Conformation of an RNA Molecule That Models the P4/P6 Junction from Group I Introns†

Jacek Nowakowski 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 September 20, 1995; Revised Manuscript Received December 8, 1995⊗

ABSTRACT: We present a three-dimensional structure of a 34-nucleotide RNA molecule determined by NMR spectroscopy. The molecule was designed to form a junction between two double-helical stems whose sequence was based on the P4/P6 domain from group I introns. There are 5′ and 3′ single-strand overhangs at the junction of the stems. Contrary to our expectations, we found that the 3′ end of the molecule is placed in the minor and not the major groove of the P4 helix. As a result of tertiary contacts and stacking interactions from nucleotides in the 3′ end, the junction helices are rotated in a left-handed fashion and do not stack coaxially. This conformation is highly dependent on the presence of single-stranded nucleotides at the 3′ overhang. When the 3′ end is removed, the molecule assumes a radically different structure with the 5′ end in the minor groove of the P6 helix and an overall right-handed rotation between the stems. Only one nucleotide at the 3′ end is sufficient to change the geometry of the junction.

Junctions between two or more double-helical regions are a common motif in RNA structures. They are found in tRNAs (Holbrook et al., 1978) and the hammerhead ribozyme (Pley et al., 1994; Scott et al., 1995), and they have been proposed to form in a majority of known RNAs including ribosomal RNAs and self-splicing group I introns (Noller, 1984; Brunel et al., 1991; Kim & Cech, 1987; Michel & Westhof, 1990). Junctions are believed to play an important role in RNA folding by orienting double-helical regions in a specific fashion and acting as a structural scaffold for folded domains. Recent evidence shows that junctions are also important in RNA−protein interactions. For example, in the TAR RNA element from HIV-1, the junction between two stems created by a bulge forms a pocket for binding arginine (Puglisi et al., 1992), and a highly conserved ribosomal protein L11 binds to a three-way junction in 23S rRNA (Ryan & Draper, 1991).

In group I introns, the junction between P4 and P6 stems [nomenclature of Burke et al. (1987)] is an essential part of the catalytic core of the ribozyme (Beaudry & Joyce, 1990; Couture et al., 1990). Earlier studies based on biochemical data, phylogenetic comparisons, and model building predicted the junction to be in the form of adjacent triple helices (Michel et al., 1990; Michel & Westhof, 1990). In this model, P4 and P6 helices stack coaxially forming an elongated domain. Because of the right-handed rotation between the stems, the 3′ end leaving P6 forms base triples in the major groove of P4 and, similarly, the 5′ end entering P4 forms base triples in the minor groove of the P6 helix.

The Michel−Westhof model of the P4/P6 junction was recently supported in our laboratory by the structure of an RNA oligonucleotide containing both stems and a 5′ overhang. Chastain and Tinoco (1992, 1993) found that single-stranded nucleotides at the 5′ end formed tertiary interactions with the minor groove of the P6 helix. NMR[1] data on that molecule also showed that the stems were coaxially stacked and rotated in a right-handed fashion. As an extension of that work, here we present the structure of the junction with both 5′ and 3′ single-stranded overhangs (see Figure 1). Our hope was to find a simultaneous set of adjacent minor and major groove triples as proposed for the P4/P6 junction (Michel & Westhof, 1990). Such a combination of base triples had not been observed in any RNA structure, and we did not find one in the molecule shown in Figure 1. Instead, we found that the addition of the 3′ end to the P4/P6 junction model had an unexpected effect on its conformation. With the presence of only one nucleotide at the 3′ end, tertiary contacts between the P6 helix and the 5′ end did not form and the molecule assumed a very different conformation.

MATERIALS AND METHODS

The Design of P4/P6. Figure 1a shows the sequence of the junction which we named P4/P6. Stems corresponding to P4 and P6 are "capped" with stable tetraloops whose structures and NMR resonance assignments are known (Varani et al., 1991; Heus & Pardi, 1991b). Although the incorporation of tetraloops into the structure significantly increases the complexity of the NMR spectra, their presence assures formation of the double-helical stems and a single molecule avoids problems with alternate conformations of two- and three-strand models. The shaded region represents a sequence commonly occurring in group I introns which was chosen from the set of 87 aligned group I introns (Michel

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

* To whom correspondence should be addressed.

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

[1] Abbreviations: NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser and exchange spectroscopy; COSY, correlation spectroscopy; NOE, nuclear Overhauser effect; FID, free induction decay; RMSD, root mean square deviation; EDTA, ethylenediaminetetraacetic acid; rMD, restrained molecular dynamics; 1-D, one dimensional.

0006-2960/96/0435-2577$12.00/0

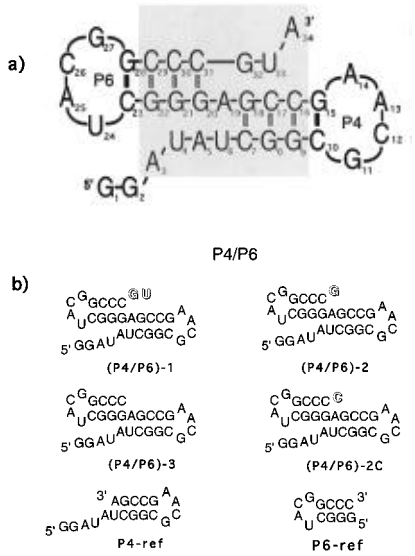


FIGURE 1: (a) Nucleotide sequence of P4/P6 showing the numbering system and naming convention for the stems. The shaded area represents the region of interest, whose sequence was derived from the P4/P6 junction of group I introns. (b) Names and sequences of P4/P6 mutants and reference molecules that are frequently used in the text. Outlined nucleotides represent the single-stranded 3′ ends.

& Westhof, 1990). The sequence of the 5′-end nucleotides was present in 18/87 introns, and the 3′-end nucleotides occurred 12/87 times. The entire sequence of the shaded region is present in three known group I introns. Residues A3 and A34 were included in the model because they are strongly conserved. The first two guanosines were added for the efficient synthesis of P4/P6 by T7 RNA polymerase.

Figure 1b shows the other molecules that were studied in order to analyze and understand P4/P6.

^{13}C-Labeled NTPs. Selectively ^{13}C-labeled nucleoside phosphates were synthesized chemically and enzymatically phosphorylated to nucleoside triphosphates by Dr. J. SantaLucia (SantaLucia et al., 1996). Guanine and adenine were labeled at C8 and cytosine and uracil at the C6 position.

RNA Synthesis and Purification. All RNA molecules were synthesized *in vitro* from nucleoside triphosphates using T7 RNA polymerase and were purified on denaturing (20%, v/w) polyacrylamide gels as described previously (Milligan et al., 1987; Wyatt et al., 1991). The yields of the transcription reactions ranged from 3 to 7 ODU/mL for various RNA sequences.

NMR Sample Preparation. Purified RNA oligonucleotides were dialyzed for 48 h against a total of 4 L of 10 mM KCl, 10 mM K_2HPO_4, and 0.1 mM EDTA, pH 6.6. After dialysis, 1 M $MgCl_2$ was added to the samples to a total Mg^{2+} concentration of 2 mM. Samples for exchangeable proton experiments were lyophilized and dissolved in 250 μL of 95% H_2O/5% D_2O. Samples for nonexchangeable proton experiments were lyophilized several times from D_2O and then suspended in 99.996% D_2O (Cambridge Laboratories). The RNA concentration was 2.5 mM for P4/P6 and 1.1−1.7 mM for other molecules. All samples were annealed once at 95 °C for 45 s and stored at 4 °C between NMR experiments.

