Articles

A Base-Triple Structural Domain in RNA†

Michael Chastain and Ignacio Tinoco, Jr.*

Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, California 94720

Received June 30, 1992; Revised Manuscript Received October 14, 1992

ABSTRACT: An oligonucleotide modeled on a proposed base-triple domain of the *Tetrahymena* group I intron has been characterized by NMR. The oligonucleotide contains two double-helix regions with adjacent single-stranded nucleotides. The NMR data show that the two helices stack coaxially, although the rotation between the two helices is approximately twice as large as the rotation between normal base pairs. The rotation between the two helices allows the single-stranded nucleotides to form U·U·G and A·G·C base triples in the minor groove. The A·G·C base triple contains a hydrogen bond between the adenine N_1 and a 2′-hydroxyl in the minor groove of the G·C pair. A similar hydrogen bond between an adenine and a 2′-hydroxyl in transfer RNA suggests that this could be a recurring tertiary interaction in RNA.

RNA molecules fold into structures which bind nucleotides (Michel et al., 1989), amino acids (Yarus, 1988), and nucleic acid helices (Doudna & Szostak, 1989). RNA molecules probably catalyze processes such as messenger RNA splicing (Guthrie, 1991) and translation (Noller et al., 1992). Base-pair formation in RNA is fairly well characterized, but very little is known about the folded three-dimensional structures of RNA molecules. An important step toward understanding the biologically active three-dimensional structures of RNA molecules is characterizing the tertiary interactions which occur between base-paired and single-stranded regions of RNA. Tertiary interactions impose strong constraints on the three-dimensional structures of RNA molecules.

One type of tertiary interaction is the base triple. Base triples occur when a single-stranded nucleotide forms hydrogen bonds with a nucleotide involved in a base pair. Three base triples occur in the crystal structure of yeast $tRNA^{Phe}$ at the junction between the anticodon helix and the D helix in the major groove of the D helix (Holbrook et al., 1978). A base triple in the structure of the TAR bulge loop complexed to arginine has been characterized by NMR[1] (Puglisi et al., 1992). Several base triples are proposed to occur in the catalytic core of the *Tetrahymena* group I intron on the basis of phylogenetic comparison (Michel & Westhof, 1990; Michel et al., 1990). Four such base triples are proposed at the junction of helices P4 and P6 (Figure 1). U·U·G and A·G·C base triples are proposed to form in the minor groove of helix P6 and U·C·G and C·G·C base triples are proposed to form in the major groove of P4. In both the group I intron and tRNA, base triples occur, or are proposed to occur, at junctions between two helices flanked by single-stranded nucleotides (Figure 2).

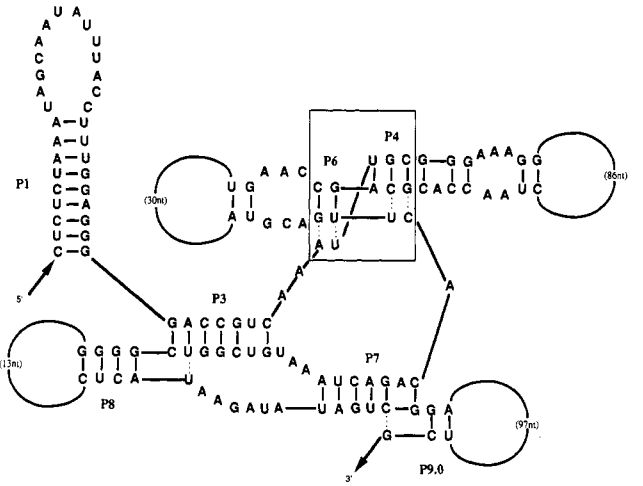

FIGURE 1: Secondary structure of the catalytic core region of the group I self-splicing intron. The tertiary interactions shown in dashed lines are proposed by Michel and Westhof (1990). The boxed region containing the P4/P6 helices and adjacent single-stranded nucleotides proposed to form base triples was used to design the oligonucleotide characterized in this study.

We have synthesized a 25-nucleotide RNA to determine if two helices with adjacent single-stranded nucleotides are sufficient to form base triples (Figure 3). The sequence is modeled after the P4/P6 region in the group I intron. The single-stranded nucleotides proposed to form the A·G·C and U·U·G base triples in the minor groove of helix P6 were included but not the nucleotides proposed to form base triples in the major groove of helix P4. Helix P6 was lengthened so that three A·G·C base triples could form consecutively. A second molecule, which contains the same helix and loop regions as the P4/P6 model, but not the single-stranded nucleotides proposed to form base triples, was synthesized as a reference. Nuclear magnetic resonance (NMR) techniques were used to determine the structure of these oligonucleotides, and the molecular dynamics program XPLOR (Brünger, 1990) was used to generate structures consistent with the NMR data.

† This research was supported in part by National Institutes of Health Grant GM 10840, by the Department of Energy, Office of Energy Research, Office of Health and Environmental Research, under 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 and BBS 87-20134. M.C. is a Howard Hughes Medical Institute Doctoral Fellow.

[1] Abbreviations: NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; COSY, correlated spectroscopy; HMQC, heteronuclear multiple-quantum correlation spectroscopy; NOE, nuclear Overhauser effect; FID, free induction decay; rmsd, root mean square deviation; EDTA, ethylenedinitrilotetraacetic acid; TSP, 3-(trimethylsilyl)-1-propanesulfonate.

0006-2960/92/0431-12733$03.00/0

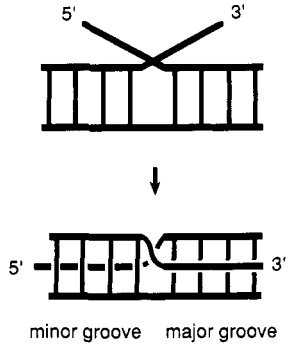

FIGURE 2: Unpaired nucleotides adjacent to coaxially stacked helices can form base triples in the major and minor grooves of the helices. The location of the single-stranded nucleotides on the 3′ strand in the major groove and on the 5′ strand in the minor groove is analogous to the two loop regions of a pseudoknot crossing the major and minor grooves (Pleij et al., 1985).

