Solution Conformation of a Five-Nucleotide RNA Bulge Loop from a Group I Intron[†,‡]

Kevin J. Luebke,* Stacy M. Landry, and Ignacio Tinoco, Jr.

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

Received January 23, 1997; Revised Manuscript Received June 16, 1997[⊗]

ABSTRACT: We present the solution conformation, determined by NMR spectroscopy, of a five-nucleotide RNA bulge loop. The bulge interrupts the stem of a 25-nucleotide RNA hairpin, and its sequence and flanking sequences are those of a conserved bulge from a Group I intron. The secondary structure of the bulge loop in the hairpin context is that predicted by the secondary structure prediction algorithm of Zuker. It differs, however, from the secondary structure deduced from sequence covariation of the bulge in the context of the functionally folded Group I introns and observed in the crystal structure of an independently folding domain of the Group I intron from *Tetrahymena thermophila*. This difference represents an exception to the heierarchical model of RNA folding in which preformed elements of secondary structure interact to form a tertiary structure. The three-dimensional structure of the bulge loop is characterized by discontinuous base stacking. Adjacent adenines stack with each other and with the flanking double helices. However, the position of the central uracil is not well defined by NOE distance constraints and is a point of discontinuity in the base stacking.

RNA bulge loops, which interrupt one strand of an otherwise continuous double helix, occur frequently in the secondary structures of large RNAs. Many bulge loops are phylogenetically conserved, indicating functional importance. In some cases, they have been shown to participate in tertiary interactions (Flor *et al.*, 1989; Cate *et al.*, 1996) or specific interactions with RNA binding proteins (Peattie *et al.*, 1981; Romaniuk *et al.*, 1987; Wu & Uhlenbeck, 1987; Weeks & Crothers, 1991; Puglisi *et al.*, 1992; Aboul-ela *et al.*, 1995).

NMR[1] and crystallographic studies of single-nucleotide bulges in RNA (van den Hoogen *et al*., 1988b) and DNA (Hare *et al.*, 1986; Joshua-Tor *et al.*, 1988; Kalnik *et al.*, 1989, 1990; Miller *et al.*, 1988; Morden *et al.*, 1983, 1990; Nikonowicz *et al.*, 1989; van den Hoogen *et al.*, 1988a; Woodson & Crothers, 1988a,b, 1989) have revealed an equilibrium between conformations in which the extra base is intercalated into the helical stack or extruded from the helix. The position of this equilibrium depends upon the identities of the extra base and the flanking base pairs as well as temperature. In the three-nucleotide DNA bulges of 5′ATA3′ (Rosen *et al.*, 1992) and 5′$A_3$3′ (Aboul-ela *et al.*, 1993), the bulge nucleotides have been shown by NMR to be intercalated into the helical stack. NMR studies (Puglisi *et al.*, 1992; Aboul-ela *et al.*, 1996) of the three-nucleotide RNA bulge (5′UCU3′) from the *trans*-activation response (TAR) element of HIV have, in the absence of ligand, shown continuity of stacking between the two 5′ nucleotides in the bulge and the base-paired nucleotides flanking the bulge at its 5′ side. The 3′ uracil is partially or transiently extrahelical. On binding argininamide, the bulge loop rearranges to a conformation in which all of the nucleotides are extrahelical (Puglisi *et al.*, 1992; Aboul-ela *et al.*, 1995).

Less information is available about the conformations of larger bulge loops. Chemical and enzymatic probing of five-nucleotide DNA bulges (Bhattacharyya & Lilley, 1989) and a six-nucleotide RNA bulge from 16S rRNA (Moazed *et al.*, 1986) have suggested conformational complexity and diversity for large bulges, but only one high-resolution structure is available for a bulge loop of more than three nucleotides. The recently reported crystal structure of the P4-P5-P6 domain of the Group I intron from *Tetrahymena thermophila* (Cate *et al.*, 1996) includes a five-nucleotide bulge loop. The adenosine-rich sequence of the bulge and its position approximately 23 base pairs from the catalytic core of the intron are both conserved within the P5A extension in the introns of Groups Ib and Ic (Figure 1) (Collins, 1988; Michel & Westhof, 1990; Michel & Cummings, 1985). Murphy and Cech (1993, 1994) have shown that this bulge loop is necessary for the tertiary folding of the structural domain that includes helices P4, P5, and P6. In the crystal structure, all the bulge nucleotides reside outside of the helix they interrupt and form a network of tertiary hydrogen bonds and stacking interactions with residues in the P4 helix and the P5a,b,c junction.

[†] This research was supported in part by National Institute of Health Grant GM 10840 and by Department of Energy Grant DE-FG03-86ER60406 and by instrumentation grants from the Department of Energy (DE-FG05-86ER752810) and the National Science Foundation (DMB 86-09305 and DMB 87-20134). K.J.L. was supported in part by a postdoctoral fellowship from the American Cancer Society. S.M.L. was supported in part by a National Institute of Health Molecular Biophysics Training Grant T32 GM08295.

[‡] The coordinates of representative structures have been deposited in the Brookhaven Protein Data Bank under the identification codes 1ajl and 1ajt.

* Author to whom correspondence should be addressed at Hewlett-Packard Co., 3500 Deer Creek Rd., Building 25U-5, Palo Alto, CA 94304.

[⊗] Abstract published in *Advance ACS Abstracts,* August 1, 1997.

[1] Abbreviations: NMR, nuclear magnetic resonance; UV, ultraviolet; TSP, 3-(trimethylsilyl)propionate-d_4; TPPI, time proportional phase incrementation; ppm, parts per million; NOE, nuclear Overhauser effect; FID, free induction decay; NOESY, nuclear Overhauser enhancement spectroscopy; COSY, correlated spectroscopy; HMQC, heteronuclear multiple quantum coherence spectroscopy; TOCSY, total correlation spectroscopy; RMSD, root mean square deviation.

S0006-2960(97)00154-2 CCC: $14.00

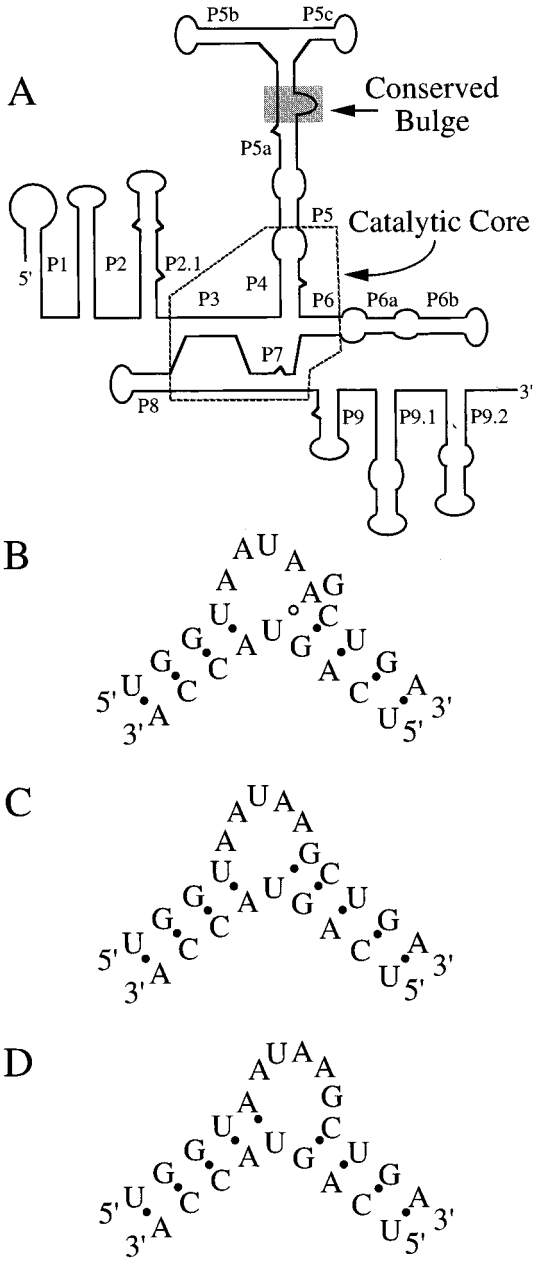

FIGURE 1: (A) Schematic representation of the secondary structure of the Group I intron from *T. thermophila*. The conserved five-nucleotide bulge is shaded, and the catalytic core is outlined. (B−D) Bulge and flanking sequences with alternative secondary structures. The pairing in B is predicted by sequence covariation of the bulge in the context of the functionally folded Group I introns and observed in the crystal structure of an independently folding domain of the Group I intron from *T. thermophila* (Cate *et al*., 1996). The pairing in C is predicted by the algorithm of Zuker (1989).

