7 ppm in a similar hairpin. All C1′ carbon resonances were assigned with the exception of three H1′−C1′ peaks which are superimposed in the proton dimension and are therefore indistinguishable. These peaks are the H1′−C1′ peaks of A_6, A_8, and C_{16}; the C1′ resonances are seen at 89.8, 90.6, and 91.9 ppm.

The ^{1}H−^{31}P HETCOR determines the sugar sequence along the backbone. This spectrum was assigned using the previously determined H3′ and H5′/H5″ assignments. A strong sequential H3′(i)-^{31}P(i+1) cross peak and weak intranucleotide H5′/H5″−^{31}P cross peaks were seen at the ^{31}P resonances. These scalar connectivities unambiguously confirmed the nucleotide sequence, especially in the loop region where structural assumptions cannot be made.

The proton, carbon, and phosphorus assignments are summarized in Table 1.

Apart from the assignment of the nonexchangeable proton resonances, the spectra introduced in this section also contain valuable structural information about the torsion angles. There are only a few weak cross peaks seen in the H1′−H2′ region of the DQF-COSY, indicating that for the entire molecule, the sugar pucker is largely C3′-*endo*. The cross peaks seen in this region of the COSY spectrum correspond to the residues in the loop. The magnitude of their H1′−H2′ couplings indicate that while the sugar pucker is in equilibrium between C3′-*endo* and C2′-*endo*, the equilibrium is favored toward C3′-*endo*.

The sequential H8/H6−H1′/H2′/H3′ connectivities are consistent with A-like character from G_1 to U_7 and C_{12} to C_{16}. This indicates that the A-form stacking continues for three nucleotides on the 5′ side and one nucleotide on the 3′ side of the stem. In an A-form helix, the adenine H2 protons lie in the minor groove and are close to the H1′ protons of the residue immediately 3′ to it and to the residue 5′ to it on the complementary strand (Wüthrich, 1986). Characteristically A-like cross-strand NOEs were observed from A_5H2 to G_{13}H1′ and from A_6H2 to C_{12}H1′. The H2s of the loop adenines showed only the sequential AH2−H1′ connectivity as did A_5 and A_6. This supports the evidence from the H8/H6−H1′/H2′/H3′ connectivities that the A-form helix extends beyond the G·U pair.

Unusual, long-range NOEs were seen in the 150- and 400-ms NOESY experiments from the aromatic proton of A_9 to the sugars of both U_7 and U_{11}. The NOEs seen were from A_9H8 to U_7H1′, U_7H2′, and U_{11}H1′. These NOEs suggest that A_9 is angled inward so that its H8 is close to the other nucleotides in the loop.

The sequential H8/H6−H1′ connectivities in the 400 ms NOESY were continuous except for a break between U_7H1′ and A_8H8. The sequential H8/H6−H2′ and H8/H6−H3′ connectivities were continuous for the entire molecule (except in the case of overlapped and unassigned resonances where we cannot be certain). These connectivities, as well as largely C3′-*endo* sugar pucker, would suggest a gradual turn in the hairpin loop as opposed to a sharp bend seen in smaller loops (Varani et al., 1991). In the case of a sharper turn, one would expect to see one or more unusually shifted ^{31}P resonances, or C2′-*endo* sugar puckers which extend the backbone so that fewer nucleotides are needed to span the distance between the two strands, or unusual backbone

Table 1: Proton, Carbon, and Phosphorus Chemical Shifts (ppm) for the Wild-Type Phage T4 mRNA Hairpin in 10 mM Sodium Phosphate and 10 μM EDTA, pH 6.0[a]