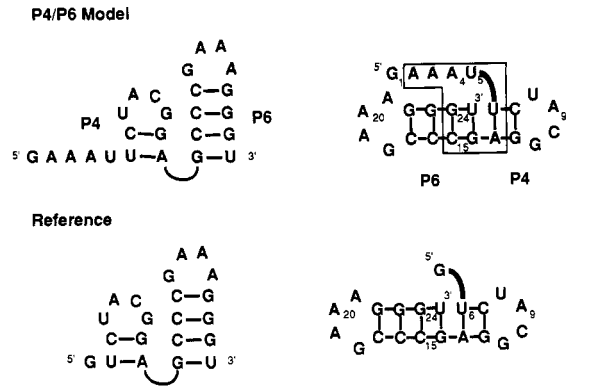

FIGURE 3: Sequences of the molecules characterized by NMR. The P4/P6 model contains two helix regions coaxially stacked with adjacent single-stranded nucleotides forming base triples in the minor groove. The reference molecule contains the same helix and loop regions but lacks the single-stranded nucleotides proposed to form base triples.

MATERIALS AND METHODS

RNA Synthesis and Characterization. RNA oligonucleotides were transcribed from DNA templates using T7 RNA polymerase and were purified using denaturing gel electrophoresis (Wyatt et al., 1991). The yield of the reactions was approximately 30 μg of RNA/mL of reaction. The sequences of the oligonucleotides were verified using base-specific RNases A, T_1, and U_2. UV melting curves were measured as a function of RNA concentration ranging from 3 to 300 μM. The melting curves did not shift over this concentration range, indicating that the structure formed is unimolecular.

NMR Spectroscopy. RNA samples used to measure NMR spectra were dialyzed for 48 h against 10 mM sodium phosphate buffer, pH 6.7, and 0.1 mM EDTA. Samples used to measure exchangeable spectra were lyophilized and then suspended in 400 μL of 90% H_2O/10% D_2O. Samples used to measure nonexchangeable spectra were lyophilized several times from D_2O and then suspended in 400 μL of 99.96% D_2O. The RNA concentration was approximately 1.5 mM. Originally, 1D spectra were taken on samples containing TSP as a chemical shift reference, but later samples were prepared without TSP and referenced to earlier spectra. All 2D NMR spectra were taken on a Bruker AMX-600 spectrometer. Phase-sensitive spectra were collected using the TPPI method (Marion & Wüthrich, 1983). 1D NMR spectra were taken on a Nicolet GN-500 spectrometer. All processing was done with FTNMR (Hare Research, Inc.).

NOESY spectra were collected at 23 °C. Approximately 300 FIDs of 2K points were taken with a sweep width of 6024 Hz. The relaxation delay was 3 s, 80 scans were collected for each FID, and the HDO peak was eliminated by presaturation during the relaxation delay and during the mixing time. For the reference molecule, only a 400-ms mixing time NOESY was measured. For the P4/P6 model, NOESY experiments were performed with mixing times of 60, 120, 200, and 400 ms. Zero quantum artifacts were shifted by incrementing the mixing time with t_1 (Macura et al., 1982). The spectra were apodized in both dimensions with a skewed sine bell shifted by 30° with a skew factor of 0.7.

Phosphorus-decoupled, double-quantum-filtered COSY experiments were done at 23 °C with a sweep width of 6024 Hz. A total of 600 FIDs of 4K points were taken with 64 scans for each FID. The HDO peak was eliminated by presaturation, and the relaxation delay was 3 s. Phosphorus was decoupled using the GARP decoupling sequence (Shaka et al., 1985). The spectrum was apodized with a sine bell shifted 60° in both dimensions.

A natural abundance ^{13}C–^1H heteronuclear multiple-quantum correlation (HMQC) experiment was done at 23 °C as described (Varani & Tinoco, 1991a). Approximately 120 FIDs of 2K points were collected without ^{13}C decoupling. The relaxation delay was 3 s, approximately 400 scans were collected for each FID, and the HDO peak was eliminated by presaturation. The total experiment time was approximately 48 h. The spectrum was apodized with a sine bell shifted 60° and with 3-Hz exponential line broadening in the t_2 dimension and with 5-Hz exponential line broadening in the t_1 dimension.

A NOESY experiment in water was done at 6 °C using a one–one pulse to replace the last pulse of the NOESY sequence. The delay between the one–one pulses was set to 45 μs. The water suppression was optimized by empirically adjusting the length of the second pulse to be slightly (approximately 0.1 μs) shorter than the first pulse. Approximately 400 FIDs of 4K points were collected with a sweep width of 12195 Hz. The relaxation delay was 2 s, and 128 scans were taken for each FID. In the t_2 dimension the spectrum was shifted left by one point and then subtracted from the unshifted spectrum in order to reduce the intensity of the water peak. The spectrum was apodized in both dimensions with a skewed sine bell shifted by 60° with a skew factor of 1.5.

Distance Constraints. The experimentally observed NOESY cross-peaks were converted to distances to serve as constraints during the molecular dynamics. Rather than convert each cross-peak to a specific distance, the cross-peaks were divided into four categories: strong, medium, weak, and very weak. Strong cross-peaks were cross-peaks present in the 60-ms NOESY and were assigned a distance range of 1.8–3.0 Å. All of the constraints derived from strong cross-peaks were internucleotide constraints from sugar $H_{2'}$ protons to the adjacent base H_8/H_6 protons. Medium cross-peaks were distinguished visually from weak cross-peaks in the 120-ms NOESY. Medium cross-peaks were assigned a distance range of 2.0–4.0 Å, and weak cross-peaks were assigned a distance range of 2.0–5.0 Å. One very weak cross-peak was included which was only present in the 400-ms NOESY. This cross-peak is between the A_3 H_2 base proton and the C_{16} $H_{1'}$ sugar proton, and it was assigned a distance range of 3.0–5.0 Å.

The NOESY cross-peaks were divided into categories of distance ranges rather than converting them into specific distances to avoid overinterpreting the data. One of the assumptions made in deriving specific distances between the protons involved in each cross-peak is that all of the protons

in the molecule have the same correlation time. Several aspects of the NMR data suggest that the single-stranded nucleotides undergo different dynamics than the molecule does as a whole. The $H_{1'}$–$H_{2'}$ coupling constants for nucleotides G_1, A_2, A_3, U_6, and U_{25} are intermediate between $C_{2'}$-endo and $C_{3'}$-endo, suggesting that the sugars are interconverting between the two forms. The cross-peak intensity buildup rate between the H_5 and H_6 protons for U_5, U_6, and U_{25} is approximately half as fast as it is for the other pyrimidines in the molecule. Finally, the A_4 H_2 resonance at 8.60 ppm is broader than are other resolved aromatic peaks.