NMR Spectroscopy. NMR spectra were collected on either a GE GN-500 or a Bruker AMX-600 spectrometer. Phase-sensitive spectra were acquired using the TPPI method (Marion & Wüthrich, 1983). All proton chemical shifts were referenced to the H_2O peak (4.8 ppm at 20 °C). The NMR data were processed and displayed with the program FELIX v. 2.31 (Biosym Technologies, Inc.).

1-D H_2O experiments were acquired at 5 °C using the 1-1 water suppression sequence (Plateau & Guéron, 1982) with the maximum excitation set at 12.0 ppm; 512 scans of 8K points and a sweep width of 12 000 Hz were used.

2-D H_2O NOESY experiments were done at 10 °C using a 1-1 suppression sequence which was optimized empirically by adjusting the length of the second pulse to be ~0.1 μs shorter than the first pulse. A total of 112 scans were taken for each FID with a 2.5 s relaxation delay between scans and a mixing time (t_{mix}) of 250 ms. A total of 420 FIDs of 4096 complex data points were collected with 12 000 Hz spectral width in both dimensions. The spectra were zero filled to 2K real points in t_1 and apodized in both dimensions with a 30° phase-shifted squared sine bell function.

2-D D_2O NOESY experiments on P4/P6 were done at 10, 20, and 25 °C with a t_{mix} of 300 ms to assign the resonances and resolve overlapped peaks. Additional data sets were acquired at 10 °C (t_{mix} = 50 ms) and 20 °C (t_{mix} = 40, 50, 80, 120, 150 ms) to quantitate the NOE volumes. The D_2O NOESY experiments on P4-ref, P6-ref, (P4/P6)-1, and (P4/P6)-3 were done at 20 °C with t_{mix} = 300 ms. In all experiments the HDO peak was suppressed by preirradiation during the 2.5 s relaxation delay. Approximately 360−490 FIDs, 32−64 scans each, of 2K real data points were recorded with a spectral width of 5000 Hz. The spectra were zero filled to 2K real points in t_1 and apodized in both dimensions using a skewed sine bell function with a 30° phase shift and a skew factor of 0.7.

Double-quantum-filtered (DQF) COSY experiments were done at 20 °C without phosphorus decoupling. The HDO peak was suppressed by preirradiation during the relaxation delay. A total of 512 FIDs of 1K real points were acquired with 32 scans for each FID. The spectra were zero filled to 1K real points in t_1 and apodized in both dimensions using a skewed sine bell function with a 10° phase shift and a skew factor of 0.7.

$^1/_2$-X-filtered NOESY experiments were done on selectively ^{13}C-labeled P4/P6 using the pulse sequence described previously (Otting & Wüthrich, 1990; SantaLucia et al., 1996). The ^{13}C carrier was set in the middle of the base carbon chemical shift range at ~140 ppm. A WALTZ-16 sequence was used for the ^{13}C−^{1}H decoupling in t_2. A total of 610 FIDs were collected with 32 scans each and 2048 complex points. Two data sets were collected for each t_1 increment, once with and once without the ^{13}C editing 180° pulse prior to the t_1 evolution. The sum of these two sets produced the subspectrum of NOEs from ^{12}C-bound protons, and the difference produced the subspectrum of NOEs from ^{13}C-bound protons. Both subspectra were zero filled to 1K real points in t_1 and were apodized in both dimensions with a skewed sine bell function with a 30° phase shift and a skew factor of 0.7.

Total correlation spectroscopy (TOCSY) spectra were acquired at 20 °C with a mixing time of 82 ms. A total of 398 FIDs, 32 scans each of 1024 real points, were collected. The suppression of the residual HDO peak was accomplished with preirradiation during the relaxation delay of 1.5 s. Spectra were zero filled to 1K real points in t_1 and apodized in both dimensions with a skewed sine bell function with a 10° phase shift and a skew factor of 0.7.

Distance and Torsion Angle Constraints. Distance constraints used in structure determination were derived from

experimentally determined NOE intensities at several mixing times. Depending on the intensity of the cross peak, the constraints were divided into four groups: strong, 1.8−3.0 Å; medium, 2.0−4.0 Å; weak, 3.0−5.0 Å; extra weak, 3.0−6.0 Å. Strong cross peaks were the most intense at t_{mix} = 50 ms, and extra weak peaks could only be detected at t_{mix} = 150 ms. A total of 86 experimental NOE constraints (4.7 constraints/residue) were used during the molecular dynamics calculations. Four distance constraints per base pair were used to maintain planarity of the base pairs and proper hydrogen bonding. These constraints were set to ±0.1 Å and were used only if an imino resonance from the base pair could be detected and assigned to a particular pair.

The ribose conformation was determined on the basis of the value of the coupling constant between H1′ and H2′ protons, measured in the DQF COSY experiment. Sugars with no observable H1′-H2′ couplings were constrained to the C3′-*endo* conformation with endocyclic torsion angles η_0, η_1, η_2, and η_3. Nucleotides with detectable couplings of 4−6 Hz were constrained to a broad pucker range encompassing C2′-*endo*, C3′-*endo*, and O4′-*endo* conformations.

The glycosidic angle χ was constrained to be *anti* for all nucleotides used in molecular dynamics protocols, based on the lack of a strong H1′−H8/H6 intranucleotide NOE. Additional torsion angle constraints were used to keep the amino groups in the plane of aromatic rings. During the refinement calculations, constraints for backbone torsion angles α, β, γ, ϵ, and ζ were used for nucleotides G8, G9, C16, C17, G20−G22, and C29−C31 to make them A-form. This assumption was based on the H1′−H2′ couplings and NOE connectivities between these nucleotides, which were consistent with the standard A-form geometry.

Molecular Dynamics. The three-dimensional structures of P4/P6 were generated using the molecular dynamics program X-PLOR (Brünger, 1990). Only nucleotides 5−9, 16−22, and 29−34 were used during the calculations to reduce computation time. Electrostatic energy terms were excluded from the force field at all times to avoid biasing the results. Twenty starting structures were created within the program with randomized backbone torsion angles and subjected to the global fold protocol. This stage of structure calculation consisted of 500 cycles of initial energy minimization, followed by 15 ps of restrained molecular dynamics (rMD) at 1000 K. Repulsive forces were initially turned off so that atoms were free to pass through each other and then gradually increased as the system was slowly cooled to 300 K. The calculation was concluded with 1000 cycles of final energy minimization. Ten structures generated by the global fold protocol were discarded because of either incorrect secondary structure or obvious violations of NMR constraints. Ten remaining structures were subjected to the refinement protocol. This stage began with 500 steps of initial energy minimization followed by 1 ps of rMD at 1000 K. At this point backbone torsion angle constraints were introduced in two steps, β, γ, and ϵ first and α and ζ second, and the system was slowly cooled to 300 K. The refinement was concluded with 2000 steps of final energy minimization. The reported RMSD values were calculated from the differences between individual structures and the calculated average structure.

RESULTS

Assignment Strategies. Exchangeable protons of P4/P6 were assigned from the H_2O NOESY experiment at 10 °C.