residue	H8/H6	H2/H5	H1′	H2′	H3′	H4′	H5′/5″ [b]
G1	8.21	na	5.78	4.85	4.73	4.55	4.51/4.26
C2	7.89	5.43	5.69	4.52	(4.47)		
C3	7.75	5.56	5.55	4.62	4.39	4.44	
U4	7.80	5.74	5.65	4.37	4.52	4.43	(4.09)
A5	8.16	7.54	5.82	4.60	4.56	4.48	4.16/4.39
A6	7.77	7.98	5.76	4.38	4.56	4.43	4.48/4.02
U7	7.42	5.42	5.45	4.22	4.45	4.25	4.20/4.01
A8	8.22	7.93	5.76	4.72	4.62	4.32	4.17/4.01
A9	8.11	8.05	5.80	4.66	4.75	4.45	4.30/4.08
C10	7.67	5.74	5.71	4.32	4.48	4.41	4.34/4.12
U11	7.74	5.71	5.47	4.48	4.37	4.41	4.25/4.12
C12	7.97	5.98	5.91	4.48	4.65	4.50	4.14/4.37
G13	7.89	na	5.67	4.70	4.42	(4.53)	
G14	7.20	na	5.65	4.58	4.51	(4.44)	(4.06)
G15	7.24	na	5.80	4.45	(4.47)		(4.03)
C16	7.48	5.25	5.76	4.00	4.12	4.17	4.47/4.01

residue	imino	amino	C8/C6	C2/C5	C1′	P
G1	(13.2)		138.2	na	90.7	na
C2	na	8.78/7.09	141.0	96.0	93.0	−4.26
C3	na	8.41/7.02	140.1	96.8	93.0	
U4	11.79	na	140.0	103.8	92.5	
A5	na		140.2	151.8	92.0	
A6	na		138.8	153.8	*c*	
U7		na	141.0	103.3	90.7	−3.98
A8	na		141.0	153.6	*c*	−3.29
A9	na		140.2	153.9	90.0	−3.82
C10	na	8.58/7.20	141.8	97.1	91.2	−4.04
U11		na	141.8	103.6	91.6	−3.73
C12	na		142.1	97.5	91.1	
G13	10.47	8.21/7.58	137.3	na	91.9	−3.67
G14	12.78	8.58/6.39	135.0	na	91.7	
G15	13.42	7.88/7.08	135.0	na	92.0	
C16	na	8.41/7.21	140.2	96.5	*c*	

[a] All proton and carbon shifts were referenced to TSP, and phosphorus chemical shifts were referenced to TMP. Imino and amino resonances were measured at 2 °C. All other proton, carbon, and phosphorus resonances were measured at 25 °C. Values in parentheses are tentative. na = not applicable. [b] For the assignment of the H5′ and H5″ resonances, the assumption made was that the H5′ proton resonates downfield from the H5″. [c] The H1′s of these nucleotides all resonate at the same frequency; hence, it is not possible to assign the C1′ chemical shifts of these residues uniquely. Resonances are seen at 89.8, 90.6, and 91.9ppm.

torsion angles, e.g., *gauche* β indicated by strong H5′/H5″−^{31}P couplings (Jaeger & Tinoco, 1993). None of these are seen in the case of the wild-type T4 RNA hairpin.

All H8/H6−H1′ NOEs are weak, indicating that the conformation about the glycosidic angle is anti for all nucleotides. χ was restricted to $-160° \pm 50°$ for all the residues in the molecule.

Natural abundance ^{1}H−^{13}C HMQC spectra were recorded at different pH values (pH 7.0, 6.0, and 4.5) to confirm either the formation or the lack of a protonated AH^{+}_{5}·C_{12} pair. The chemical shift of the C2 carbon in an adenine residue that is protonated at its N1 position moves upfield; the range from unprotonated to fully protonated is ≈8 ppm (Legault & Pardi, 1994). Our experiments (Figure 6) showed no change in the chemical shift of A_5C2 when the pH was changed from 6.0 to 7.0, indicating that at least at pH 6.0, which is the condition under which this study has been performed, there is no protonation at the N1 position of A_5. As the pH was lowered to 4.5, we saw an upfield shift of only 1.3 ppm. This indicates that the pK_a for this protonation is well below 4.5.

Backbone Conformation. The backbone conformation is determined by six torsion angles for each nucleotide (Saenger, 1984). The torsion angles were estimated on the basis of the scalar couplings of the sugar protons with each other and with ^{31}P as well as on the ^{31}P chemical shifts. The measured scalar couplings are reported in Table 2.