Molecular Dynamics. The molecular dynamics program XPLOR (Brünger, 1990) was used to generate three-dimensional structures consistent with the constraints derived from the NMR data. The force field consisted of bond distances, bond angles, improper angles (used to maintain chirality and base planarity), hard-sphere repulsion, NMR distance constraints, and NMR torsion angle constraints. No electrostatic terms were used in the force field to avoid biasing the results. The objective was to find all the structures consistent with the NMR data. In addition to the NMR distance constraints from the NOESY spectra described above, four or five distance constraints were used for each base pair which effectively maintained hydrogen bonding between the bases and made the bases coplanar. These constraints were set to ±0.1 Å. The sugar conformations were determined on the basis of the $H_{1'}$–$H_{2'}$ coupling measured in the COSY. Sugars for which the $H_{1'}$–$H_{2'}$ coupling was too small to be measured were constrained to be in the $C_{3'}$-endo family of puckers by constraining four of the endocyclic torsion angles in the sugar ring. Sugars which were intermediate between $C_{2'}$-endo and $C_{3'}$-endo were constrained to have the four endocyclic torsion angles in a broad range which encompasses $C_{2'}$-endo, $C_{3'}$-endo, and the $O_{4'}$-endo sugar conformations.

Twenty starting structures were created within XPLOR by randomizing the backbone torsion angles except for the sugar ring. The two tetraloop regions were not included in the calculations to reduce computation time. Folded structures consistent with the NMR data were generated from the random starting structures by an annealing protocol followed by a refinement protocol. The annealing protocol started at high temperature with the repulsive force turned off so that atoms were free to pass through each other. The system was cooled, and gradually the repulsive force was increased. The twenty structures generated by the annealing protocol were then examined for the expected secondary structure (Figure 3) and for close contacts between different regions of the molecule for which no NOEs were seen. Five of the structures generated during the annealing protocol were discarded because they had incorrect secondary structures, and one structure was discarded because it contained several close contacts between protons for which no NOEs were seen. The fourteen remaining structures generated in the annealing protocol were then subjected to a refinement protocol which consisted of energy minimization and further dynamics, during which the repulsive force was on and the temperature was raised slightly.

RESULTS

Assignments. The methods used to assign the resonances in the NMR spectra have been reviewed by Varani and Tinoco (1991b). Assignment of the imino proton resonances in the P4/P6 model confirmed the base pairing (Figure 4). The imino protons from the loop nucleotides, U_8, G_{11}, and G_{18}, were assigned by comparison to previous work on these loops

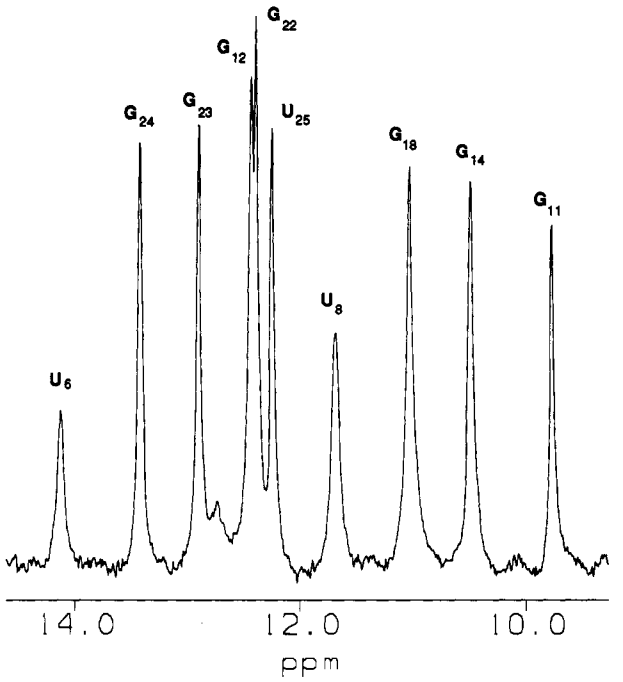

FIGURE 4: Imino proton spectrum of the P4/P6 model taken at 6 °C in 10 mM sodium phosphate, pH 6.7, and 0.1 mM EDTA. The assignments of the ten imino proton resonances are indicated. The imino protons for G_1 and U_5 exchange too quickly to be seen.

(Varani et al., 1991; Heus & Pardi, 1991). The imino protons in the helix regions were assigned from 1D NOEs measured between imino proton resonances and from the NOESY spectrum taken in water. An intense NOE measured between the U_{25} imino proton and the G_{14} imino proton confirms the formation of a G·U base pair, and the NOE measured between the U_6 imino proton and the G_{14} imino proton shows that the two helices stack coaxially. Only 10 resonances are seen in the imino region of NMR spectrum even though there are 12 imino protons. The two imino proton resonances from G_1 and U_5 are not seen because they exchange too quickly with the solvent.

The assignment of the nonexchangeable resonances of the P4/P6 model to specific protons was simplified by comparing the data to the spectra of the reference. The reference molecule contains the same helices and loops as the P4/P6 model but does not contain the adjacent single-stranded nucleotides proposed to form tertiary interactions. Starting with the loop assignments (Varani et al., 1991, Heus & Pardi, 1991), the resonances from the base H_8/H_6 protons and the sugar $H_{1'}$ protons in the helix regions of the reference were assigned as shown in Figure 5 by following the pathway in the NOESY from sugar $H_{1'}$ protons to base H_8/H_6 protons. Comparison of the NOESY spectrum of the reference to the NOESY spectrum of the P4/P6 model (Figure 5) allowed the assignment of the aromatic and $H_{1'}$ protons in the helices and loop regions. The additional connectivity pathway in the NOESY spectrum for the P4/P6 model must be due to the single-stranded nucleotides and was used to assign the base H_8/H_6 protons and the sugar $H_{1'}$ protons for these nucleotides (Figure 5).