Covariation of nucleotides between Group I introns from different species indicates a secondary structure (Figure 1B) in which the five bulged nucleotides are divided into a one-nucleotide bulge and a four-nucleotide bulge separated by a single base pair (Michel & Westhof, 1990). This arrangement is found in the crystal structure, where the base pair that divides the one- and four-nucleotide bulges is a Hoogsteen A·U base pair (adenosine in the *syn* conformation). This secondary structure differs from that predicted based on nearest-neighbor thermodynamic parameters: the secondary structure predicted by the algorithm of Zuker (1989) (Figure 1C) has U·A and G·U base pairs flanking a loop of sequence 5′AAUAA3′. A third plausible pairing arrangement flanks a bulge sequence of 5′AUAAG3′ by A·U and C·G base pairs (Figure 1D).

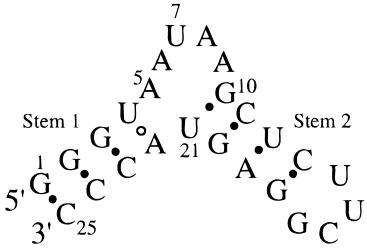

FIGURE 2: Sequence of 25-nucleotide RNA for NMR spectroscopic study. The secondary structure determined by NMR is shown.

We have used NMR spectroscopy to determine the solution conformation of the conserved five-nucleotide bulge loop from the self-splicing Group I intron of *T. thermophila* in the context of a 25-nucleotide RNA hairpin. The hairpin contains in its stem the five-nucleotide bulge and flanking sequence from P5a of the Group I intron from *T. thermophila* (Figure 2). Our spectroscopic data clearly define the secondary structure of the bulge within this context. In addition, NMR-derived distance and torsion-angle constraints define, to atomic resolution, the conformational features of this bulge loop.

MATERIALS AND METHODS

Synthesis and Purification of RNA. The oligoribonucleotide 5′GGGUAAUAAGCUCUUCGGAGUACCC3′ was enzymatically synthesized *in vitro* using T7 RNA polymerase and a chemically synthesized oligodeoxyribonucleotide template (Davanloo *et al*., 1984; Wyatt *et al.*, 1991). The transcription product was purified by denaturing polyacrylamide gel electrophoresis [20% (w/v) acrylamide, 29:1 cross-linking, 7 M urea]. Typical yields of RNA were 4−8 optical density units per milliliter of transcription reaction.

UV Absorbance-Monitored Thermal Denaturation. Unimolecular folding of the oligonucleotide was confirmed by the concentration independence of the UV absorbance *versus* temperature profile (Marky & Breslauer, 1987; Puglisi & Tinoco, 1989). Samples were analyzed in 10 mM sodium phosphate, 50 μM EDTA, pH 6.7, at oligonucleotide concentrations of 5, 62, 304, and 788 μM. Data were recorded at 260 nm on a thermoelectrically controlled Gilford spectrophotometer. The temperature was increased at the rate of 1 °C/min.

NMR Spectroscopy. NMR samples were dialyzed for 2−4 days against 10 mM sodium phosphate, 50 μM EDTA, pH 6.7, at 4 °C. Following dialysis, samples were lyophilized to dryness. To measure spectra of exchangeable protons, the sample was dissolved in a 10:1 mixture of H_2O and D_2O. To measure spectra of non-exchangeable protons, the sample was dissolved in D_2O and lyophilized to dryness several times before being dissolved in 650 μL of 99.96% D_2O (Aldrich). The concentration of RNA was 2.5 mM. Before each experiment, the sample was warmed to 70 °C for approximately 1 min and then allowed to cool to room temperature. All spectra were acquired on either a GE GN-500 spectrometer operating at 500 MHz proton frequency, a BRUKER AMX-600 spectrometer operating at 600 MHz proton frequency, or a BRUKER AMX-300 spectrometer operating at 300 MHz proton frequency. Spectra were

acquired at 25 °C except where otherwise indicated. NMR data were processed and displayed using the program FELIX, version 2.30 (Biosym Technologies, Inc.).

One-dimensional exchangeable proton spectra in H_2O were collected with a 1-1 water suppression sequence and a 125 μs delay between pulses. One-dimensional NOE data were collected with a 1331 solvent suppression scheme (Hore, 1983), the excitation maximum set in the center of the imino region of the spectrum. Preirradiation was for 200 ms, and a 16K data set was collected for irradiation of each peak. Reference data were subtracted from data collected following irradiation, and the resulting difference FIDs were apodized with 2 Hz line broadening prior to Fourier transformation.

All multidimensional NMR spectra were recorded in the phase-sensitive mode using the TPPI method (Marion & Wüthrich, 1983). A NOESY spectrum of the sample in H_2O was acquired at 10 °C (600 MHz spectrometer). Water suppression was achieved with a 1-1 pulse sequence having a 150 μs delay between pulses. 80 scans were taken for each of 371 FIDs of 1024 complex points. The spectral width in both dimensions was 12 500 Hz. In all spectra of nonexchangeable protons, the residual HDO resonance was suppressed by presaturation during a 2.5 s relaxation delay. NOESY spectra of the sample in D_2O were acquired with mixing times of 60, 100, 120, 150, and 400 ms at 25, 20, and 10 °C.

A double-quantum filtered COSY spectrum was acquired with a spectral width of 4000 Hz (500 MHz spectrometer). 946 FIDs were collected with 32 scans of 1024 complex points for each FID. A high-resolution double-quantum filtered COSY spectrum was acquired by decreasing the spectral width to 2000 Hz (600 MHz spectrometer). 735 FIDs were collected with 32 scans of 2048 complex points for each FID.

A natural abundance heteronuclear $^1H-^{13}C$ multiple quantum coherence (HMQC) spectrum was acquired (600 MHz spectrometer) using a standard pulse sequence (Bax *et al.*, 1983; Varani & Tinoco, 1991a) and the GARP1 sequence for carbon decoupling during acquisition. The spectral width was 5000 Hz (8.3 ppm) in the ^1H dimension and 7500 Hz (50 ppm) in the ^{13}C dimension. 84 FIDs were collected with 160 scans of 2048 complex points. A proton-detected heteronuclear $^1H-^{31}P$ correlated spectrum was acquired (300 MHz spectrometer) as proposed by Sklenar *et al.* (1986). Spectral widths were 1000 Hz in the ^1H dimension and 700 Hz in the ^{31}P dimension. 80 scans of 2048 complex data points were collected for each of 160 t_1 increments.

A TOCSY spectrum (Griesinger *et al.*, 1988) was acquired (500 MHz spectrometer) with a compensated MLEV17 sequence for mixing (Bax *et al.*, 1985) and GARP decoupling of ^{31}P, and 48 scans of 2048 complex data points were collected for each of 494 t_1 increments. A mixing time of 60.6 ms was used. A three-dimensional homonuclear TOCSY−NOESY was acquired (600 MHz spectrometer) as described previously (Wijmenga *et al.*, 1994). The spectral width in each dimension was 3311 Hz. Eight scans of 512 real points were collected for each of 17 740 FIDs. Total points acquired for each dimension were 512 for t_3, 136 for t_2, and 128 for t_1. A 73 ms TOCSY mixing time and a 175 ms NOESY mixing time were used.