The torsion angles adjacent to the phosphorus atom, i.e., α (O3′−P) and ζ (P−O5′), were estimated from the chemical shifts of the ^{31}P resonances (Gorenstein, 1984). The small chemical shift dispersion of the ^{31}P resonances (≈1 ppm) ruled out any nonstandard conformation for these torsion angles, and they were both constrained to gauche conformations. Taking a conservative approach, no constraints were used for the angles around A_8′s phosphorus even though its ^{31}P resonance is shifted downfield by only ≈0.4 ppm.

The conformations around β, γ, and ϵ were obtained from scalar couplings. Torsion angles β and ϵ were estimated from the heteronuclear ^{1}H−^{31}P couplings (H5′−P and H5″−P for β and H3′−P for ϵ) obtained from the ^{1}H−^{31}P HETCOR experiment. γ was estimated from the H4′−H5′ and H4′−H5″ couplings obtained from the high-resolution DQF-COSY experiment. Scalar couplings, both homonuclear and heteronuclear, involving H5′ and H5″ were obtained without stereospecific assignments. For the stem nucleotides, β was constrained to be trans ($180° \pm 20°$), as is expected in standard A-form RNA. Residues A_8, A_9, and C_{10} showed very small H5′−P and H5″−P couplings (Table 2). β was constrained to be trans for these residues as well; however, a much broader range ($\pm 40°$) was used. γ was

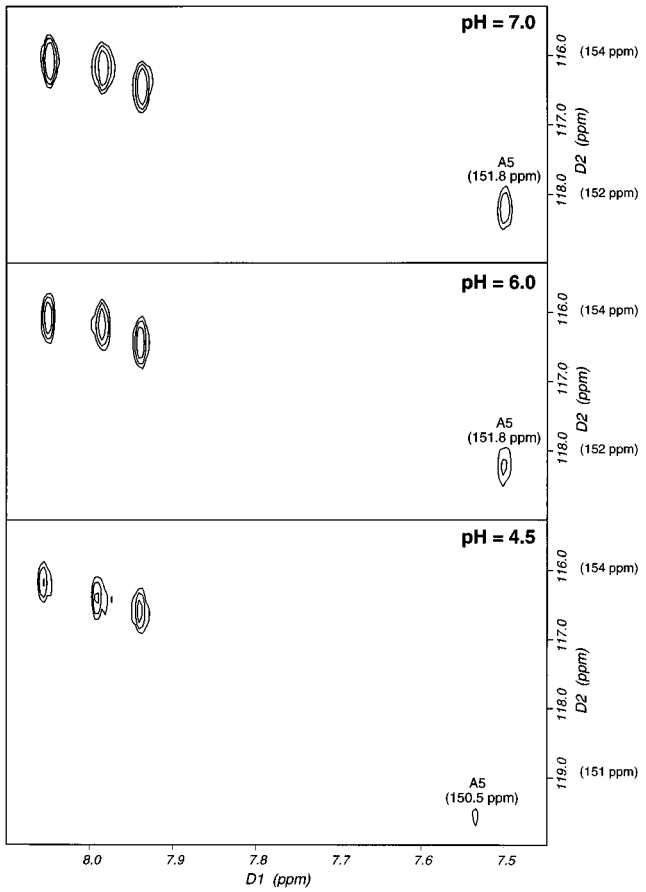

FIGURE 6: Adenine H2 (horizontal dimension) to C2 (vertical dimension) region of natural abundance $^{1}H-^{13}C$ HMQC spectra for the wild-type hairpin in D_2O. These spectra were recorded at 25 °C on an AMX-600 spectrometer. The pH conditions under which the spectra were recorded are indicated on each spectrum. The choice of carrier frequency and spectral width resulted in the folding of these cross peaks in the carbon dimension. The chemical shift positions at which they actually resonate are indicated in parentheses.

restricted to the standard A-form range of gauche$^+$ only for the stem residues. Most of the loop nucleotides showed H4′−H5′ and H4′−H5″ couplings >3 Hz, indicating that γ is in equilibrium between the different conformations and hence γ was not constrained for these nucleotides. The angle ϵ is characterized by the H3′−P coupling. Although it is known that the gauche$^+$ form is never populated due to steric hindrances, the other two conformers, gauche$^-$ and trans, are both allowed and cannot be distinguished on the basis of the H3′−P coupling alone (Altona, 1982). This torsion angle was constrained to the standard A-form value of −155° ± 20° (Saenger, 1984) only for the residues in the stem.