Sugar protons other than the $H_{1'}$ were mostly assigned from the NOESY data although, whenever possible, the assignments were confirmed with the COSY data. Unfortunately, spectral overlap prevented the use of the COSY for many of the sugar assignments. The $H_{1'}$ proton has cross-peaks to the $H_{2'}$, $H_{3'}$, and $H_{4'}$ protons in the 400-ms mixing time NOESY. $H_{2'}$ protons were assigned by their strong cross-peaks to the $H_{1'}$

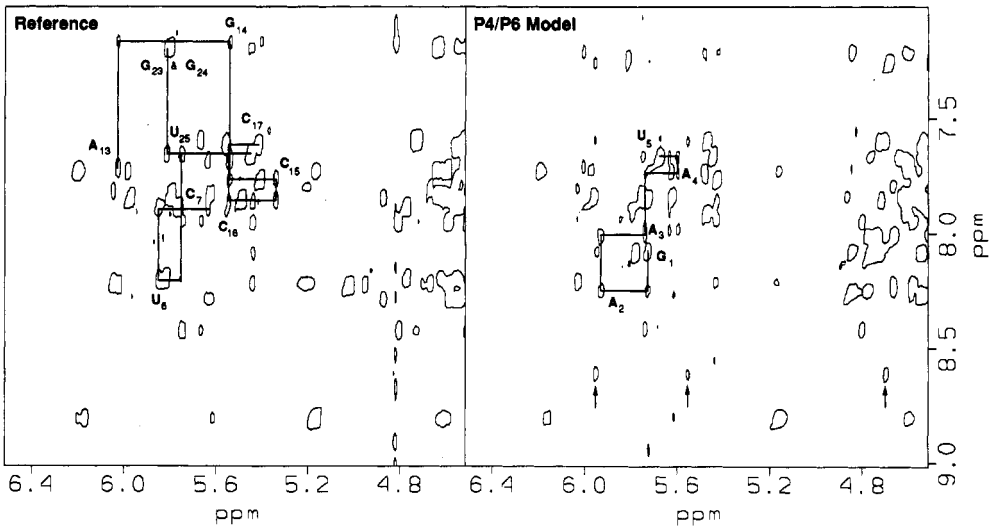

FIGURE 5: NOESY spectra for both molecules showing the $H_8/H_6/H_2$ base proton to $H_{1'}/H_5$ sugar and base proton region. In the NOESY on the reference molecule the H_8/H_6–$H_{1'}$ cross-peaks from the two helix regions are connected by lines. In the NOESY on the P4/P6 model the H_8/H_6–$H_{1'}$ cross-peaks for the single-stranded nucleotides are connected by lines, and the cross-peaks between the A_4 H_2 base proton and the G_{24} $H_{1'}$, G_{24} $H_{2'}$, and U_{25} $H_{1'}$ sugar protons are marked with arrows. These cross-peaks were crucial in determining the structure. Spectra were taken at 23 °C in 10 mM sodium phosphate, pH 6.7, and 0.1 mM EDTA.

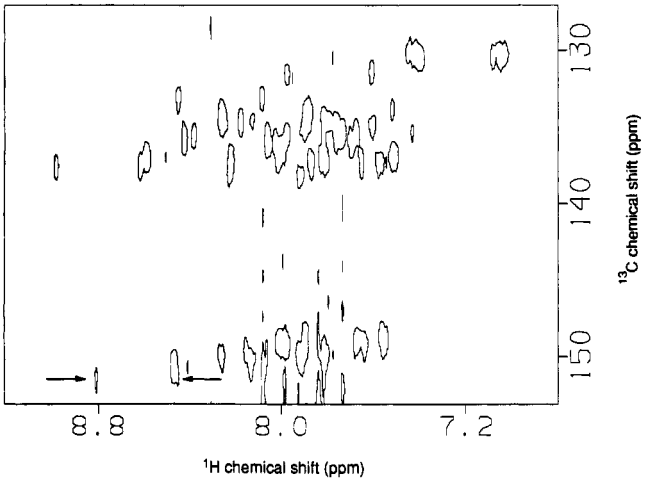

FIGURE 6: H_8/H_6 and H_2 region of the natural abundance ^{13}C–^1H HMQC experiment. Each proton bound to ^{13}C gives two cross-peaks separated by the ^{13}C–^1H coupling constant. The cross-peaks at the top of the spectrum shown are from H_8 and H_6 protons bound to ^{13}C while the cross-peaks at the bottom are due to H_2 protons bound to ^{13}C. The arrows show the two cross-peaks resulting from the proton resonance at 8.60 ppm. This proton must be an adenine H_2 proton since it is bound to a ^{13}C whose chemical shift is downfield from the C_8/C_6 chemical shifts. The spectrum was taken at 23 °C in 10 mM sodium phosphate, pH 6.7, and 0.1 mM EDTA.

protons in the 60-ms mixing time NOESY spectrum. $H_{3'}$ protons were assigned by their intranucleotide cross-peaks to base H_8/H_6 protons in the 120-ms mixing time NOESY. The $H_{4'}$ protons were the remaining unassigned cross-peaks to the $H_{1'}$ protons at long mixing time after the $H_{2'}$ and $H_{3'}$ protons were assigned. $H_{5'}$ and $H_{5''}$ sugar protons were not assigned.

Adenine H_2 protons are difficult to assign by conventional methods unless the adenine is base paired. In base-paired adenines there is a strong NOE between the H_2 proton and the uracil imino proton. Recently it was shown that adenine H_2 protons can be distinguished from H_8/H_6 resonances since the C_2 chemical shift is significantly different from C_8/C_6 chemical shifts (Varani & Tinoco, 1991a). Figure 6 shows the $H_8/H_6/H_2$ region of a natural abundance ^{13}C–^1H HMQC experiment. The H_2 resonances present in the ^{13}C–^1H HMQC spectrum of the P4/P6 model but not in the spectrum of the reference must be from the adenines in the single-stranded region. One of the resonances in the NOESY spectrum which is not found in the spectrum of the reference is at an unusual chemical shift, 8.60 ppm, and has cross-peaks to several sugar protons (Figure 5). The ^{13}C–^1H HMQC experiment confirms that this resonance is from an adenine H_2 proton (Figure 6). Cross-peaks from the single-stranded adenine H_2 protons to duplex sugar protons were crucial in determing the structure of the P4/P6 model and are discussed further below. The complete list of assignments is given in Table I.