Structure Determination. The three-dimensional structure of the RNA bulge loop was calculated using a restrained molecular dynamics protocol incorporating NMR-derived distance and torsion angle constraints (Wimberly, 1992). The UUCG hairpin loop and the terminal G·C base pair were excluded from the calculation. Where a Watson−Crick or G·U base pair was indicated by the observation of an imino proton resonance, distance constraints were applied to maintain the appropriate hydrogen bonding distances and coplanarity of the bases. Four constraints were used for each Watson−Crick base pair, and five constraints were used for the G·U base pair.The distance between imino protons for which an NOE was observed was constrained to the range 3.0−5.0 or 2.5−4.5 Å. NOESY cross-peaks observed between imino and amino protons in H_2O were assigned a distance range of 1.8−5.0 Å.

NOE cross-peak intensities were used semiquantitatively to assign distance ranges to nonexchangeable protons. Cross-peak intensities were characterized as strong, medium, weak, and very weak according to the NOESY mixing times at which they were observable. Strong cross-peaks, observable with a mixing time of 60 ms, were assigned the range 1.8−3.0 Å. Medium-intensity cross-peaks, observable with a NOESY mixing time of 100 ms or longer, were assigned the range 2.0−4.0 Å. Weak cross-peaks, observable with mixing times of 150 ms or longer, were assigned the range 2.5−5.0 Å. Very weak cross-peaks were only seen in NOESY spectra with a 400 ms mixing time and were assigned the range 2.5−6.0 Å. Some pairs of nonexchangeable protons for which NOESY cross-peaks were unambiguously absent were constrained to a distance >5.0 Å.

The sugar conformations were characterized by the H1′−H2′ scalar couplings evident in COSY experiments. Residues for which no H1′−H2′ coupling was observed were constrained to the C3′-*endo* conformation with the endocyclic torsion angles ν_0, ν_1, ν_2, and ν_3. Residues for which a small (<5.5 Hz) H1′−H2′ coupling was observed were constrained to the range of conformations including C2′-*endo*, O4′-*endo*, and C3′-*endo*. Sugars with a large (>5.5 Hz) H1′−H2′ coupling were constrained to the C2′-*endo* conformation.

The glycosidic torsion angle, χ, was constrained to the *anti* conformation in each stem residue. It was also constrained to the *anti* conformation in the bulge loop residues for which the intranucleotide NOE cross-peak between H1′ and aromatic resonances was similar in intensity to the intranucleotide NOE cross-peak between H1′ and aromatic resonances in the stem nucleotides. This condition was not clearly met for U7 and A8, and the glycosidic torsion angles of these residues were not constrained. Backbone torsion angles β, γ, ϵ, α, and ζ were not constrained for loop nucleotides A5 through A9 and were constrained to A-form values for the stem nucleotides. β, γ, and ϵ, were specified $\pm 20°$, and α and ζ were specified $\pm 30°$. Torsion angle constraints were applied to fix exocyclic amino groups in the plane of the bases.

The molecular dynamics program XPLOR 2.2 (Biosym Technologies, Inc.) was used to generate three-dimensional structures consistent with the NMR data (Wimberly, 1992). Forty-two starting structures with randomized backbone torsion angles were created and subjected to a simulated annealing protocol. This stage of structure calculation began with 500 cycles of initial energy minimization followed by 15 ps of restrained molecular dynamics at 1000 K. Constraints on backbone torsion angles beta, gamma, epsilon, alpha, and ζ were omitted from the simulated annealing. The repulsive van der Waals forces were initially turned off so

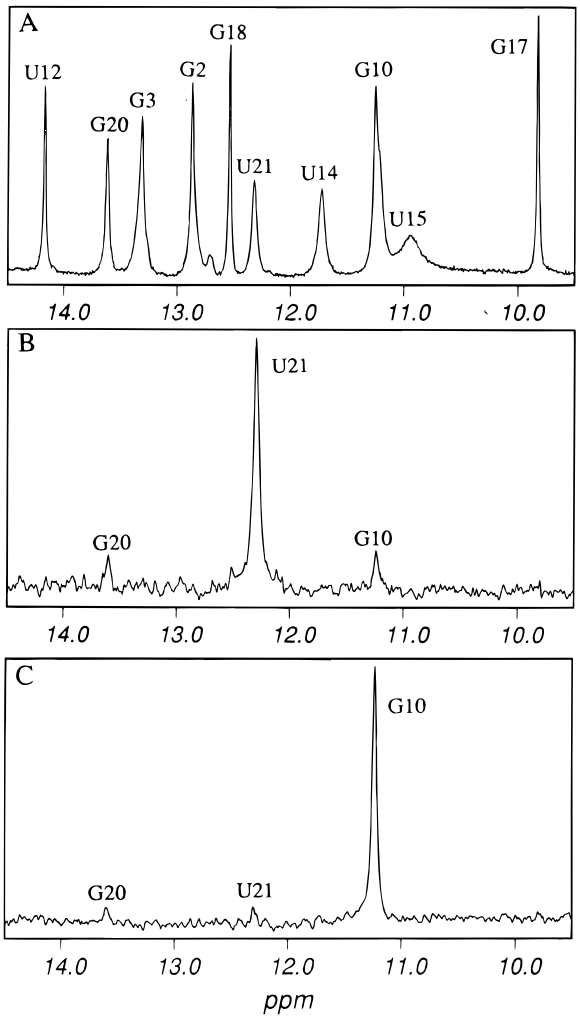

FIGURE 3: (A) NMR spectrum of imino protons. The peaks are labeled according to the assignments described in the text. Conditions were as described in the Materials and Methods section. (B) One-dimensional NOE difference spectrum with irradiation of the imino proton of U21 at 12.31 ppm. (C) One-dimensional NOE difference spectrum with irradiation of the imino proton of G10 at 11.27 ppm.

that atoms could pass through each other but were gradually increased as the system was cooled to 300K. The calculation was concluded with 1000 cycles of energy minimization.

The twenty structures with the lowest NOE violation energies were subjected to a refinement protocol. This stage began with 500 cycles of energy minimization followed by 1 ps of restrained molecular dynamics at 1000 K. Backbone torsion angle constraints were introduced in two steps: β, γ, and ϵ first, followed by α and ξ. The system was cooled to 300 K, and refinement was concluded with 2000 steps of energy minimization. Structures were displayed using the program INSIGHT II (Biosym Technologies, Inc.) on a Silicon Graphics workstation.

RESULTS

Assignment of Exchangeable Protons and Secondary Structure. The imino proton region of the NMR spectrum of the oligonucleotide is shown in Figure 3. The sharp peaks at 9.8 ppm and 11.7 ppm and a broad peak at 10.9 ppm are characteristic of the 5′UUCG3′ hairpin structure and can be assigned by comparison to previous work (Varani *et al.*, 1991). The melting temperature of the structure, measured by UV absorbance-monitored thermal denaturation, is 56 °C and independent of oligonucleotide concentration under the conditions of the NMR experiment, also consistent with a unimolecular hairpin conformation (Marky & Breslauer, 1987; Puglisi & Tinoco, 1989). A series of sequential NOEs between imino protons, starting from the imino proton of G17 in the hairpin loop, allowed assignment of resonances due to imino protons of G18, U12, and G20. The chemical shifts of these protons were consistent with their participation in the A·U and G·C base pairs anticipated. At chemical shifts corresponding to imino protons involved in non-Watson−Crick base pairing, two imino protons have NOEs to each other as well as to the imino proton of G20 (Figure 3B and 3C). This pattern of NOE connectivities is consistent with formation of a G·U base pair between G10 and U21.