The angle δ was not used to restrain the sugar pucker since it has been noticed that the sugar pucker is not well determined when only this one degree of freedom is restrained (Wimberly, 1992). Instead, nucleotides with weak or absent H1′−H2′ couplings (residues C_2 to A_5 and G_{13} to C_{16}) were constrained C3′-*endo* by constraining the four endocyclic torsion angles in the sugar ring (ν_0, ν_1, ν_2, and ν_3). The four endocyclic torsion angles for the other nucleotides were constrained in a broad range which encompassed C2′-*endo*, C3′-*endo*, and the O4′-*endo* sugar conformations. The percent C3′-*endo* conformation (=% N) reported in Table 2 was calculated from the magnitude of the H1′−H2′ couplings (de Leeuw & Altona, 1982).

Figure 7 summarizes the data obtained from the different NMR experiments.

Structure Calculations. Twenty structures with random torsion angles were refined as described in Materials and Methods using the torsion angles determined as above and 251 interproton distance constraints obtained from the NOESY spectra at different mixing times (see Materials and Methods). An additional four constraints for the C_2·G_{15} and C_3·G_{14} and five constraints for the U_4·G_{13} pairs were used to maintain the hydrogen bonding between the paired bases and to keep them planar. The entire 16-nucleotide sequence of the wild-type T4 RNA hairpin was used to generate the starting structures with randomized backbone torsion angles. These were then subjected to the global fold protocol, followed by the refinement protocol and finally the energy minimization step as described in Materials and Methods. Eleven of the initial 20 converged to low-energy structures. None of the 11 final structures violate any of the experimentally derived NOE constraints by more than 0.2 Å, although there are minor violations of the distances used to constrain the hydrogen bonding between the base pairs.

In order to assess how precisely the structures were determined by the NMR data, these 11 structures were superimposed, and an average structure was calculated. The root mean square deviation (RMSD) for each structure compared to the average structure was calculated. The average RMSD when all 16 nucleotides are superimposed is 1.24 Å (standard deviation = 0.19 Å) for the 11 converged structures. When the first and last nucleotides (G_1 and C_{16}), which were not constrained very much by the NMR data, were left out of the calculation, the average RMSD dropped to 1.04 Å (±0.21 Å). The three base pairs, C_2·G_{15}, C_3·G_{14}, and U_4·G_{13}, are very precisely defined. Unlike a TAR hairpin loop (Jaeger & Tinoco, 1993), the loop structure of the T4 hairpin is well defined. Superimposing just the loop nucleotides (A_5 to C_{12}) of all 11 structures gave an average RMSD value of 1.12 Å (±0.27 Å). Figure 8a shows a ribbon representation of the phosphate backbones of the 11 converged structures superimposed from residue C_2 to G_{15}.

The bases of loop residues A_8, A_9, and C_{10} are out in solution in all the 11 structures. In the majority of the structures, they are also stacked on one another with their Watson−Crick faces pointing out into the solution. These structures were examined for the conformation of the loop nucleotides after the refinement step, and the results obtained were similar. The final minimization step, in which the electrostatics term was added to the force field, had little or no impact on the structure of the loop. The average RMSD values obtained for the superimposing of the loop nucleotides are very similar for both the post-refinement (i.e., pre-final minimization) and post-final minimization structures.

Figure 8b shows the overlap of 6 of the 11 final structures of the T4 hairpin. These 6 structures all have similar loop conformations, with A_8, A_9, and C_{10} stacked on each other. This is the predominant loop conformation. In the other five structures, while the bases of A_8, A_9, and C_{10} are out in solution, there are slight variations in the extent to which they stack on each other. The structures shown in Figure 8b superimpose with an average RMSD of 0.81 Å (±0.18 Å) for nucleotides C_2 to G_{15} and 0.77 Å (±0.09 Å) for the loop nucleotides.