Structural Features Determined by NMR. The NMR data measured for the tetraloops in the P4/P6 model, UACG and GAAA, are consistent with the structures determined for these two tetraloop families (Varani et al., 1991; Heus & Pardi, 1991). The NOE connectivities in the double-helix regions are consistent with A-form geometry. There are strong internucleotide NOEs between the $H_{2'}$ sugar protons and the adjacent base H_8/H_6 protons. There are intranucleotide NOEs between the $H_{3'}$ sugar protons and the base H_8/H_6 protons. The H_2 proton from A_{13} has NOEs to the cross-strand $H_{1'}$ proton of C_7 and to the $H_{1'}$ proton of G_{14}. The sugar puckers in the helices are in the $C_{3'}$-endo conformation with the exception of the sugars for U_6 and U_{25} at the junction between the two helices. The H_1–$H_{2'}$ coupling constants for these two sugars are both 4 Hz, indicating that the sugars are in equilibrium between the the $C_{2'}$- and $C_{3'}$-endo conformations with approximately 50% of each conformer present.

Although the NMR data are consistent with the two helices adopting A-form geometry, there is a perturbation at the junction of the two helices. The normal NOE from the $H_{2'}$ sugar proton to the H_6 base proton is missing between U_{25} and U_6. There is a normal NOE pathway on the other strand, between A_{13} and G_{14}, and there is an NOE between the U_6 imino proton and the G_{14} imino proton. The two helices are stacked, but the geometry between the two helices where the single-stranded nucleotides leave the helix is unusual. An NOE between the U_{25} sugar $H_{2'}$ proton and the U_6 base H_5 proton shows that the two helices are rotated with respect to each other (Figure 7). Instead of U_{25} stacking on U_6 so that the two helices form a continuous helix through the junction, the two helices are rotated and U_5, the single-stranded nucleotide proposed to interact with G_{14}, stacks on U_6. There is a strong NOE between the $H_{2'}$ sugar proton of U_5 and the

Table I: Chemical Shifts (ppm) of Assigned Protons in the P4/P6 Model Relative to TSP[a]

	H_8/H_6	H_2/H_5	$H_{1'}$	$H_{2'}$	$H_{3'}$	$H_{4'}$	imino	amino
G_1	8.08	na	5.73	4.84	4.71	4.54		
A_2	8.25		5.92	4.65	4.77	4.58	na	
A_3	8.00	7.98	5.74	4.34	4.73	4.51	na	
A_4	7.73	8.60	5.59	4.53	4.38	4.16	na	
U_5	7.66	5.67	5.68	4.76				
U_6	8.08	5.78	5.72	4.70	4.53	4.31	14.17	na
C_7	7.85	5.73	5.62	4.58	4.32	4.19	na	8.37/6.33
U_8	7.83	5.74	5.61	3.86	4.17		11.76	na
A_9	8.79	8.30	6.16	5.14	4.15	4.24	na	
C_{10}	7.71	6.19	5.97	4.12			na	7.14/6.40
G_{11}	7.84	na	5.97	4.86	5.60	4.40	9.80	7.85/6.64
G_{12}	8.26	na		4.59			12.47	6.95
A_{13}	7.70	7.72	6.01	4.84	4.60	4.50	na	
G_{14}	7.21	na	5.48	4.58	4.39	4.48	11.07	6.50
C_{15}	7.76	5.43	5.46	4.29	4.51		na	8.46/6.99
C_{16}	7.91	5.46	5.64	4.48	4.60	4.42	na	8.52/6.91
C_{17}	7.66	5.44	5.74	4.32			na	8.23/6.69
G_{18}	7.59	na	5.67	4.46			10.69	
A_{19}	8.41	7.85	5.74	4.78			na	
A_{20}	7.94	7.81	5.42	4.33			na	
A_{21}	8.21	8.10	6.03	4.66			na	
G_{22}	7.79	na		4.44			12.44	
G_{23}	7.23	na	5.82	4.67	4.46	4.43	12.95	8.70/6.26
G_{24}	7.26	na	5.96	4.66			13.47	8.56/7.17
U_{25}	7.59	5.46	5.55	4.11			12.28	na

[a] na = not applicable. Spectra were taken in 10 mM sodium phosphate, pH 6.7, at 6 °C for imino and amino protons and at 23 °C for all other protons.

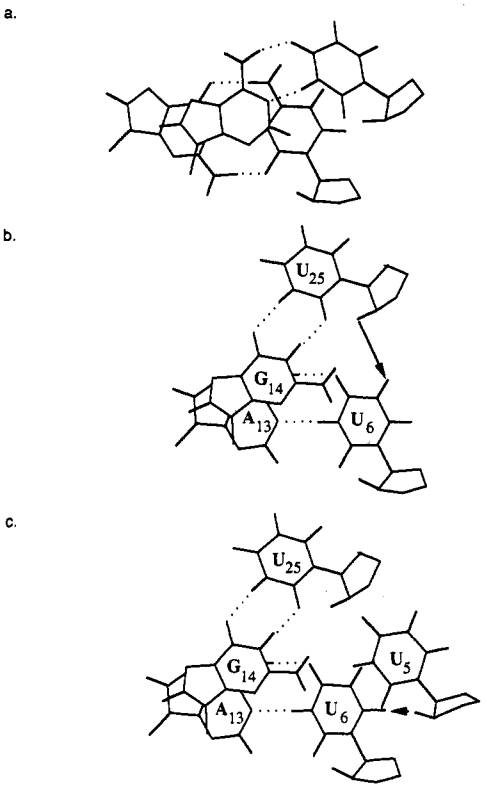

FIGURE 7: (a) Rotation between two base pairs in A-form geometry. (b) Rotation between the two base pairs A_{13}·U_6 and G_{14}·U_{25} in the P4/P6 model. The rotation between these two base pairs is approximately twice as large as the rotation between base pairs in A-form helices. The unusual NOE which indicates the rotation is shown by an arrow. (c) Single-stranded nucleotide U_5 shown stacking above U_6 in normal A-form geometry leaving G_{14}, U_{25}, and U_5 roughly coplanar. The NOE between U_5 and U_6 is shown by an arrow.

H_6 proton of U_6. There are no NOEs between U_5 and either U_{25} or G_{14}. The imino proton of U_5 is not observed (Figure 4), presumably due to fast exchange with the solvent. The possible formation of a U_5·U_{25}·G_{14} base triple will be discussed in the next section.