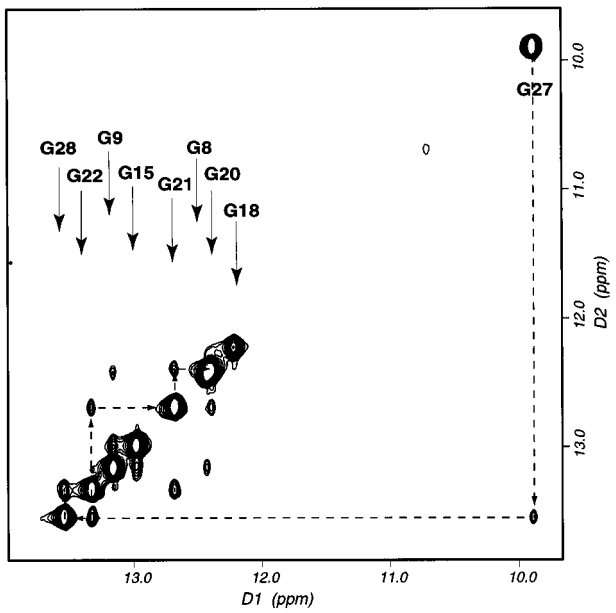


FIGURE 2: Portion of the H_2O NOESY spectrum showing NOEs between imino protons in P4/P6. Eight sharp resonances in the 12.0−13.5 ppm region confirm the formation of two double-helical stems each containing four G·C base pairs. The imino peak at 9.95 ppm from the G27·U24 pair is diagnostic of the UACG loop formation. Dashed lines connect resonances from the P6 helix. The spectrum was acquired in 10 mM KCl, 10 mM potassium phosphate, 2 mM $MgCl_2$, and 0.1 mM EDTA, pH = 6.6 at 10 °C.

The chemical shift of the G27 imino proton at 9.95 ppm (Varani et al., 1991) served as a starting point for sequential assignments of the P6 stem (Figure 2). Four remaining imino peaks were assigned to the P4 stem on the basis of the NOEs to other imino peaks and nonexchangeable protons. Amino protons were assigned from their NOE patterns to imino peaks characteristic for A-form RNA helices (Heus & Pardi, 1991a).

^{13}C-edited, $^1/_2$-X-filtered NOESY data on selectively labeled P4/P6 proved invaluable in identifying each aromatic resonance by nucleotide type. The results of this experiment are shown in Figure 3. Full-length P4/P6 was selectively labeled with ^{13}C at C8 of all adenines and at C6 of all cytosines. The ^{13}C subspectrum (Figure 3a) displays the NOE signals originating only from ^{13}C-bound protons, whereas the spectrum in Figure 3b contains NOEs from all ^{12}C-bound protons. In effect, the D2 dimension of the ^{13}C subspectrum contains H8 protons of A's and H6 protons of C's, and the ^{12}C subspectrum contains H8 of G's, H6 of U's, and H2 of A's. Pyrimidine's H6 protons are easily distinguished from the rest by the presence of strong H5−H6 cross peaks, which also can be detected in a DQF COSY experiment. The remaining ambiguity between guanine H8's and adenine H2's was resolved by acquiring another ^{13}C-edited NOESY data set on P4/P6 with selectively labeled guanines (data not shown). In this case resonances that appeared in the ^{13}C subspectrum came only from guanine H8 protons.

Once the identity of each aromatic resonance was established, the assignments of nonexchangeable protons in the helix regions of P4/P6 followed standard procedures based on sequential NOE connectivities and through-bond COSY correlations (Varani & Tinoco, 1991b). In order to facilitate the assignments and cross check the results, two smaller RNA oligonucleotides were synthesized whose sequence corresponded to the P6 stem (molecule P6-ref, Figure 1b) and

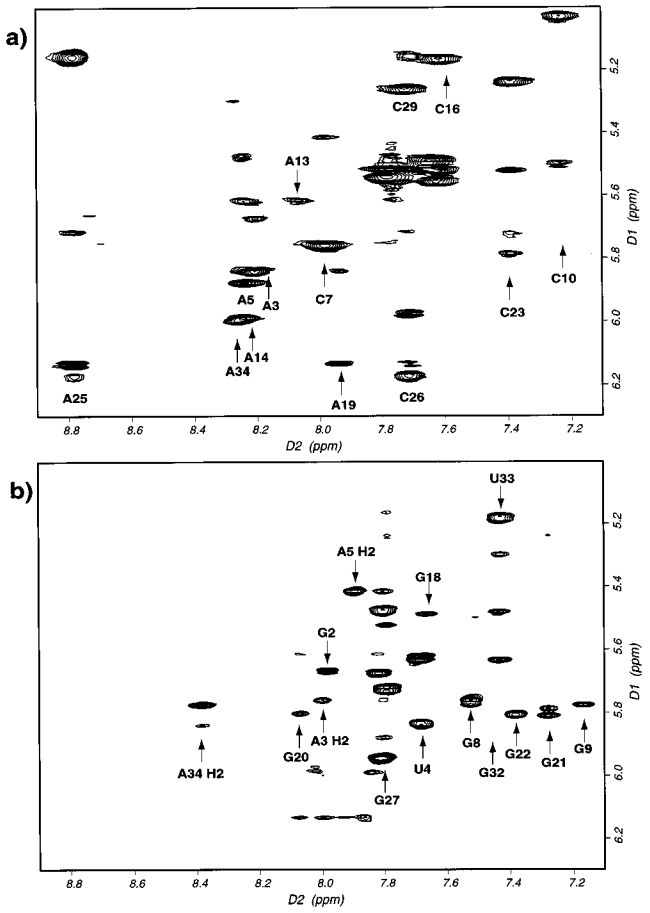


FIGURE 3: $^1/_2$-X-filtered NOESY spectra of P4/P6 showing the H8/H6/H2 to H1′/H5 region. P4/P6 was selectively ^{13}C labeled at H8 of adenines and H6 of cytosines. (a) ^{13}C subspectrum containing NOEs originating only from protons bound to ^{13}C (H8 of A's and H6 of C's), (b) ^{12}C subspectrum displaying NOEs from ^{12}C-bound aromatic protons (H8 of G's, H6 of U's, and H2 of A's). The analysis of the two subspectra allowed an unambiguous identification of each aromatic resonance by nucleotide type.

P4 stem (molecule P4-ref, Figure 1b). Starting with the known tetraloop data, resonances for these molecules were assigned separately and compared where possible with the full-length P4/P6. Any ambiguities caused by overlapped peaks were resolved by comparing NOESY spectra acquired at different temperatures. Several nucleotides displayed a significant amount of C2′-*endo* ribose conformation, which allowed the identification of H2′ and some H3′ and H4′ protons from the TOCSY experiment. The complete list of assignments is given in Table 1.