Consecutive NOEs were found between two imino protons involved in G·C base pairs, and those protons were assigned to G2 and G3, consistent with the formation of stem 1. NOEs between these protons and the H5 protons of C23 and C24 (see "Assignment of Nonexchangeable Protons" below), allowed distinction between the imino proton resonances. Resonances are not observed for imino protons of G1, U4, and U7, presumably due to exchange with solvent. Cross-peaks due to some amino proton resonances were visible in the NOESY spectrum in H_2O and could be assigned from NOEs to imino protons.

The imino proton region of the spectrum did not change substantially upon addition of $MgCl_2$ to a total concentration of 5 mM. The peaks were broadened, and the resonance assigned to the imino proton of G2 was shifted downfield by 0.2 ppm. There was little change, however, in the chemical shifts of any of the other resonances. This result indicates that the secondary structure does not change in response to magnesium ion at this concentration.

Assignment of Nonexchangeable Protons. Pyrimidine H5 and H6 resonances were identified by their strong cross-peaks in the double-quantum filtered COSY spectrum of the molecule. Cytosines were further distinguished from uridines by the chemical shifts of their C5 carbons, determined in a natural abundance $^1H-^{13}C$ HMQC (Varani & Tinoco, 1991a). This information provided a starting point for the assignment of nonexchangeable protons following standard procedures based on sequential NOE connectivities and through-bond correlations (Varani & Tinoco, 1991b). NOE connectivities were observed in a 400 ms mixing-time NOESY experiment with identifiable purine−pyrimidine patterns, leading to the sequential assignment of aromatic and H1′ protons for nucleotides G1 through C13 and A19 through C25. The aromatic and H1′ protons of U14 through G18 were assigned by comparison to chemical shifts from previous studies of that hairpin loop (Varani *et al.*, 1991). Overlapping H1′−H6 cross-peaks were separated in a third dimension on the basis of the through-bond correlation of the H5 and H6 protons in a homonuclear TOCSY−NOESY spectrum (Wijmenga *et al.*, 1994).

The 2′ protons were assigned by their cross-peaks with 1′ protons in a short mixing-time (60 ms) NOESY spectrum. These proton assignments were confirmed in the bulge loop by cross-peaks in the DQF-COSY spectrum and the TOCSY spectrum. The 2′ assignments in the stems were confirmed by sequential H6/H8−H2′ NOE connectivities in the short mixing-time NOESY spectrum. 3′ protons were identified from a $^1H-^{31}P$ correlated spectrum and a high-resolution

Table 1: Chemical Shifts (ppm) of Assigned Protons[a]

	H8/H6	H2/H5	H1′	H2′	H3′	H4′	imino[b]	amino[c]
G1	8.14	na[d]	5.84	4.93	4.72	4.56		
G2	7.58	na	5.94	4.74	4.68	4.45	12.88	7.08/8.30
G3	7.22	na	5.80	4.54	4.43	4.48	13.33	
U4	7.56	5.14	5.56	4.58	4.36	4.42		na
A5	7.95	6.87	5.73	4.45	4.13	4.58	na	
A6	7.74	7.86[e]	5.48	4.08	4.41	4.12	na	
U7	7.34	5.33	5.55	4.05	4.40	3.98		na
A8	8.13	7.80[e]	5.80	4.73	4.44		na	
A9	8.11	7.99	5.85	4.82	4.65	4.60	na	
G10	7.65	na	5.45	4.64		4.50	11.27	6.53
C11	7.81	5.51	5.52	4.30	4.54		na	7.06/8.60
U12	8.01	5.45	5.63	4.60	4.45	4.53	14.18	na
C13	7.80	5.67	5.60	4.46	4.33	4.47	na	6.97/8.47
U14	7.79	5.76	5.55	3.75	4.54	4.36	11.70	na
U15	8.03	5.87	6.10	4.66	4.01	4.48	10.92	na
C16	7.68	6.13	5.96	4.09	4.48	3.78	na	
G17	7.86	na	5.96	4.85	5.64	4.40	9.83	
G18	8.31	na	4.46				12.53	6.34/8.22
A19	7.78	7.58	6.01	4.67	4.50		na	6.65/8.02
G20	7.16	na	5.69	4.42	4.74	4.50	13.63	6.30/8.42
U21	7.71	5.49	5.94	4.43	4.75		12.31	na
A22	8.37	7.36	6.04	4.61	4.27		na	
C23	7.58	5.21	5.45	4.30	4.40	4.48	na	6.92/8.42
C24	7.80	5.46	5.54	4.28	4.50	4.39	na	6.89/8.52
C25	7.69	5.53	5.78	4.02	4.19	4.16	na	6.55/8.28

[a] Assignments of nonexchangeable protons are at 25 °C. All chemical shifts are reported relative to TSP. [b] Assignments of imino protons are at 5 °C. [c] Assignments of amino protons are at 10 °C. [d] na, not applicable. [e] Assigned during structure calculation.

COSY spectrum. Assignments of H3′ were made from COSY 2′−3′ peaks and sequential H6/H8−H3′ NOE connectivities. H4′ assignments were made from the H1′−H2′/H3′/H4′/H5′/H5″ region of a NOESY spectrum with a mixing time of 150 ms. These assignments were confirmed from the three-dimensional homonuclear TOCSY-NOESY spectrum and H3′−H4′ cross-peaks in high-resolution COSY data.

Adenosine H2 resonances were identified by the chemical shifts of bound carbons, determined in a natural abundance $^{1}H-^{13}C$ HMQC. Assignments were made for H2 of A5, A19, and A22 on the basis of the standard NOEs to 1′ protons observed in A-form double helices. Another resonance was assigned to the H2 of A9 based on a strong NOE to H1′ of G10 and a very weak NOE to its own H1′. Of the remaining H2 resonances, one had an NOE to the H1′ of A9 and no NOEs were observed for the other. These resonances were assigned in the course of structure calculation on the basis of consistency with the other experimental constraints. The proton assignments are summarized in Table 1.

Conformational Features of the Bulge Loop Determined by NMR. Stems 1 and 2 are A-form double-helices as judged by the relative intensities of internucleotide NOEs at mixing times ≤150 ms (shown for H1′−H8/H6 in Figures 4A and 4B). NOEs consistent with A-form geometry include U4 and A22 in stem 1. Thus, base pairing between U4 and A22 is suggested, despite the absence of an imino proton resonance due to U4. NOEs between H1′, H2′, and H6 of U21 and H8 of A22 are weak (Shown for H1′−H8/H6 in Figure 4C), indicating distortion of the helix between those residues and supporting the secondary structure in which G10 and U21 are base paired.

Within the bulge loop, discontinuous stacking is indicated by the relative intensities of sequential internucleotide NOEs (Figure 4C). The relative intensities of sequential internucleotide NOEs between U4, A5, and A6 are suggestive of continuous base stacking similar to that in an A-form geometry. The internucleotide NOEs between A6 and U7 are weak, and the only internucleotide NOE observed between U7 and A8, even with a long mixing time (400 ms), is between H4′ of U7 and H8 of A8. Thus, base stacking within the bulge loop is discontinuous around U7. Sequential NOE connectivities are resumed between A8 and A9 and continue between A9 and G10, suggesting base stacking between these nucleotides.

The phosphorus spectrum indicates that there are no sharp turns in the backbone except those associated with the UUCG hairpin loop. The chemical shift of a phosphorus for which its alpha or ζ torsion angle is *trans* is downfield with respect to the chemical shift of a phosphorus for which those torsion angles are both *gauche* (Gorenstein, 1981). All but three of the phosphorus nuclei in the molecule resonate within a range of approximately 1 ppm. Those three resonances can be assigned by comparison to chemical shifts from previous studies of the UUCG hairpin loop (Varani *et al.*, 1991) to U15, G17, and C16.