Table II: Unusual NOEs Measured

NOE measured	shortest mixing time seen (ms)	distance constraint range (Å)
U_5H_2–U_6H_6	60	1.8–3.0
U_5H_6–U_6H_6	400	not used
$U_{25}H_2$–U_6H_5	120	2.0–4.0
A_4H_2–$U_{25}H_{1'}$	120	2.0–5.0
A_4H_2–$G_{24}H_{1'}$	120	2.0–5.0
A_4H_2–$G_{24}H_{2'}$	120	2.0–4.0
A_4H_2–$U_{25}H_6$	400	not used
A_4H_2–A_3H_2	400	not used
A_3H_2–$A_4H_{1'}$	120	2.0–5.0
A_3H_2–$C_{16}H_{1'}$	400	3.0–5.0

The conformation of the single-stranded nucleotides is determined both by sequential intrastrand NOEs and by NOEs between the single-stranded nucleotides and the duplex region. The five single-stranded nucleotides (5′-GAAAU-) are roughly stacked since there are sugar $H_{1'}$ proton to base H_8/H_6 proton NOEs. The $H_{1'}$–$H_{2'}$ coupling constants for the first three nucleotides (5′-GAA-) are 4–6 Hz, intermediate between $C_{2'}$-endo and $C_{3'}$-endo coupling constants. These coupling constants indicate that the sugars interconvert between the two sugar conformations, with approximately 50% of each conformer present. The positions of the next two single-stranded nucleotides, A_4 and U_5, are reasonably well determined as described below.

The NOEs between the single-stranded nucleotides and the duplex consist mainly of NOEs between single-stranded adenine H_2 protons and protons in the helix analogous to P6; a complete list of these NOEs is in Table II. The A_4 H_2 proton has NOEs to the $H_{1'}$ and $H_{2'}$ sugar protons of G_{24} and to the $H_{1'}$ proton of U_{25} (Figure 5). In the long mixing time NOESY there are also NOEs between the A_4 H_2 proton and the A_3 H_2 proton, suggesting that these two nucleotides are stacked. The A_3 H_2 proton has an NOE to the $H_{1'}$ proton of C_{16}. No NOEs between A_2 and the helix were observed. The NOEs between the A_3 and A_4 H_2 protons and $H_{1'}$ sugar protons of the helix show that the single-stranded nucleotides are in the minor groove of the helix. The different pattern of NOEs

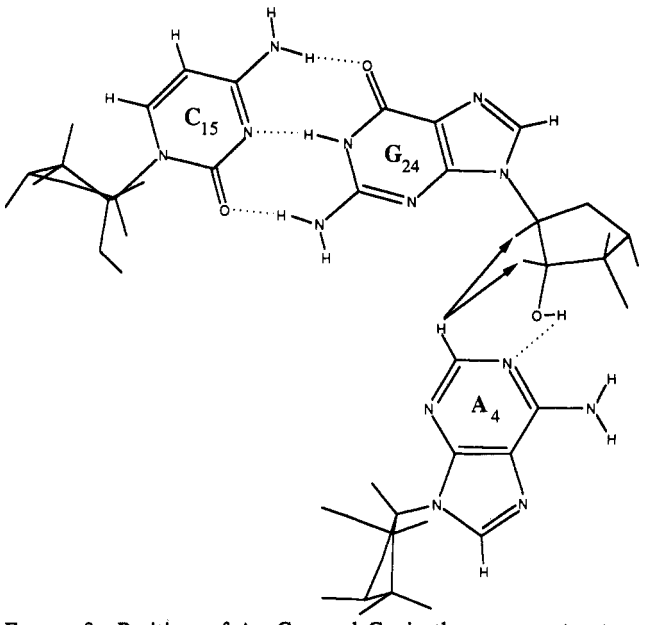

FIGURE 8: Positions of A_4, G_{24}, and C_{15} in the average structure calculated during molecular dynamics. The N_1 position of A_4 is always within 3.5 Å of the 2′-hydroxyl oxygen of G_{24}. The NOEs measured between A_4 and G_{24} are shown by arrows.

between A_3 and A_4 and the helix suggests that a repeating triple-helix structure does not form; the interactions between the single-stranded adenines and the duplex characterized here are probably different than they are in the polynucleotide triple helix, poly(rA)·poly(rG)·poly(rC) (Chastain & Tinoco, 1992).

Molecular Dynamics. Restrained molecular dynamics were done using distance and torsion angle constraints derived from the NMR data in order to find structures consistent with the NMR data. Fourteen folded structures were generated from random starting structures by molecular dynamics using XPLOR (Brünger, 1990). None of the fourteen structures violated the distance constraints derived from the NOE data, although there were minor violations of the distance constraints used to constrain the hydrogen bonding between the base pairs. In order to assess how precisely the structures are determined by the NMR data, the fourteen structures generated by molecular dynamics were superimposed, and an average structure was calculated from these structures. The root mean square deviation (rmsd) for each structure compared to the average structure was calculated. The average rmsd when the molecules are superimposed on the basis of all the atoms in the structure is 4.3 Å. If G_1, A_2, and A_3 are removed from the calculation, then the rmsd is lowered to 1.5 Å. The positions of these three nucleotides are not determined; however, the positions of the single-stranded nucleotides proposed to form base triples, A_4 and U_5, are fairly well determined by the NMR data.

The positions of A_4, G_{24}, and C_{15} in the average structure are shown in Figure 8. The three bases are approximately coplanar. Arrows show the measured NOEs between A_4 and G_{24}. The N_1 of A_4 is within hydrogen-bonding distance of the 2′-hydroxyl of G_{24}. The average distance in the 14 different structures between the N_1 and the 2′-hydroxyl oxygen is 3.0 Å, and the largest distance is only 3.5 Å. Superimposing the A·G·C base triples in the 14 different structures based on the positions of A_4 and G_{24} shows that the position of A_4 relative to G_{24} is well-defined (Figure 9). The direct NMR data constraining A_4 to be near G_{24} consist of three NOEs (Table II); however, there are also NOEs between A_3 and C_{16} and between U_5 and U_6 so that backbone is heavily constrained as well. Observation of the 2′-hydroxyl proton resonance in the NMR spectrum would provide additional evidence for the N_1 to 2′-hydroxyl hydrogen bond. Normally 2′-hydroxyl protons are not seen because of fast exchange with the solvent, however, and no new cross-peak from the A_4 H_2 proton is seen in the water NOESY experiment presumably for this reason. In addition to the hydrogen bond between the N_1 and the 2′-hydroxyl, the structure shown in Figure 8 could be stabilized by hydrogen bonds between the adenine amino proton and the $O_{2'}$ and $O_{3'}$ of G_{24} sugar. The average distances between the adenine N_6 and the $O_{2'}$ and $O_{3'}$ of G_{24} are 4.7 and 5.4 Å, respectively. In several of the structures the distance between the adenine N_6 and the $O_{2'}$ is less than 4 Å.