In all 11 final structures, the loop nucleotide, A_9, is very well defined. Its H2 proton points directly out in solution

Table 2: Scalar Couplings (in Hz) for the Wild-Type Phage T4 mRNA Hairpin at 25 °C in 10 mM Sodium Phosphate and 10 μM EDTA, pH 6.0[a]

residue	$J_{1'2'}$	% N	$J_{2'3'}$	$J_{3'4'}$	$J_{4'5'/5''}$	$J_{5'5''}$[b]	$J_{P(i+1)3'}$[b]	$J_{P(i)5'/5''}$[b]
G1	2.5	80	4.3	8.3	4.4/<2	11.8	8.2	
C2	2.3	83						
C3	2.6	79	4.2	9.8				
U4	2.3	83	4.4	9.4				
A5	2.4	82	4.5	8.9				
A6	2.8	76	4.5[b]	8.3	4.7[b]/<2		7.8	
U7	3.1	71	4.6[b]	7.1	4.0/4.0	11.4	7.8	3.6/4.7
A8	3.3	68	4.7[b]	6.4	4.2/4.3	11.7	6.8	2.3/2.4
A9	3.1	71	4.9[b]	7.3	4.1/4.1	11.5	7.2	2.2/2.2
C10	3.0	73	4.0	7.5	4.4/4.1	11.7	7.6	2.2/2.4
U11	2.8	76	4.7	9.0			6.2	
C12	3.0	73	4.6[b]	7.3			6.2	
G13	2.3	83	4.6	9.6				
G14	<2	>90	4.8	9.7				
G15	<2	>90						
C16	2.2	84	4.5	8.8	3.9/<2	10.9	10.9	10.9

[a] All uncertainties are ±0.5 Hz unless otherwise mentioned. For the H4′−H5′/H5″ and the P(*i*)−H5′/H5″ couplings, the couplings to the downfield proton are reported first. [b] Uncertainties are ±1 Hz.

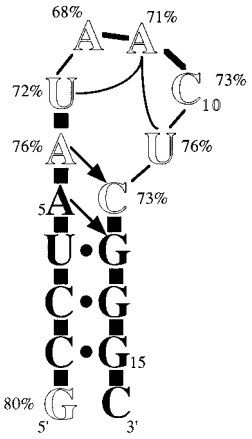

FIGURE 7: Schematic summary of the NMR data in this study. Nucleotides shown in solid type have C3′-*endo* sugar puckers while those in outline type are in equilibrium between C3′-*endo* and C2′-*endo* conformations. The percent (±7%) C3′-*endo* is shown next to the base. Standard A-like cross-strand NOEs seen from the adenine H2s in this molecule are indicated by arrows. The solid line shows the nucleotides between which unusual, long-range NOEs are seen. The dots represent hydrogen-bonded imino protons. The black boxes between the bases indicate standard A-form stacking. In the loop region, the presence of connectivities from one nucleotide to the next are shown by lines between the residues. The thick lines between A_8, A_9, and C_{10} represent the stacking seen between them.

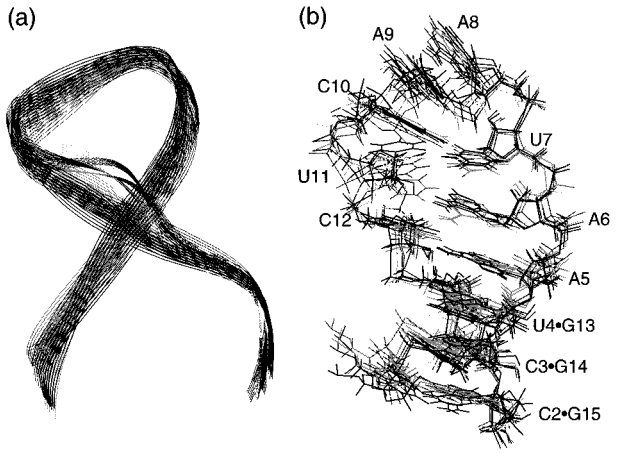

FIGURE 8: (a) Ribbon representation of the phosphate backbones of the 11 converged structures superimposed from residue C_2 to G_{15}. (b) Overlap of six of the structures of the T4 hairpin with the nucleotides numbered.