Structural Features Determined by NMR. (A) Secondary Structure. The P4/P6 molecule forms a unimolecular structure containing two double-helical stems capped with stable tetraloops. The formation of tetraloops is confirmed by the presence of several characteristic resonances including imino peaks of U24, G27, and G11 (Figure 8a), downfield-shifted H8 of A25 from the UACG loop, and upfield-shifted H1′ of G15 from the GCAA loop. There are a total of eight G·C base pairs formed, four in each stem. Unlike the P6 helix, which shows an uninterrupted walk between hydrogen-bonded iminos from G27 to G20, there is no NOE between G18 and G8 imino protons in the P4 stem, although both peaks are sharp and of equal intensity. This break is a result of a local distortion of the A-form geometry which is caused by a tertiary contact from the 3′ end of the molecule. The absence of any imino resonances in the 13.5−14.5 ppm region eliminates the possibility of forming the central

Table 1: Chemical Shifts (ppm) of Assigned Protons in P4/P6[a]

	H8/H6	H2/H5	H1′	H2′	H3′	H4′	imino	amino
G1	7.85	na[b]	5.70	4.70	4.42			
G2	8.00	na	5.69	4.79	4.45			
A3	8.22	7.98	5.87	4.69	4.76	4.50		
U4	7.71	5.67	5.85	4.36	4.64	4.46		
A5	8.27	7.94	5.92	4.77	4.57			
U6	7.82	5.23	5.47	4.43				
C7	8.00	5.81	5.75	4.68				
G8	7.57	na	5.76	4.57			12.43	
G9	7.24	na	5.20	4.29			13.16	
C10	7.31	5.07	5.23	4.37				8.20/6.65
G11	7.56	na	5.74	4.50	4.73		10.67	
C12	7.76	5.50	5.63	4.68	4.20			
A13	8.07	8.13	5.77	4.63	4.41			
A14	8.25	7.88	6.03	4.75	5.16			
G15	7.83	na	3.70	4.20			12.98	
C16	7.65	5.14	5.50					8.72/6.80
C17	7.68	5.56	5.51					8.32/6.82
G18	7.64	na	5.79	4.50			12.22	
A19	7.96	8.00	6.14	4.49	4.84			
G20	8.06	na	5.79	4.75			12.40	
G21	7.41	na	5.82	4.62			12.68	8.24
G22	7.29	na	5.79	4.55			13.33	8.60
C23	7.42	5.21	5.54					8.62/6.97
U24	7.81	5.68	5.53				11.86	
A25	8.79	8.30	6.15	5.16	4.15			
C26	7.73	6.19	5.99	4.14	4.51			
G27	7.83	na	5.96	4.85	5.63	4.42	9.95	
G28	8.30	na	5.62				13.54	8.92
C29	7.74	5.23	5.54					8.86/6.94
C30	7.80	5.49	5.53					8.59/6.92
C31	7.65	5.61	5.51	4.56				8.22/6.82
G32	7.49	na	5.65	4.37				
U33	7.46	5.24	5.37	3.90	4.13	4.38		
A34	8.27	8.27	6.00	4.58	4.23			

[a] Spectra were taken at 20 °C for nonexchangeable protons and at 10 °C for exchangeable protons. [b] na = not applicable.

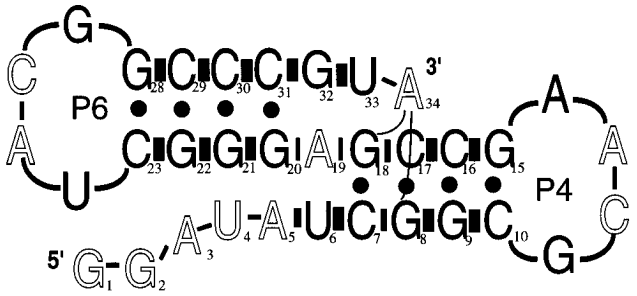


FIGURE 4: Summary of the secondary structure of P4/P6 based on NMR data. Dots represent hydrogen bonding between complementary base pairs. Black vertical bars indicate base stacking between nucleotides. The amount of stacking is proportional to the thickness of each bar. Nucleotides which have a significant amount of C2′-*endo* ribose pucker are printed in outline, and tertiary NOEs are marked with thin lines.

U6·A19 pair. This base pair, crucial for the overall structure of the junction, is formed when the 3′ end of the P4/P6 molecule is removed as discussed later in this section. The entire P6 and most of the P4 stem are predominantly in A-form geometry as judged by the intensities of intranucleotide NOE signals at short mixing times and by the C3′-*endo* ribose conformation. Base stacking in P4/P6 extends between nucleotides U6−C10, G15−G18, G20−C23, and G28−U33 with an obvious break at both the 3′ and the 5′ side of A19. The summary of the secondary structure is shown in Figure 4.

(B) Tertiary Structure. Contrary to initial expectations, no long-range NOEs were detected between the single-stranded 5′ end and the P6 stem. The first five nucleotides

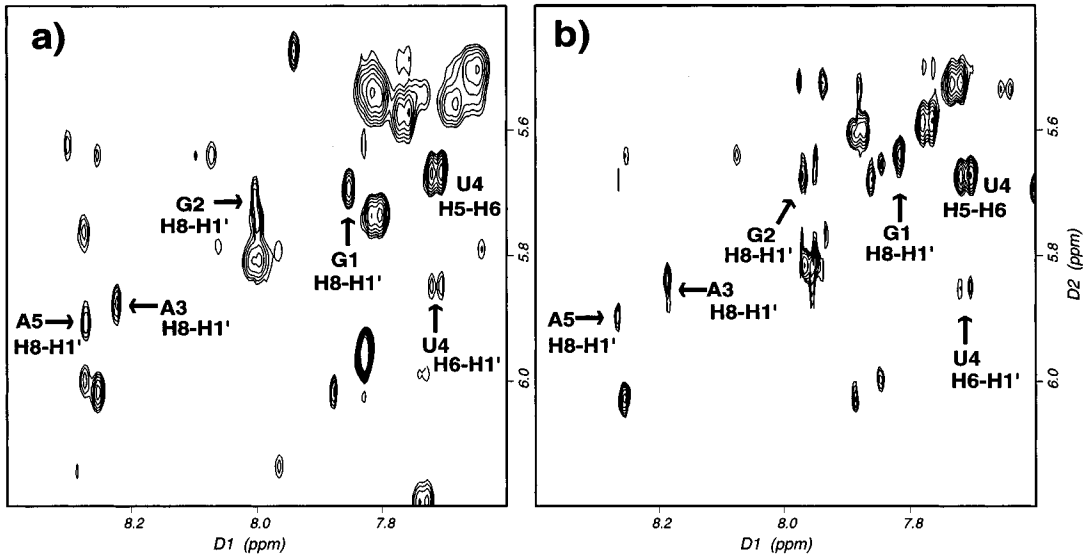


FIGURE 5: Comparison of chemical shifts and NOE patterns between the first five nucleotides of (a) P4/P6 and (b) P4-ref. The P4-ref molecule has the first 19 nucleotides of P4/P6 and forms a stem−loop structure corresponding to the P4 stem. The figure shows the portion of the NOESY spectrum displaying the NOEs between aromatic and H1′/H5 protons. NOEs originating from the first five nucleotides are labeled.

Table 2: Summary of the Ribose Sugar Puckers for P4/P6, (P4/P6)-3, and P4-ref

	% of C3′-*endo* sugar pucker ± 10%[b]		
nucleotide[a]	P4/P6	(P4/P6)-3	P4-ref
G1	<10	<10	<10
G2	<10	<10	<10
A3	<10	>90	<10
U4	<10	80	<10
A5	25	20	<10
C12	20	20	20
A13	30	20	20
A19	50	>90	50
A25	<10	<10	
C26	<10	<10	
A34	40		

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

of P4/P6 have a high percentage of ribose C2′-*endo* pucker, indicating flexibility of this region (Table 2). Residues G1 to U4 have no intranucleotide NOEs at short mixing times and make no tertiary contacts with either of the stems. A5 is partially stacked on U6, based on the NOEs from A5 H2 to U6 H1′ and A5 H1′ to U6 H6. Figure 5 compares NOESY data of P4/P6 with a reference molecule containing only residues 1 through 19 (P4-ref, Figure 1b). In this short oligonucleotide, the 5′ end is free to assume a wide distribution of conformations characteristic of a short single-stranded RNA chain. Since the aromatic, H1′ and H5 protons from the first five nucleotides in both molecules show identical NOE patterns and resonate at similar frequencies, we conclude that the 5′ end of P4/P6 also exists in a similar unstructured state. The conformations of the ribose groups of the 5′ end nucleotides in both molecules are also identical (Table 2).