The downfield chemical shift of the A_4 H_2 proton, 8.60 ppm, is further evidence for the structure of the A·G·C base triple. A downfield chemical shift can result from an unstacked adenine as in the case of A_9 in the UACG loop. This nucleotide is known to be unstacked from the structure of the UNCG family of tetraloops (Varani et al., 1991), and the chemical shifts of the A_9 H_8 and H_2 protons are 8.79 and 8.30 ppm, respectively. In the case of A_4, however, the H_2 proton resonates at 8.60 ppm but the H_8 proton resonates at 7.73 ppm. The H_8 chemical shift is consistent with A_4 stacking between A_3 and U_5. The unusual chemical shift of the A_4 H_2 is explained by the structure shown in Figure 8. The A_4 H_2 proton is in the plane of G_{24} pointing at N_3. A proton above or below the plane of a base is shifted upfield, while protons in the plane of a base are shifted downfield. The A_4 H_2 resonance would also be shifted downfield if it were involved in a C–H···N hydrogen with the N_3 of G_{24}. A number of crystallographic studies suggest that such hydrogen bonds exist (Jeffrey & Saenger, 1991), but the distance between the A_4 C_2 and the G_{24} N_3 is long for a hydrogen bond. The shortest distance between these atoms in the 14 structures is 4 Å, and the average distance is 5.1 Å. We attribute the unusual chemical shift of the A_4 H_2 to the ring current effect of G_{24} although there could be a weak hydrogen bond to the G_{24} N_3 as well.

Another feature of the structure which is well-defined by the NMR data is the stacking between the two helices. The two helices are coaxially stacked, but the rotation between the two helices is greater than the rotation between two normal base pairs (Figure 7). The rotation between two base pairs can be calculated as the angle between the line formed between the two $C_{1'}$ atoms of one base pair and the line formed between the two $C_{1'}$ atoms of the next base pair. The average rotation between the two base pairs at the junction of the two helices in the 14 structures is 59° compared to 33° for the rotation between two base pairs in A-form geometry. The average rotation between the other base pairs in the structures is 26°. The two helices in the reference molecule are not rotated as the two helices are in the P4/P6 model. A normal NOE between the U_{25} sugar $H_{2'}$ proton and the U_6 base H_6 proton indicates that the two helices maintain A-form geometry at the junction between the two helices. The cause of the rotation in the P4/P6 model is not simply the stacking of U_5 on U_6 since one single-stranded nucleotide is not sufficient to cause the helices to rotate in the reference molecule. The rotation is probably a result of the interactions between the single-stranded nucleotides and the minor groove of the helix.

The stacking of U_5 onto U_6 and the rotation of the two helices result in U_5, U_{25}, and G_{14} being approximately coplanar. The position of U_5 relative to U_{25} is not as well determined as the position of A_4 is relative to G_{24} (Figure 9). In most of

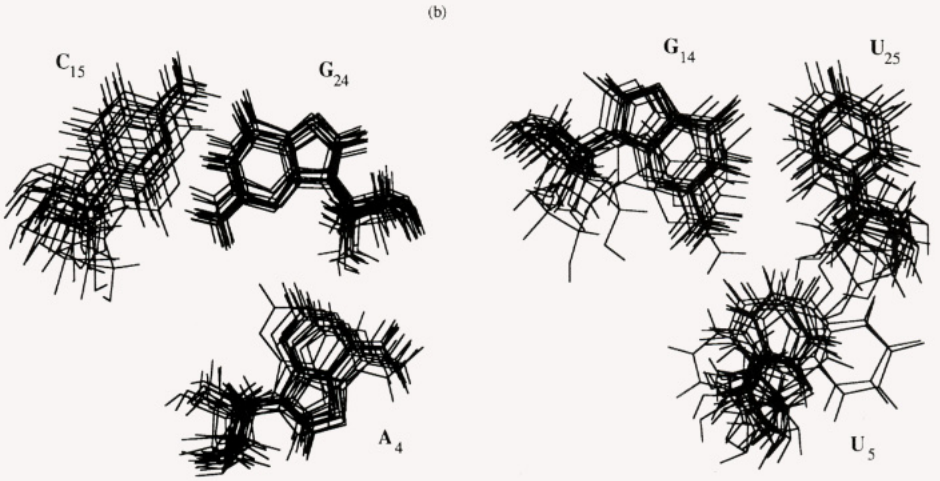

FIGURE 9: (a) Positions of A_4, G_{24}, and C_{15} in the fourteen structures generated during molecular dynamics are superimposed. The structures are superimposed only on the basis of the positions of A_4 and G_{24} in order to show that the relative orientation of these nucleotides is well determined by the NMR data. (b) Positions of U_5, U_{25}, and G_{14} in the fourteen structures generated during molecular dynamics are superimposed. The structures are superimposed only on the basis of the positions of U_5 and U_{25}.

the structures, U_5 is positioned so that the U_5 O_4 is between the G_{14} amino proton and the U_{25} 2′-hydroxyl. This provides an opportunity for one or two hydrogen bonds. The average distances between the the U_5 O_4 and the U_{25} 2′-hydroxyl oxygen and between the U_5 O_4 and the G_{14} amino nitrogen are 4.1 and 5.4 Å, respectively. The distance from the O_4 to the 2′-hydroxyl oxygen is less than 3 Å in five of the structures, and the largest distance is 6 Å. The distance from the O_4 to the amino nitrogen is less than 4 Å in only three of the structures, and the largest distance is 10.7 Å. Only one amino proton was seen in the water NOESY for G_{14}, suggesting that the two amino protons are in fast exchange. Considering both the distance between the U_5 O_4 and the G_{14} amino and the exchange behavior of the amino, it seems more likely that there is a hydrogen bond from the U_5 O_4 to the 2′-hydroxyl than to the G_{14} amino.