so that the H8 proton is angled in toward the minor groove face of the loop. This nucleotide was well constrained with a total of 30 NOE-derived distance constraints. Of these, three are intrasugar constraints and, hence, of not much importance. Most (21) of the remaining 27 important constraints are in the "weak" or "very weak" ranges. This directly demonstrates that residues can be well constrained even with weak constraints as long as there are sufficiently many of them and indirectly demonstrates the importance of the complete assignment of the different resonances in the molecule. The three unusual and long-range NOEs seen from U_7H1′, U_7H2′, and U_{11}H1′ to the H8 proton of A_9 play an important role in constraining this nucleotide such that its H8 proton is angled in toward the other nucleotides in the loop. These were not tight constraints (two of them were in the 2.5−5 Å range and the third was in the 3−7 Å range). However, when the structures were calculated with just these three distance constraints omitted, A_9 was clearly more flexible.

Figure 9 shows the three-dimensional structure of one of the converged structures of the wild-type phage T4 RNA hairpin. In this figure, as well as in Figure 8b, the stacking of loop nucleotides A_5, A_6, U_7, and C_{12} can be seen. It is likely that the compact loop structure is stabilized by hydrogen bond interactions within the loop. The O2 of residue U_7 is in the right geometry for hydrogen bond formation with the amino proton of C_{10}. This geometry of these two bases was seen in all of the six structures shown in Figure 8b. Also, as can be seen from Table 1, one of the amino protons of C_{10} is shifted considerably downfield (8.58 ppm), indicating its involvement in hydrogen bond formation. The average distance between U_7O2 and C_{10}N4 is 2.7 Å and the average angle between the heteroatoms is 167° ± 7°. Similarly, there is potential for hydrogen bond formation between C_{12}O2 and the A_5 amino proton (average distance between the N6 and O2 is 2.7 Å and the average angle is 158° ± 11°), U_{11}O2 and the imino proton (H3) of U_7 (average distance between U_7N3 and U_{11}O2 is 2.5 Å and the average angle is 166° ± 6°), and A_9O4′ and $A_8$2′OH (the average distance between the two oxygen atoms is 2.7 Å and the average angle is 151° ± 19°). The maximum

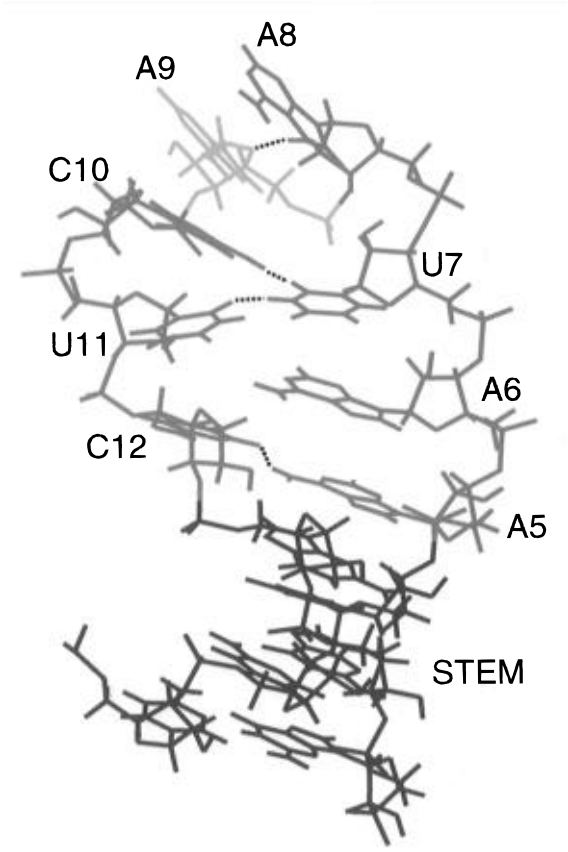

FIGURE 9: Representative structure of the wild-type phage T4 RNA hairpin. The A_9H2 proton is seen pointing directly out in solution. The putative hydrogen bonds in the loop (between U_7O2 and the amino proton of C_{10}, C_{12}O2 and the A_5 amino proton, U_{11}O2 and the imino proton of U_7, and A_9O4′ and $A_8$2′OH) are shown by black dashed lines.

range for these distances is ±0.1 Å except for one structure in which A_8O2′ and A_9O4′ are 3.4 Å apart.