The most significant structural determinants for P4/P6 came from the contacts between the 3′ end and the P4 stem. Two long-range NOEs, A34 H1′ to G18 H1′ and A34 H2 to G8 H1′, place the very last adenosine in the minor groove of the P4 helix. Because of the key importance of these constraints, an additional test was performed to verify the

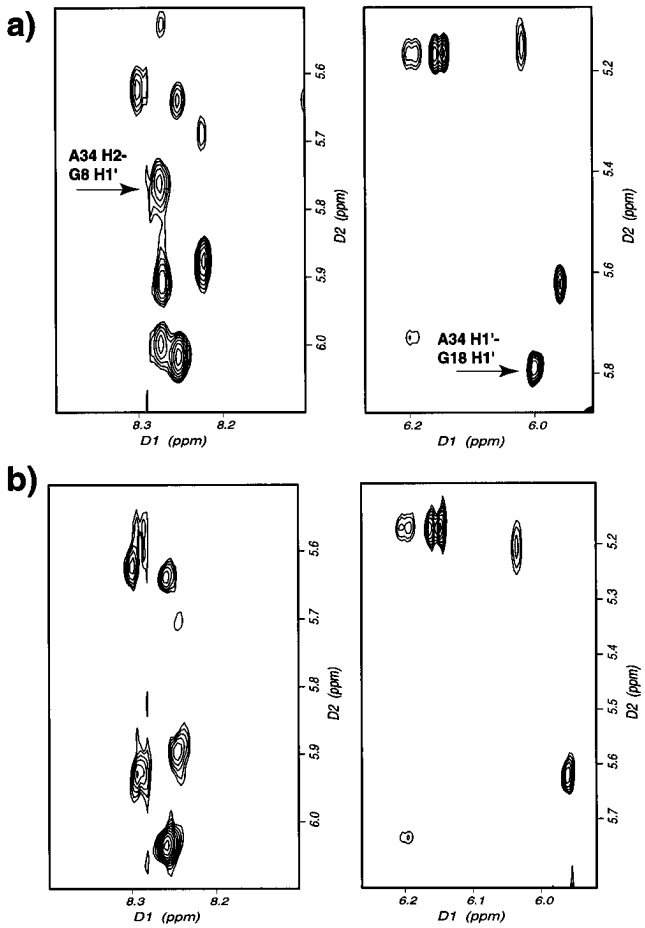


FIGURE 6: Regions of the NOESY spectra of P4/P6 (a) containing two crucial tertiary NOEs (marked with arrows). Lower panels (b) show the same regions of NOESY spectra of (P4/P6)-1, a mutant missing the A34 residue. Note that both of the crucial NOEs involving A34 are absent in the spectrum of the mutant.

identity of these NOEs. In Figure 6, we compare the crucial regions of NOESY spectra of P4/P6 (upper panels) with a molecule missing the last adenosine residue [(P4/P6)-1; lower panels]. As expected, both tertiary NOEs are absent in the spectrum of (P4/P6)-1. The remaining two nucleotides in the 3′ end, G32 and U33, have no long-range NOEs with

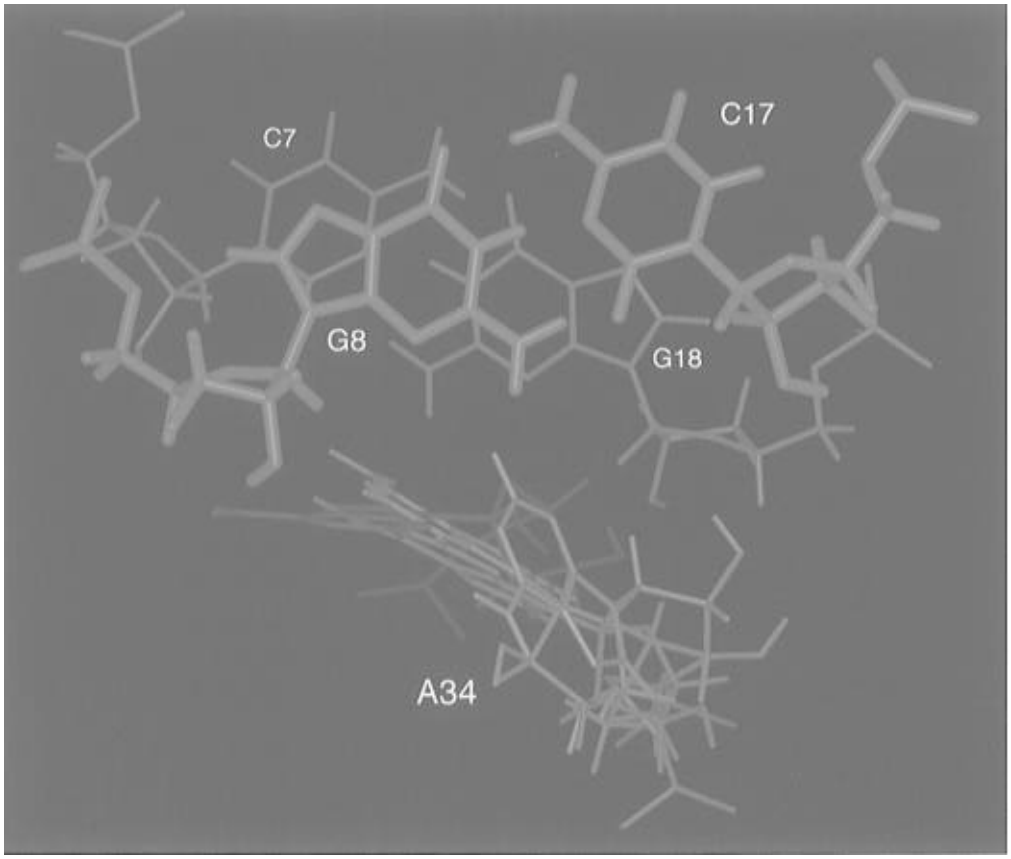


FIGURE 7: Superposition of four lowest energy structures showing the position of A34 with respect to the G8·C17 and C7·G18 base pairs of the P4 stem. To improve the clarity of the figure, only one set of base pairs is displayed. The view is down the P4 helix axis from the GCAA tetraloop side.

the P4 helix. Both of these nucleotides display a regular A-form NOE walk between H1′, H2′, and aromatic protons which extends to C31. At long mixing times, there are also NOEs between C31 H6−G32 H8 and G32 H8−U33 H6. These data suggest an uninterrupted stacking between C31 and U33. On the contrary, no intranucleotide NOEs were found between U33 and A34. Figure 7 shows the position of the last nucleotide with respect to the P4 stem as determined in four lowest energy structures obtained from molecular dynamics calculations. A34 is positioned in the minor groove of P4 with its ribose close to the sugar ring of G18. In the average structure, the 2′-hydroxyl groups of A34 and G18 are within hydrogen-bonding distance of each other, but the low number of structural constraints in this region can only suggest such interaction.