Four of the nucleotides in the bulge loop have significant C2′-*endo* character. H1′−H2′ cross-peaks in the DQF-COSY spectrum for A6, U7, A8, and A9 indicate significant contributions of the C2′-*endo* conformation to the conformational equilibria of those nucleotides. An H1′−H2′ COSY cross-peak is also present for U21. The equilibrium percentage of each of these sugars in the C2′-*endo* conformation was estimated from the magnitude of the 1′−2′ scalar coupling with the relation proposed by van den Hoogen (1988). Table 2 lists the results.

Structure Calculation. Twenty structures were calculated using 160 torsion angle constraints and 218 distance constraints. Of the distance constraints, 157 were internucleotide distances and 61 were intranucleotide distances. Lower bound constraints of 5.0 Å account for 27 of the distance constraints. The five nucleotides of the bulge (A5, A6, U7,

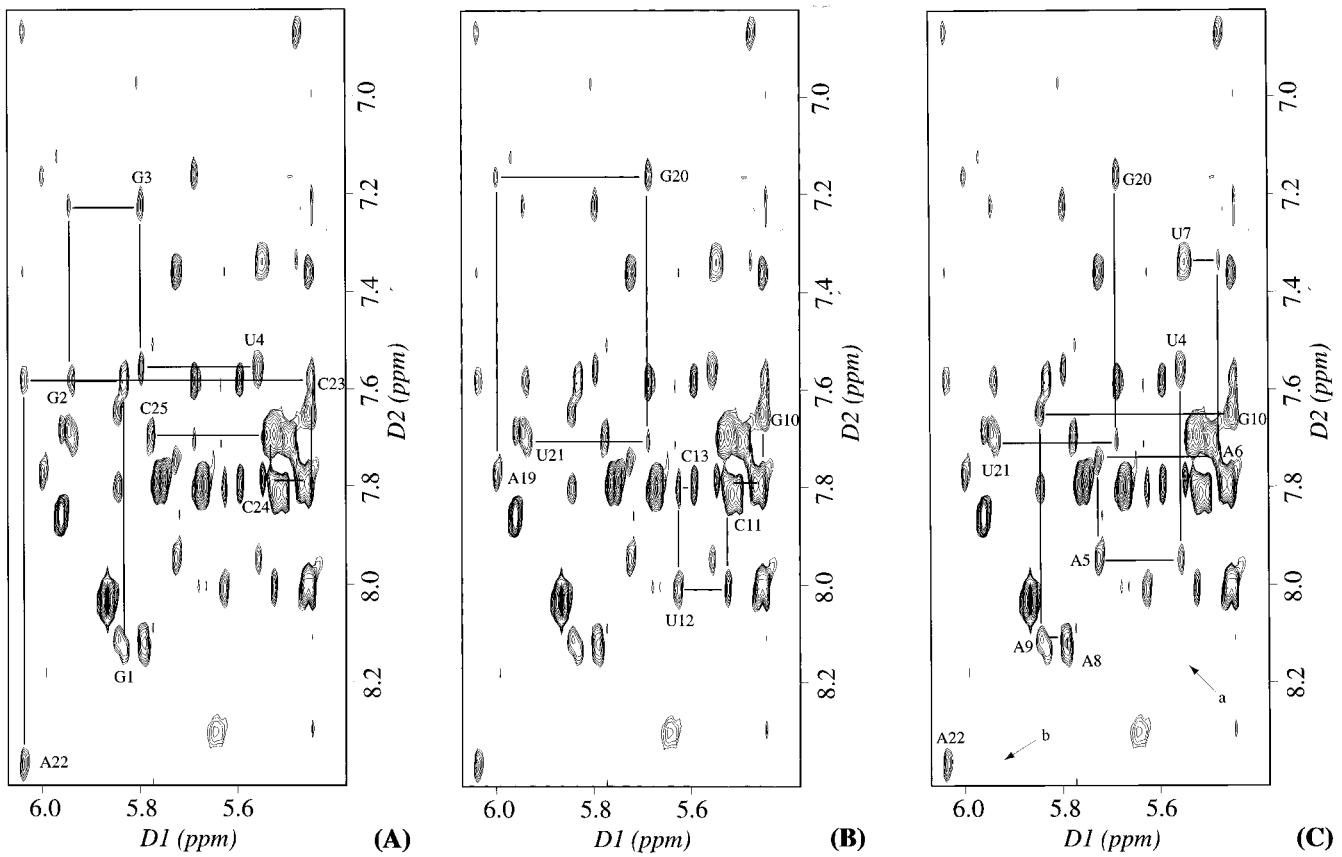

FIGURE 4: Portion of the NOESY spectrum showing NOEs between H8/H6/H2 (6.8−8.4 ppm) and H1′/H5 (5.4−6.1 ppm) protons. Data were collected on a GE GN-500 spectrometer. The mixing time was 150 ms. 300 FIDs of 1024 points were collected with a spectral width of 4000 Hz. 80 scans were averaged for each FID. The data were zero-filled to 1024 real points in the t_1 dimension and apodized in both dimensions with a skewed sine bell function shifted 30° and skewed by a factor of 0.7. Labeled NOE cross-peaks indicate NOEs from a nucleotide aromatic proton to the H1′ proton of its own sugar. (A) H8/H6−H1′ connectivity pathways for stem 1. (B) H8/H6−H1′ connectivity pathways for stem 2. (C) H8/H6−H1′ connectivity pathways within and around the bulge loop. Arrows indicate positions where cross-peaks between H1′ of U7 and H8 of A8 (a) and H1′ of U21 and H8 of A22 (b) would appear if present.

Table 2: Summary of Ribose Sugar Conformations

nucleotide[a]	% C3′-*endo* (±10%)[b]
G1	50
A6	50
U7	25
A8	25
A9	50
U15	<10
C16	<10
U21	20
C25	50

[a] Nucleotides not included in the table are >90% in the C3′-*endo* conformation. [b] Values were calculated from $J_{1'-2'}$ using the empirical equation of van den Hoogen (1988): %C3′-*endo* = 114.9 − 14.5($J_{1'-2'}$).

A8, and A9) and the four nucleotides of the flanking base pairs (U4, A22, G10, and U21) were constrained with a total of 69 internucleotide distances, 24 intranucleotide distances, and 59 torsion angles. The structures have a root mean square deviation (RMSD) of 2.1 Å. They can be divided into two families, distinguished principally by the orientation of A8 and A9 with respect to stem 2. In one of the families (family A, Figure 5), comprising sixteen of the structures, A8 and A9 are stacked on each other but bent away from stem 2. In the other family (family B, Figure 6), comprising four of the structures, A8, A9, and G10 are continuously stacked, making the stack of A8 and A9 colinear with stem 2 but requiring the G·U base pair at the end of stem 2 to buckle from planarity. The RMSD of the structures in family A alone is 1.3 Å, and the RMSD of the structures in family B alone is 1.0 Å. Features of both structural families are illustrated schematically in Figure 7.

The effect on the structure calculation of constraining G10 and U21 to form a G·U base pair was investigated. The structure was calculated using only distance constraints derived directly from NOEs for those two nucleotides. No assumptions about hydrogen bonding of G10 and U21 were included. Ten structures were calculated starting from ten different randomized structures. The ten calculated structures did not differ significantly from the structures of family A calculated with the G·U pair constrained, though they were determined with less precision (RMSD of 2.1 Å). In three of the ten structures, the G·U pair was apparent. In the other seven structures, G10 and U21 were too far from each other to imply base pair hydrogen bonds but were oriented with respect to each other as they would be in a base pair and had no other hydrogen bonding partners. The sharp peaks for G10 and U21 imino protons in the NMR spectrum indicate that those protons are protected from exchange with water by participating in hydrogen bonds, thus this calculation supports the inclusion of restraints to fix the G·U base pair in the final structure determination.