DISCUSSION

The wild-type phage T4 hairpin plays an important role in the translational regulation of the phage's DNA polymerase gene. By binding to a region of its mRNA which includes this hairpin sequence, the polymerase prevents the ribosome from binding the mRNA and thereby achieves autoregulation (Andrake et al., 1988). This hairpin is crucial for the binding of the polymerase to the mRNA, and mutations in this hairpin greatly diminish the affinity of the polymerase for the RNA (Tuerk et al., 1990). Why is this hairpin structure so important and what information about the protein−RNA interaction can we get from its structure?

This study shows that the wild-type hairpin, despite having an eight-base loop, is fairly rigid. There is overwhelming evidence, as discussed below, to indicate that the stem is not confined to four base pairs but is extended on both the 5′ and 3′ sides.

• The results from the 1D temperature dependence of the imino spectrum indicate that not only is the U_4·G_{13} pair formed but that it is the last one to melt. This is indicative of the presence of bases that stack over this pair and protect its imino protons from rapid exchange with the solvent.

• The sequential H8/H6−H1′/H2′/H3′ connectivities are consistent with A-like character from G_1 to U_7 and C_{12} to C_{16}, indicating that the A-form stacking on the stem continues for three nucleotides on the 5′ side and one nucleotide on the 3′ side.

• The characteristically A-like cross-strand NOEs observed from A_5H2 to G_{13}H1′ and from A_6H2 to C_{12}H1′ are, once again, clear evidence that the A-form helix extends up to and includes A_5, A_6, and C_{12}.

• In an A-form helix, the AH2 proton is located directly above the plane of the base pair below it. This results in a shielding effect and hence the proton resonates more upfield. This upfield shift is seen in many molecules studied in this laboratory. In this study, we see that the A_5H2 proton resonates upfield compared to the other AH2s in the molecule. This further corroborates the stem−extension hypothesis.

• The preciseness with which the three base pairs, C_2·G_{15}, C_3·G_{14}, and U_4·G_{13}, superimpose is further indication that these pairs are shielded from both ends. In the case of a four-base-pair TAR hairpin, only the central two base pairs superimpose well; the pair at the stem−loop junction is much less defined due to the flexibility of the loop (Jaeger & Tinoco, 1993). In the case of the T4 hairpin, not only are these three pairs very well defined but the residues included in the extended stem are also well defined. Residues C_2 to U_7 and C_{12} to G_{15} superimpose with an average RMSD of 0.69 Å (±0.15 Å).

The extension of the stem and hence the shortening of the loop make the loop less flexible. This could well serve to constrain the loop in a conformation recognized by the protein. It is also feasible that hydrogen bonds, like the ones we propose between U_7O2 and the amino proton of C_{10}, between C_{12}O2 and the A_5 amino proton, between U_{11}O2 and the imino proton of U_7, and between A_9O4′ and $A_8$2′OH, help keep the loop rigid.

The phosphorus resonances for this molecule are all within 1 ppm of each other, suggesting that none of the α or ζ torsion angles are trans. The sugar pucker, even in the loop region, is predominantly C3′-*endo*. There is no evidence for any sharp bend or turn in the loop. We believe that even though some of the loop nucleotides stack in an A-like conformation on the stem, there are sufficient nucleotides left in the loop for it to turn and bridge the gap between the two strands. There is a gentle turn in the loop region as contrasted to sharp bends seen in other hairpin loops (Varani et al., 1991; Jaeger & Tinoco, 1993).

Selection for hairpin loops that best bind the polymerase (Tuerk & Gold, 1990) produced a "consensus sequence" of ARYAACYY (corresponding to residues 5−12 in the hairpin in this study). It is noteworthy that the AAC sequence corresponding to the bases we see pointing out in solution was nearly invariant. Even though in a total of 2 out of 20 clones A8 was changed to a G, no instances of residue 9 being anything but an adenine were found. Similarly, residue 10 was always a cytosine. In this hairpin, these three bases are out in solution with their amino groups, N1s or N3s, and the carbonyl group (on C_{10}) all available for hydrogen-bonding interactions. It is possible that it is this sequence of three nucleotides that the protein recognizes and uses to bind the mRNA.