Previous NMR studies of the P4/P6 models showed that molecules missing the single-stranded 3′ end exist in a conformation radically different than that of P4/P6 (Chastain & Tinoco, 1992, 1993). In order to investigate the effect of the 3′ end on the structure of the junction, we synthesized a mutant molecule, (P4/P6)-3, which has the same sequence as P4/P6 but without the last three nucleotides. The junction sequence of this mutant is almost identical to the main variant molecule whose structure was previously determined in our laboratory (Chastain & Tinoco, 1993). On the basis of the similarity of the NOESY data of (P4/P6)-3 and the main variant, we concluded that these two molecules have the same structure.

Structural Differences between P4/P6 and (P4/P6)-3. The imino spectrum of (P4/P6)-3 has a clear resonance at 14.2 ppm (Figure 8d) indicating formation of the central A19·U6 base pair. The ribose of A19, which shows a mixed C2′/C3′-*endo* character in P4/P6, becomes fully C3′ *endo* in (P4/P6)-3. With the A19·U6 pair formed, the P4 helix is longer and there is no extra nucleotide between the two stems. Unlike in P4/P6, the NOE connectivities consistent with the A-form geometry (and base stacking) extend between residues G15 and C23 without interruption. Additional evidence for base stacking between A19 and G20 in (P4/P6)-3 comes from the chemical shift of G20 H8. In P4/P6, where these two residues are not stacked, G20 H8 resonates at a downfield location of 8.1 ppm. In (P4/P6)-3, a large ring current effect of stacked A19 shifts this resonance upfield to 7.1 ppm. Several tertiary NOEs including A5 H2−G20 H1′, A5 H2−A19 H2, and U4 H1′−G22 H1′ unambiguously place 5′ end nucleotides in the minor groove of the P6 stem. A strong NOE between A5 H1′ and U6 H1′ confirms the unusual upside down geometry of the A5 ribose seen in the structure of the major variant. The C3′-*endo* ribose conformation of nucleotides A3 and U4 is different from the same nucleotides in P4/P6 which assume a C2′-*endo* conformation (Table 2). The structural differences between P4/P6 and (P4/P6)-3 as detected by NMR are contrasted in Table 3.

Since the absence of the last three nucleotides in P4/P6 results in a large change in the conformation of the junction, it is important to identify the number and type of crucial nucleotides in the 3′ end necessary to maintain the P4/P6 geometry. The two conformations can be readily distinguished by the presence of the downfield-shifted imino peak from the central A19·U6 base pair. The presence of this resonance is indicative of a conformation similar to (P4/P6)-

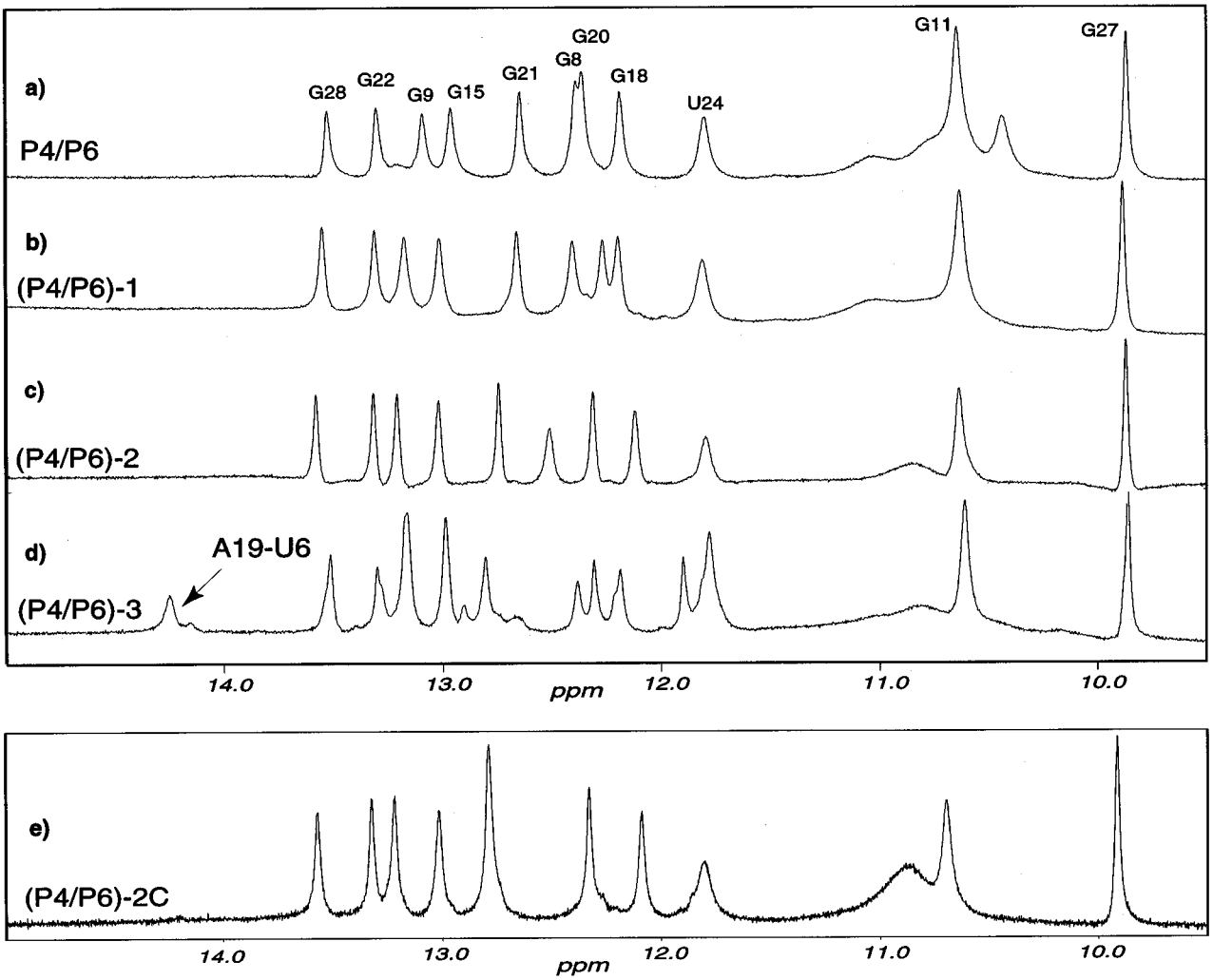
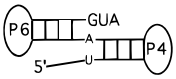
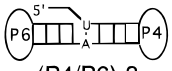


FIGURE 8: 1-D H_2O NMR spectra showing the imino regions of P4/P6 (a), (P4/P6)-1 (b), (P4/P6)-2 (c), (P4/P6)-3 (d), and (P4/P6)-2C (e). The presence of the downfield resonance at 14.2 ppm is indicative of A19·U6 base pair formation; this distinguishes between two conformations of the junction.