In addition to stacking of A8 and A9, several features are common to both structural families. A single non-Watson−Crick hydrogen bond is formed between O2 of U4 (acceptor) and N6 of A22 (donor). The average distance between donor

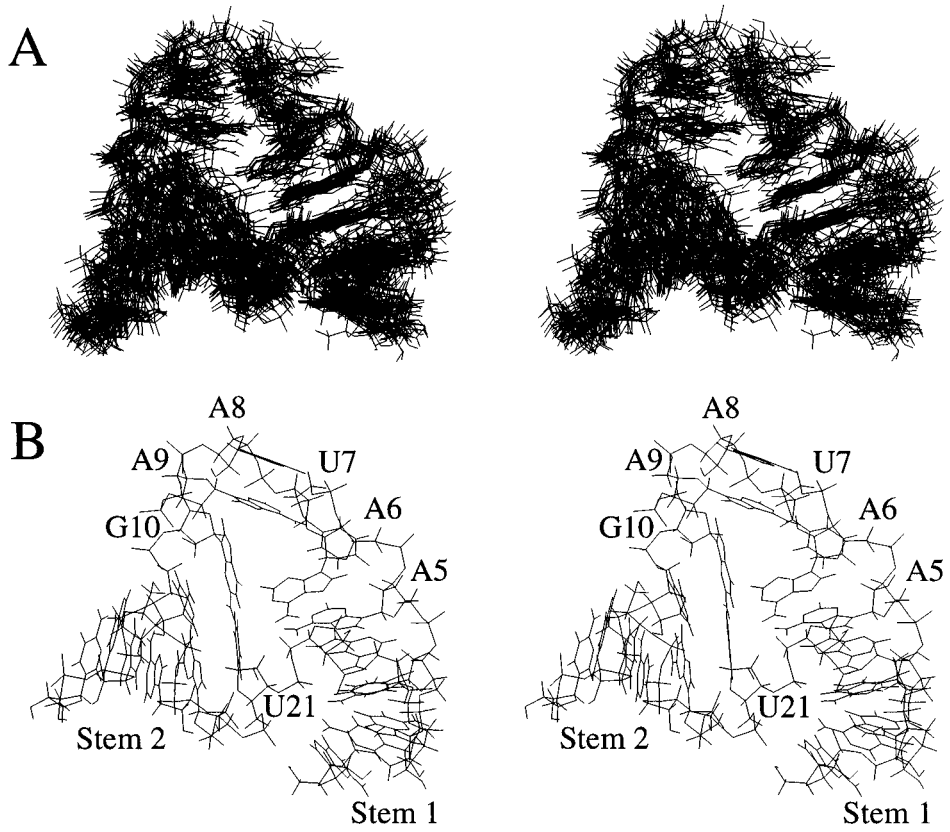

FIGURE 5: Stereoviews of structural family A: (A) 16 superimposed structures. (B) One representative structure from family A.

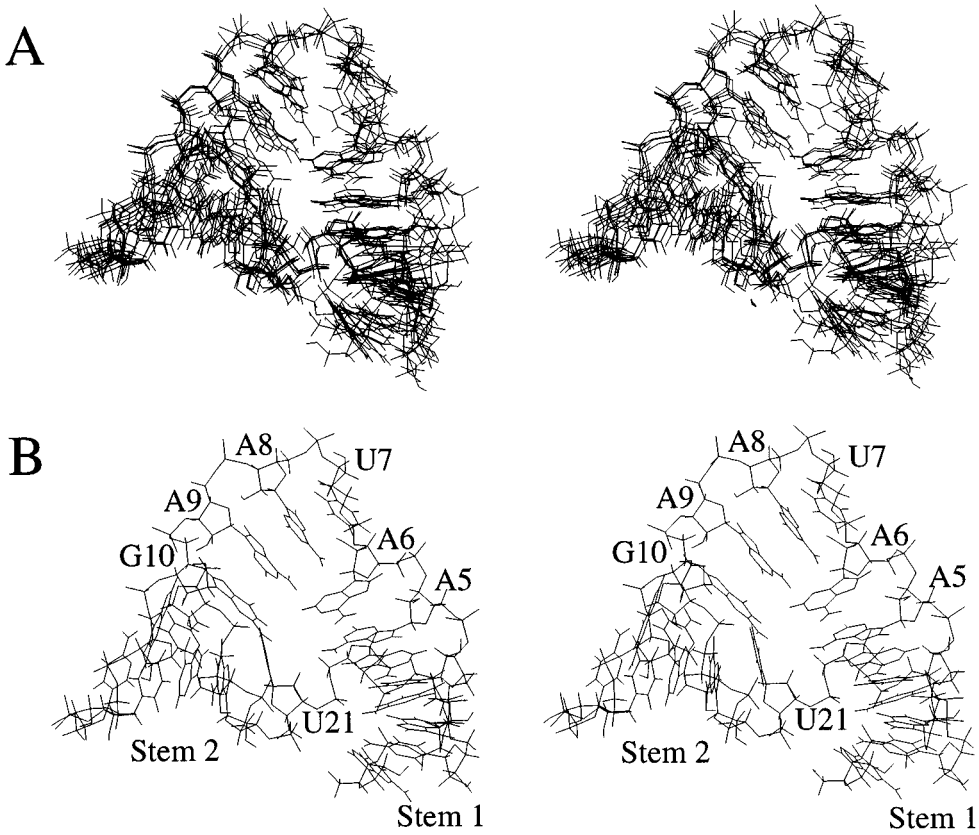

FIGURE 6: Stereoviews of structural family B: (A) four superimposed structures. (B) One representative structure from family B.

and acceptor in all 20 calculated structures is 2.5 Å. A5 and A6 are stacked on each other, and A5 is stacked cross-strand on stem 1 with A22. Few NOEs were observed to U7, and the conformation of that nucleotide is not well defined. There is no clear stacking interaction between U7 and the surrounding nucleotides. In all of the calculated structures, the bulge loop induces a bend between the flanking helices. The range of bend angles is 67−115°, the

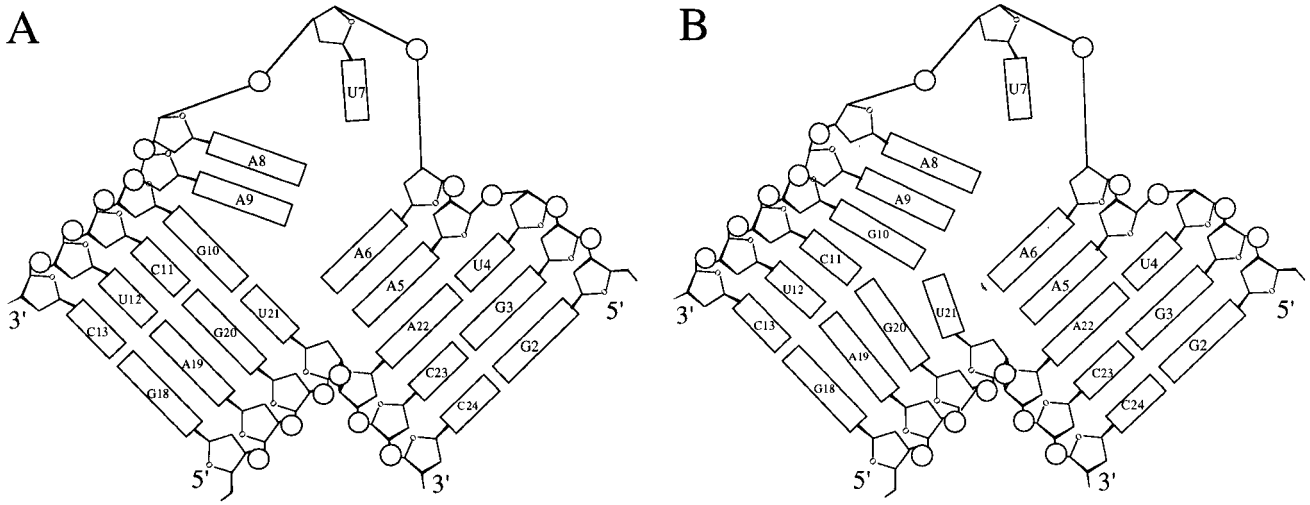

FIGURE 7: Schematic illustration of base stacking within and around the bulge loop. Rectangles represent the bases of the indicated nucleotides. Base stacking is indicated by adjacent, parallel rectangles. (A) Structural family A. (B) Structural family B.

mean being 90° (standard deviation of 14°). Bend angle was not correlated with calculated energy or structural family.