CONCLUSIONS

The solution structure of a bacteriophage T4 RNA hairpin involved in the translational repression of its DNA polymerase has been determined by NMR spectroscopy. 1D and 2D, homonuclear and heteronuclear NMR experiments were

used to assign the proton, carbon, and phosphorus resonances and to obtain the distance and torsion angle constraints used to calculate the structure. The hairpin is well defined by these constraints. The eight-nucleotide loop is shortened by stacking on both the 5′ and 3′ side of the stem. The absence of C2′-*endo* sugar pucker, unusual torsion angles, or shifted phosphorus resonances suggests a gently turning loop. Three bases are out in solution with the groups capable of hydrogen bond formation exposed to the solution. The position of A_9 is well defined by the unusual long-range NOEs seen to the sugar protons of U_7 and U_{11}. The H2 proton and the amino group of A_9 point directly out in solution so that the H8 is pointing downward and is angled in toward the loop.

The structure of this hairpin has been determined to a high degree of precision. Six of the 11 converged structures superimpose to an average RMSD of about 0.8 Å and all 11 to approximately 1 Å for residues C_2-G_{15}.

ACKNOWLEDGMENT

We thank Ms. Barbara Dengler for managing the laboratory and Mr. David Koh for synthesizing the DNA templates. Our sincere thanks to Dr. Gabriele Varani for his critical reading of the manuscript and his valuable comments.

REFERENCES

Altona, C. (1982) *Recl. Trav. Chim. Pays-Bays 101*, 413−433.
Andrake, M., Guild, N., Tien, H., Gold, L., Tuerk, C., & Karam, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7942−7946.
Bax, A., Griffey, R. H., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301−315.
Bax, A., Davis, D. G., Gregory, R. J., Cahill, P. B. Thurlow, D. L., & Zimmermann, R. A. (1985) *J. Magn. Reson. 65*, 295−307.
Brünger, A. T. (1990) *X-PLOR: A System for Crystallography and NMR, Version 2.1*, Yale University, New Haven, CT.
de Leeuw, F. A. A. M., & Altona, C. (1982) *J. Chem. Soc., Perkin Trans. 2*, 375−384.
Derome, A. E. (1987) *Modern NMR Techniques for Chemistry Research*, Pergamon Press, New York.
de Waard, A., Paul, A., & Lehman, I. R. (1965) *Proc. Natl. Acad. Sci. U.S.A. 54*, 1241−1248.
Gorenstein, D. G. (1984) *Phosphorus-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) Academic Press, New York.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.
Heus, H. A., & Pardi, A. (1991) *J. Am. Chem. Soc. 113*, 4360−4361.
Jaeger, J. A., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 12522−12530.
Legault, P., & Pardi, A. (1994) *J. Am. Chem. Soc. 116*, 8390−8391.
Macura, S., Wüthrich, K., & Ernst, R. R. (1982) *J. Magn. Reson. 46*, 269−282.
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.
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.
Russel, M. (1973) *J. Mol. Biol. 79*, 83−94.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94−98.
Tuerk, C., & Gold, L. (1990) *Science 249*, 505−510.
Tuerk, C., Eddy, S., Parma, D., & Gold, L. (1990) *J. Mol. Biol. 213*, 749−761.
Varani, G., & Tinoco, I., Jr. (1991a) *J. Am. Chem. Soc. 113*, 9349−9354.
Varani, G., & Tinoco, I., Jr. (1991b) *Q. Rev. Biophys. 24*, 479−532.
Varani, G., Cheong, C., & Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
Warner, H. R., & Barnes, J. E. (1966) *Virology 28*, 100−107.
Wimberly, B. (1992) Ph.D. Thesis, University of California, Berkeley.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wyatt, J. R., Chastain, M., & Puglisi, J. D. (1991) *BioTechniques 11*, 764−769.

BI960414Y