Table 3: Structural Differences between P4/P6 and (P4/P6)-3 As Determined by NMR

 P4/P6	 (P4/P6)-3
A19·U6 base pair not formed	A19·U6 base pair formed
5′ end is unstructured and makes no contacts with P6 stem	tertiary contacts between 5′ end and P6 stem in the minor groove: A5H2−G20H1′, A5H2−A19H2, A5H1′−U6H1′, U4H1′−G22H1′
ribose sugar pucker at 5′ end; G1−A5 are C2′ *endo* (mix)	ribose sugar pucker at 5′ end; G1, G2, and A5 are C2′ *endo* (mix) and A3 and U4 are C3′ *endo*
stacking between G15−G18 and G20−C23, interrupted at A19	continuous stacking from G15 to C23
tertiary contacts between 3′ end and P4 stem in the minor groove: A34H1′−G18H1′, A34H2−G8H1′	not applicable
chemical shift of G20 H8 at 8.1 ppm	G20 H8 at 7.1 ppm

3, whereas the absence points toward the P4/P6 conformation. Figure 8a−d displays imino regions of 1-D NMR spectra of the family of P4/P6 molecules with an increasingly shorter 3′ end. Molecules P4/P6, (P4/P6)-1, and (P4/P6)-2 have three, two, and one single-stranded nucleotides in the 3′ end, respectively (see Figure 1b). None of these mutants forms the central A·U base pair (Figure 8a−c). Since the number of base-paired imino resonances and their chemical shift patterns are the same, it is likely that these molecules have similar structures. The conformation of the junction is switched only when the entire single-stranded 3′ end is absent, as is the case in (P4/P6)-3 (Figure 8d). An additional mutant, (P4/P6)-2C, was designed to test whether a single nucleotide at the 3′ end of (P4/P6)-2 other than guanosine can maintain the P4/P6 conformation. We decided to make a purine to pyrimidine substitution in order to assure a large difference in both hydrogen-bonding and stacking properties of the last nucleotide. Uridine was not used because of the possibility of forming an alternate secondary structure (possible base pair A19·U32). When the single 3′ G32 residue of (P4/P6)-2 was changed to C, the junction still retained the P4/P6 geometry (Figure 8e).

DISCUSSION

Structure of the Junction. A representative three-dimensional structure of P4/P6 is shown in Figure 9. Helices

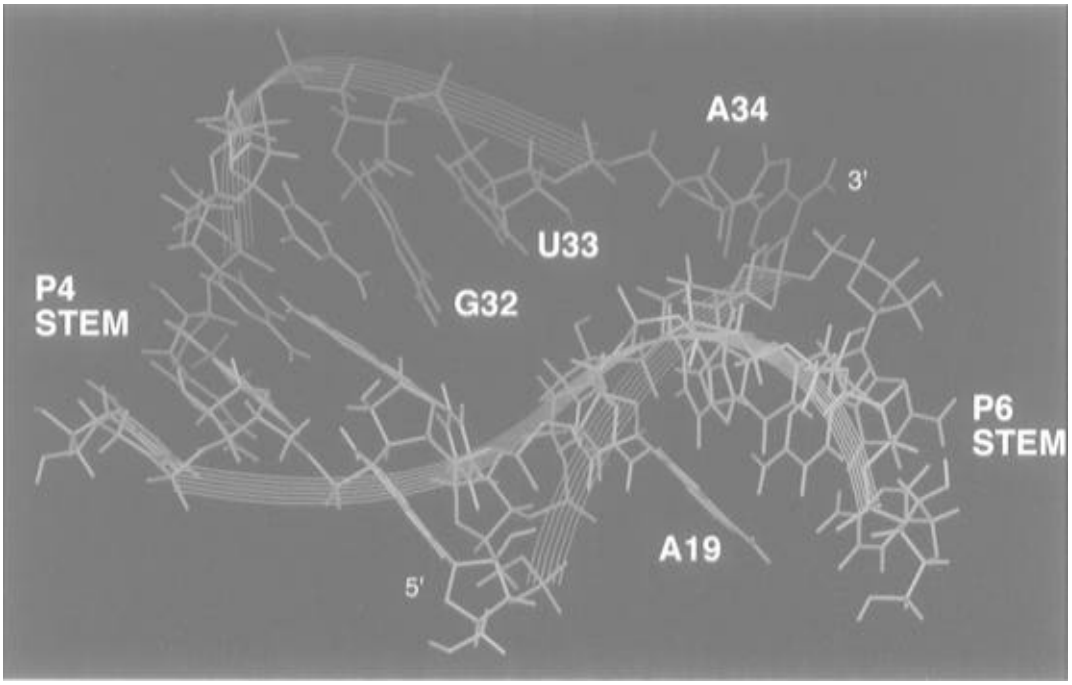


FIGURE 9: Representative three-dimensional structure of the P4/P6 junction: blue, P4 helix; green, P6 helix; red, 3′ end; yellow, 5′ end. The figure displays only nucleotides 5−9, 16−22, and 29−34.

P4 and P6 are bent with respect to each other and are not coaxially stacked. The twist at the junction of the two helices is left-handed. The central residue A19 is not involved in stacking interactions with either of the stems but acts as a flexible joint between them. In 8 out of 10 structures obtained from molecular dynamics calculations this nucleotide was bulged out of the junction. The 3′-end nucleotides G32 and U33 are stacked on the P6 helix and extend toward the minor groove of the P4 stem. The last residue, A34, is involved in tertiary contacts in the minor groove with G18 and G8, which cause a local distortion of the P4 helix. The average distance between the imino protons of G18 and G8 measured in ten low-energy structures was 4.6 Å. This distance is in agreement with the absence of the NOE between these protons in the H_2O NOESY experiment (Figure 2). The 5′ end of P4/P6 is dynamic and does not make any tertiary contacts with either of the stems. For simplicity, this part of the molecule is not displayed in Figure 9. Although the general geometry of P4/P6 is well-defined, the structure presented here is only a medium resolution one. The RMSD values for the superposition of ten low-energy structures are 2.37 Å for all atoms, 0.82 Å for the P4 helix (residues 7−9, 16−18), and 1.14 Å for the P6 helix (residues 20−22, 29−31). The left-handed twist angle at the junction of the stems (−40° ± 20°) and the bending between the stems (135° ± 20°) are not well-defined due to a low number of structural constraints in the junction region.

Our experiments with deletion mutants of P4/P6 have shown that the structure of the junction is dependent on the presence of single-stranded nucleotides at the 3′ end. We found that molecules containing two [(P4/P6)-1] and one [(P4/P6)-2] nucleotides at the 3′ end still retained the original conformation of P4/P6, although the A34 residue which formed tertiary contacts in P4/P6 was not present. The single nucleotide at the 3′ end of the (P4/P6)-2 mutant is not specific, since a guanine to cytosine substitution does not change the geometry of the junction. Once the entire 3′ single-stranded region is removed [(P4/P6)-3], the junction changes its conformation, assuming a right-handed rotation between

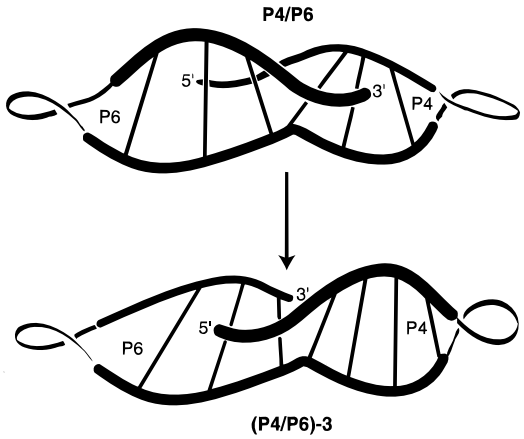


FIGURE 10: Drawing of the differences between P4/P6 and (P4/P6)-3 geometries.

the stems and forming tertiary contacts between the 5′ end and the P6 helix. Only one nonspecific nucleotide at the 3′ end is required to convert one conformation into the other. This rearrangement is graphically presented in Figure 10.