DISCUSSION

Secondary Structure. Our NMR data support the secondary structure shown in Figure 1C, in which a G·U base pair flanks a bulge loop of sequence 5′AAUAA3′. Sequential internucleotide NOEs and chemical shifts of the observed imino proton resonances are consistent with this structure and rule out the alternative secondary structures shown in Figure 1B and 1D. Though the observed secondary structure is predicted by the secondary structure prediction program of Zuker (1989), it is different from the secondary structure predicted by comparative sequence analysis of the bulge loop in the Group I introns (Figure 1B). Crystallographic data for the structural domain comprising stems P4, P5, and P6 of the Group I intron from *T. thermophila* confirm the secondary structure predicted for the bulge by comparative sequence analysis within the context of that domain (Cate *et al*., 1996). The insensitivity of the secondary structure of the bulged hairpin we studied to the presence of magnesium ion indicates that the difference between this structure and the structure of the bulge in the tertiary domain is not due to the presence of magnesium ions required to stabilize tertiary structure. Thus, the secondary structure of the isolated bulged duplex is different from the secondary structure of the bulge within the context of the higher order folding of the domain.

Loop Conformation. In several of the bulge loops of one to three nucleotides previously studied by NMR, all of the bulged nucleotides are either intercalated into the helical stack or extruded from the helix. Within the five-nucleotide bulge studied here, these limiting extremes are both represented. The adenosines within the bulge are at least partially stacked on the flanking helical stems, whereas discontinuity of sequential NOEs around the uridine suggests its extrahelicity. Discontinuous base stacking characterizes the loop conformation. The five-nucleotide bulge loop is thus similar to the three-nucleotide bulge loop from the TAR element of HIV, in which two of the bulged nucleotides are continuously stacked with the adjacent helix and one bulged nucleotide is partially or transiently extrahelical (Aboul-ela *et al.*, 1996). The five-nucleotide bulge is also like the three-nucleotide bulge loop from the TAR element in its transition to a distinct, largely extrahelical, conformation upon forming functional interactions: tertiary interactions with RNA in the intron's bulge and interactions with peptide ligand in the TAR bulge. Hydrogen bonding of bulge nucleotides with each other and with flanking nucleotides might be anticipated for large bulge loops. However, we find no evidence for hydrogen bonding interactions within the five-nucleotide bulge studied here.

The cross-strand stacking of nucleotide A5 on A22 and the non-Watson−Crick base pair between A22 and U4 are noteworthy features of this structure. A consequence of this arrangement is placement of A5H2 over the aromatic ring system of A22. Consistent with the expected shielding of that proton by the ring current of A22, the H2 of A5 resonates at a chemical shift (6.87 ppm) in the upfield extreme of the typical range for that type of proton. In this structure, the chemical shift of A5′s H2 is the most upfield of all of the H8, H6, and H2 protons (Table 1). Another consequence of this arrangement is proximity of A5′s H2 with A22′s H8. The mean distance between these protons in the 20 calculated structures is 3.77 Å (standard deviation of 0.37 Å). A weak NOE is observed between these protons in a NOESY spectrum with mixing time 400 ms, though due to its weakness, this NOE was not used in the structure calculation as the basis of a distance constraint. To the extent that the bulge loop is flexible, A5 is like a 3′ dangling end. Unpaired nucleotides at the 3′ ends of RNA double helices are known to have a stabilizing effect, and the sequence-dependence of this effect suggests that it is due to stacking of the unpaired nucleotide on the 5′ terminal base of the opposite strand (Freier *et al.*, 1985). The cross-strand stacking found in this structure is consistent with that suggestion.

The apparent stacking of A5 above A22 and the stacking of A22 and U4 on the adjacent base pair would be expected to protect the imino proton of U4 from exchange with water if it were involved in a Watson−Crick hydrogen bond with A22. However, in the non-Watson−Crick pairing apparent in the calculated structures, this proton is not involved in a hydrogen bond and is accessible to exchange with solvent. In fact, the imino proton of U4 is not visible in the NMR spectrum in H_2O. It is not surprising that no peak could be assigned to H6 (amino proton) of A22, because these protons

are often not observed, even when hydrogen bonded, due to rotation about the C6−N6 bond in adenosine at a rate comparable to the NMR time-scale. The molecular forces that result in a non-Watson−Crick base pair between A22 and U4 are not obvious, though destabilization of base pairs adjacent to bulge loops has been observed previously (Aboul-ela *et al.*, 1993, 1996). It is likely that the base pair configuration formed allows optimal stacking of A5 on A22 while concurrently allowing a backbone conformation between U21 and A22 that bridges the bulge loop.

The RMSD is not a useful measure of the precision of a structure arising from only four independent calculations, such as structures of family B. However, the large decrease in RMSD for both structural families when they are considered separately indicates that the structure calculation indeed generates two discrete structural families rather than a continuum between two extremes. Though examples of family A were arrived at more frequently by our modeling protocol, our data do not preferentially support either of the two structures. The significant contribution of both C2′-*endo* and C3′-*endo* sugar puckers to the conformational equilibria of the bulge loop ribose moieties indicates that the bulge conformation is not rigid, and both structural families might contribute significantly to the equilibrium ensemble of conformations.

Helix Bending. RNA bulge loops are known to induce local, directional bends between the helical segments that flank them (Bhattacharyya & Lilley, 1989; Bhattacharyya *et al.*, 1990; Hsieh & Griffith, 1989; Rice & Crothers, 1989; Riordan *et al.*, 1992; Tang & Draper, 1990, 1994; Wang & Griffith, 1991; Wang *et al.*, 1992; Gohlke *et al.*, 1994; Zacharias & Hagerman, 1995; Lilley, 1995). The bulge loop conformation that we observe requires such a bend. However, the short-range distance constraints provided by NMR data do not precisely define the extent of helix bending, and a range of angles consistent with the NMR-derived contraints is obtained. Electrophoretic experiments have been used to estimate the average bend angle induced by this bulge loop (Luebke & Tinoco, 1996). Under conditions of the NMR experiments described here, the bulge is found to induce a bend of approximately 90° (Luebke and Tinoco, unpublished data). This angle is identical, within experimental error, to the average of the bend angles obtained in structures calculated from the NMR data.

Comparison to Crystal Structure. Different loop conformations underlie the different secondary structures observed in this work and in the crystal structure of the bulge within the P4−P6 domain (Cate *et al.*, 1996). In the crystal structure, a Hoogsteen A·U base pair is formed between the adenosine corresponding to A9 and the uridine corresponding to U21. The adenosine in this base pair is in the *syn* conformation, whereas the relative intensities of intranucleotide NOE cross-peaks indicate that A9 is in the *anti* conformation within the hairpin context. Adenines corresponding to A8 and A9 are not stacked on each other in the crystal structure as they are in the NMR structure, and adenines corresponding to A5 and A6, while stacked on each other, are not stacked on the adjacent helix as they are in the NMR structure. Instead, they are looped out to make tertiary contacts. In the crystal structure of the tertiary domain, the helices flanking the bulge are nearly co-axial, whereas the bulge loop in the context of the hairpin induces a bend of approximately 90° between the flanking helices.

In a hierarchical view of RNA architecture (Westhof & Michel, 1992), tertiary structure is determined by the interactions of pre-formed elements of secondary structure. The P5a bulge, which alters its secondary structure in response to the formation of tertiary interactions, is an exception to the hierarchical model. The structure we observe for the isolated bulged duplex is thermodynamically favored in the absence of tertiary interactions, so tertiary interactions must energetically compensate for distortion of that structure. Moreover, the compensating tertiary interactions require the higher energy secondary structure found in the functional RNA for their formation. Distortion of the bulge conformation to that found in the crystal structure of the P4-P6 domain allows the formation of specific tertiary interactions by changing the orientation of nucleotides within the bulge and by changing the orientation of the helices surrounding the bulge. By changing the orientation of nucleotides within the bulge, hydrogen bond donors and acceptors within the loop are positioned for tertiary interactions. By changing the orientation of the helices surrounding the bulge from a bent to a co-axial arrangement, hydrogen bond donors and acceptors remote from the bulge, such as those in the L5b tetraloop (Cate *et al*., 1996), are positioned for concurrent tertiary interactions.

ACKNOWLEDGMENT

We thank Mr. David Koh for chemical synthesis of DNA and Ms. Barbara Dengler for managing the laboratory.

REFERENCES

Aboul-ela, F., Murchie, A. I. H., Homans, S. W., & Lilley, D. M. J. (1993) *J. Mol. Biol. 229*, 173−188.
Aboul-ela, F., Karn, J., Varani, G. (1995) *J. Mol. Biol. 253*, 313−332.
Aboul-ela, F., Karn, J., Varani, G. (1996) *Nucleic Acids Res. 24*, 3974−3981.
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.
Bhattacharyya, A., & Lilley, D. M. J. (1989) *Nucleic Acids Res. 17*, 6821−6840.
Bhattacharyya, A., Murchie, A. I. H., & Lilley, D. M. J. (1990) *Nature 343*, 484−487.
Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., & Doudna, J. A. (1996) *Science 273*, 1678−1685
Collins, R. A. (1988) *Nucleic Acids Res. 16*, 2705−2714.
Crothers, D. M., Haran, T. E., & Nadeau, J. G. (1990) *J. Biol. Chem. 265*, 7093−7096.
Davanloo, P., Rosenberg, A. H., Dunn, J. J., & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035−2039.
Flor, P. J., Flanegan, J. B., & Cech, T. R. (1989) *EMBO J. 8*, 3391−3399.
Freier, S. M., Alkema, A., Sinclair, T., Neilson, T., & Turner, D. H. (1985) *Biochemistry 24*, 4533-4539.
Gohlke, C., Murchie, A. I. H., Lilley, D. M. J., & Clegg, R. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11660−11664.
Gorenstein, D. G. (1981) *Annu. Rev. Biophys. Bioeng. 10*, 355−386.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870- 7872.
Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7456−7464.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283−300.
Hsieh, C.-H., & Griffith, J. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 4833−4837.

Joshua-Tor, L., Frolow, F., Appella, E., Hope, H., Rabinovich, D., & Sussman, J. L. (1992) *J. Mol. Biol. 225*, 397−431.
Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989) *Biochemistry 28*, 294−303.
Kalnik, M. W., Norman, D. G., Li, B. F., Swann, P. F., & Patel, D. J. (1990) *J. Biol. Chem. 265*, 636−647.
Lilley, D. M. J. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7140−7142.
Luebke, K. J., & Tinoco, I. (1996) *Biochemistry 35*, 11677−11684.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Marky, L. A., & Breslauer, K. J. (1987) *Biopolymers 26*, 1601−1620.
Michel, F., & Cummings, D. J. (1985) *Curr. Genet. 10*, 69−79.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 85-86.
Mills, J. B., Cooper, J. P., & Hagerman, P. J. (1994) *Biochemistry 33*, 1797−1803.
Moazed, D., Stern, S., & Noller, H. F. (1986) *J. Mol. Biol. 187*, 399−416.
Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I. (1983) *Biochemistry 22*, 5557−5563.
Morden, K. M., Gunn, B. M., & Maskos, K. (1990) *Biochemistry 29*, 8835−8845.
Murphy, F. L., & Cech, T. R. (1993) *Biochemistry 32*, 5291−5300.
Murphy, F. L., & Cech, T. R. (1994) *J. Mol. Biol. 236*, 49−63.
Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714-8725.
Peattie, D. A., Douthwaite, S., Garrett, R. A., & Noller, H. F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 7331−7335.
Puglisi, J. D., & Tinoco, I. (1989) *Methods Enzymol. 180*, 304−325.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76-80.
Rice, J. A., & Crothers, D. M. (1989) *Biochemistry 28*, 4512−4516.
Riordan, F. A., Bhattacharyya, A., McAteer, S., & Lilley, D. M. J. (1992) *J. Mol. Biol. 226*, 305-310.
Romaniuk, P. J., Lowary, P., Wu, H. N., Stormo, G., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 1563−1568.
Rosen, M. A., Shapiro, L., & Patel, D. J. (1992) *Biochemistry 31*, 4015−4026.
Sklenar, V., Miyoshiro, H., Zon, G., & Bax, A. (1986) *FEBS Lett. 208*, 94−98.
Tang, R. S., & Draper, D. E. (1990) *Biochemistry 29*, 5232−5237.
Tang, R. S., & Draper, D. E. (1994) *Nucleic Acids Res. 22*, 835−841.
van den Hoogen, F. (1988) Ph.D. Thesis, University of Leiden, The Netherlands.
van den Hoogen, Y. T., van Beuzekom, A. A., de Vroom, E., van den Elst, G. A., van Boom, J. H., & Altona, C. (1988a) *Nucleic Acids Res. 16*, 5013−5030.
van den Hoogen, Y. T., van Beuzekom, A. A., de Vroom, E., van der Marel, G. A., van Boom, J. H., & Altona, C. (1988b) *Nucleic Acids Res. 16*, 2971−2986.
Varani, G., & Tinoco, I. (1991a) *J. Am. Chem. Soc. 113*, 479−532.
Varani, G., & Tinoco, I. (1991b) *Q. Rev. Biophys. 29*, 479−532.
Varani, G., Cheong, C., & Tinoco, I. (1991) *Biochemistry 30*, 3280−3289.
Wang, Y.-H., & Griffith, J. (1991) *Biochemistry 30*, 1358−1363.
Wang, Y.-H., Barker, P., & Griffith, J. (1992) *J. Biol. Chem. 267*, 4911−4915.
Weeks, K. M., & Crothers, D. M. (1991) *Cell 66*, 577−588.
Westhof, E., & Michel, F. (1992) in *Structural Tools for the Analysis of Protein−Nucleic Acid Complexes* (Lilley, D. M. J., Heumann, H., & Suck, D., Eds.) Advances in Life Sciences, pp 255−267, Birkhauser, Boston.
Wimberly, B. G. (1992) Ph.D. Thesis, University of California, Berkeley, CA.
Wijmenga, S. S., Heus, H. A., Werten, B., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1994) *J. Magn. Reson. 103B*, 134−141.
Woodson, S. A., & Crothers, D. M. (1988a) *Biochemistry 26*, 904−912.
Woodson, S. A., & Crothers, D. M. (1988b) *Biochemistry 27*, 3130−3141.
Woodson, S. A., & Crothers, D. M. (1989) *Biopolymers 28*, 1149−1177.
Wu, H. N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221−8227.
Wyatt, J. R., Chastain, M., & Puglisi, J. D. (1991) *BioTechniques 11*, 764−769.
Zacharias, M., & Hagerman, P. J. (1995) *J. Mol. Biol. 247*, 486−500.
Zuker, M. (1989) *Science 244*, 48−52.

BI9701540