The change in junction geometry caused by a single nucleotide can be explained in terms of energetics of base stacking. Thermodynamic studies on short RNA duplexes showed that a single dangling nucleotide at the 3′ end can significantly increase duplex stability. For example, a 3′ dangling guanosine in the sequence 5′-GGCC-**G** -3′ (underlined nucleotides are base paired) decreases the free energy of duplex formation by $\Delta G^\circ_{37} = -1.7$ kcal/mol compared to the same sequence without the free end (Freier et al., 1983). The additional stability is the result of an enhanced base stacking between the dangling residue and the preceding base pair (Petersheim & Turner, 1983). In the case of P4/P6, (P4/P6)-1, and (P4/P6)-2 the enhanced stacking from the 3′-end nucleotides can be thermodynamically more favorable than stacking between two stems at the junction. In the structure of (P4/P6)-3, the two helices are stacked but the right-handed rotation between them is twice as large as the rotation between two base pairs in a A-form geometry

(Chastain & Tinoco, 1992). Such arrangement results in a poor overlap between the G20·C31 and A19·U6 junction base pairs. When the more favorable 3′-end stacking interaction takes place, there is no longer room for the coaxial stack of the stems, and the junction assumes an alternate structure. Even a single C residue at the 3′ end [(P4/P6)-2C] can provide enough stacking to change the junction conformation, although the stabilizing effect of 3′ dangling pyrimidines is smaller than for purines (Sugimoto et al., 1987).

We do not believe that the structure of P4/P6 represents the structure of the P4/P6 element in group I Intron for several reasons. Existing evidence clearly indicates a sequence covariation between the nucleotides proposed to form base triples in the junction (Michel et al., 1990; Green & Szostak, 1994). Our structure is not consistent with such data. In P4/P6, the 5′-end nucleotides, which are equivalent to the J3/4 stretch in the intron, are not involved in any tertiary interactions with the P6 helix, and their sequence is probably not relevant for the overall conformation of the model. The two nucleotides from the 3′-end, G32 and U33 which correspond to the proposed site of triples in the major groove of P4 helix, are involved in nonspecific stacking interactions. Also, the central A19·U6 base pair which is very conserved in group I introns is not formed in P4/P6. Finally, the only residue forming nucleoside triples in our model, A34, corresponds to the single, universally conserved adenine in all known introns (A261 in *Tetrahymena*). In the Michel−Westhof model, this nucleotide joins the main P4−P5−P6 and P7−P3−P8 domains by creating a sharp bend in the backbone. If the location of this residue were in the minor groove of the P4 helix as it is in P4/P6, stems P7 and P4 would be too close to avoid unfavorable repulsions between the phosphate groups. On the other hand, there is no experimental evidence for a right-handed rotation between P4 and P6 stems in the Michel−Westhof model. This particular arrangement is based on the assumption that an interrupted RNA helix would continue its natural right-handed twist at the break point. The structure of P4/P6 shows that RNA is capable of forming a junction with a left-handed rotation between two double-helical stems. Left-handed geometry would interchange the grooves of the proposed base triples, but it can still produce a model which satisfies the sequence specificity of the triples.

The results presented in this report show that a change in a single nucleotide unit can cause a large rearrangement of an RNA structure. There are a number of known systems where a single-nucleotide mutation in an RNA sequence drastically reduces or abolishes the biological function. Examples include the group I intron (Couture et al., 1990) and a frame-shifting pseudoknot (Chen et al., 1995) among many others. The conformational flexibility of RNA makes the interpretation of the mutagenesis data difficult since the mutation can change either the structure or the function, or both. Our findings emphasize that in order to understand the function of RNA in terms of structure it is crucial to assess the integrity of the structure after the mutation.

The large change in conformation with a single nucleotide change may also be helpful in understanding ribozyme function. The crystal structures (Pley et al., 1994; Scott et al., 1995) of a hammerhead ribozyme place several catalytically essential groups far away from the cleavage site. The catalytically active form is presumably very different. Thermal fluctuations involving energy differences of order 1 kcal mol^{-1} can cause large changes in RNA conformation.

ACKNOWLEDGMENT

We thank Dr. J. SantaLucia for kindly providing the selectively ^{13}C-labeled nucleotides and for his help with the design of the P4/P6 model and his assistance with NMR experiments. We also thank Mr. David Koh for the synthesis of DNA oligonucleotides and Ms. Barbara Dengler for managing the laboratory.

REFERENCES

Beaudry, A. A., & Joyce, G. F. (1990) *Biochemistry 29*, 6534−6539.
Brunel, C., Romby, P., Westhof, E., Ehresmann, C., & Ehresmann, B. (1991) *J. Mol. Biol. 221*, 293−308.
Brünger, A. T. (1990) *X-PLOR*: *A System for Crystallography and NMR, Version 2.1*, Yale University, New Haven, CT.
Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W., & Tabak, H. W. (1987) *Nucleic Acids Res. 15*, 7217−7221.
Chastain, M., & Tinoco, I., Jr. (1992) *Biochemistry 31*, 12733−12741.
Chastain, M., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 14220−14228.
Chen, X., Chamorro, M., Lee, S. I., Shen, L. X., Hines, J. V., Tinoco, I., Jr. & Varmus, H. E. (1995) *EMBO J. 14*, 842−852.
Couture, S., Ellington, A. D., Gerber, A. S., Cherry, J. M., Doudna, J. A., Green, R., Hanna, M., Pace, U., Rajagopal, J., & Szostak, J. W. (1990) *J. Mol. Biol. 215*, 345−358.
Freier, S. M., Burger, B. J., Alkema, D., Neilson, T., & Turner, D. H. (1983) *Biochemistry 22*, 6198−6206.
Green, R., & Szostak, J. W. (1994) *J. Mol. Biol. 235*, 140−155.
Heus, H. A., & Pardi, A. (1991a) *J. Am. Chem. Soc. 113*, 4360−4361.
Heus, H. A., & Pardi, A. (1991b) *Science 253*, 157−163.
Holbrook, S. R., Sussman, J. L., Warrant, R. W., & Kim, S.-H. (1978) *J. Mol. Biol. 123*, 631−660.
Kim, S. H., & Cech, T. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8788−8792.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967− 974.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
Michel, F., Ellington, A. D., Couture, S., & Szostak, J. W. (1990) *Nature 347*, 578−580.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119−162.
Otting, G., & Wüthrich, K., (1990) *Q. Rev. Biophys. 23*, 39−96.
Petersheim, M., & Turner D. H. (1983) *Biochemistry 22*, 256−263.
Plateau, P., & Guéron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994) *Nature 372*, 68−74.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76−80.
Ryan, P. C., & Draper, D. E. (1991) *J. Mol. Biol. 221*, 1257.
SantaLucia, J., Jr., Shen, L. X., Cai, Z., Lewis, H., & Tinoco, I., Jr. (1996) *Nucleic Acids Res. 23*, 4913−4921.
Scott, W. G., Finch, J. T., & Klug, A. (1995) *Cell 81*, 991−1002.
Sugimoto, N., Kierzek, R., & Turner, D. H. (1987) *Biochemistry 26*, 4554−4558.
van den Hoogen, F. (1988) Ph.D. Dissertation, University of Leiden.
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.
Wyatt, J. R., Chastain, M., & Puglisi, J. D. (1991) *BioTechniques 11*, 764−769.

BI952263P