DISCUSSION

Several base triples are proposed to occur in the group I intron on the basis of phylogenetic comparison (Michel & Westhof, 1990). All of the base triples which occur in tRNA, or are proposed to occur in the group I intron, form in a region of the RNA where two adjacent helices are flanked by single-stranded nucleotides. The simplest explanation for these base triples is that the two helices stack coaxially with the single-stranded nucleotides on the 5′ strand bound in the minor groove of one of the helices and the single-stranded nucleotides on the 3′ strand bound in the major groove of the other helix (Figure 2). The structure of the P4/P6 molecule characterized here is consistent with this model for the formation of base triples. The NMR data show that the two helices are coaxially stacked but rotated relative to each other. The rotation is about twice that between normal base pairs and allows the single-stranded nucleotides on the 5′ strand to stack onto one helix and to interact with the other helix in the minor groove (Figure 10).

The structure of the oligonucleotide P4/P6 model supports the proposed base triples in the minor groove at the P4/P6 junction in the group I intron (Michel & Westhof, 1990), but there is no direct evidence that the structure of the P4/P6 region of the group I intron is the same as the structure of the oligonucleotide. The presence of other tertiary interactions in the group I intron could result in a different structure. However, preliminary NMR data collected on several oligonucleotides not described in this paper show that those oligonucleotides which contain conserved sequences at the P4/P6 junction form base triples, but oligonucleotides containing other sequences do not (Chastain and Tinoco, unpublished results). This suggests that the minor groove base triples do form in the intron.

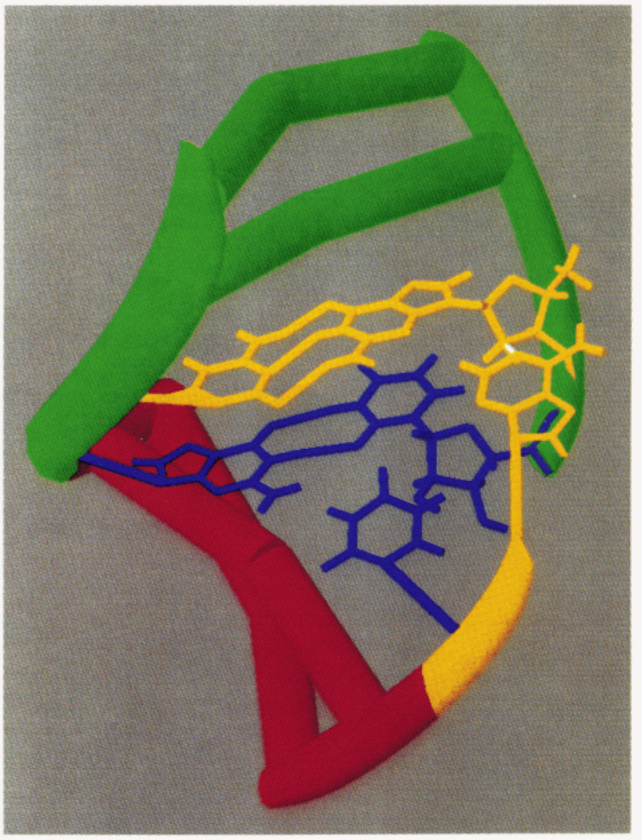

FIGURE 10: A picture of the average structure of the P4/P6 model. The first three nucleotides (5′-GAA) and the two tetraloops are not shown. The P4 helix is at the bottom in red, the P6 helix is at the top in green, the nucleotides in the U_5·U_{25}·G_{14} base triple are in blue, and the nucleotides in the A_4·G_{24}·C_{15} base triple are in yellow. Nucleotides not involved in base triples are shown as cylinders. The hydrogen bond between the A_4 N_1 and the G_{24} 2′-hydroxyl is shown in white.

The tertiary interactions involving the P6 helix are important in the function of the intron. Changing the intron sequence

so that base pairing in P6 is disrupted greatly reduces the activity of the intron. Mutations which restore base pairing but change the sequence of base pairs in the P6 helix also greatly reduce activity (Ehrenman et al., 1989; Couture et al., 1990). If the tertiary interactions in the group I intron involving helix P6 are the base triples characterized here, their functional role is probably to orient other parts of the molecule which are responsible for catalysis. The formation of these base triples in the core of the intron is a strong constraint on its structure since there are only a few unpaired nucleotides between the P4/P6 stacked helices and the P3/P7/P8 stacked helices which contain the G-binding site (Michel et al., 1989). In addition to aligning these two helix domains, the formation of base triples may cause a rotation between the P4 and P6 helices which could also serve to align functional groups within the catalytic core of the intron.

Although the structure of the oligonucleotide P4/P6 model is consistent with the formation of a base-triple domain as proposed on the basis of phylogenetic comparison by Michel and Westhof (1990), the hydrogen bonds between the single-stranded nucleotides and the helices in the proposed structure are different from those found by NMR. This is not surprising since the proposed structure was not based on experimental structural data. In the structure determined by NMR, the N_1 ring nitrogen of A_4 is hydrogen bonded to the G_{24} 2′-hydroxyl, whereas in the structure proposed by Michel and Westhof (1990), the corresponding adenine forms hydrogen bonds with the guanine base. The proposed structure is inconsistent with the observed NOEs between the A_4 H_2 proton and G_{24} sugar protons since the distances between these protons are about 8 Å in the proposed structure. The other base triple proposed in the minor groove is a U·U·G base triple with a hydrogen bond between the uracil O_4 and the guanine amino proton. Although the interactions between U_5 and G_{14}·U_{25} could not be rigorously determined, the structures generated from the NMR data suggest that the U_5 O_4 could form hydrogen bonds to both the G_{14} amino proton and the 2′-hydroxyl proton.

Tertiary interactions involving 2′-hydroxyl groups are a common feature of RNA tertiary structure. In the crystal structures of yeast $tRNA^{Phe}$ and $tRNA^{Asp}$, 2′-hydroxyl groups form hydrogen bonds to bases, sugars, and phosphates (Holbrook et al., 1978; Westhof et al., 1985). One of the hydrogen bonds involving a 2′-hydroxyl group in yeast $tRNA^{Phe}$ is very similar to the one found here. A single-stranded adenine, A_{21}, binds in the minor groove of a tertiary base pair, U_8·A_{14}. The three bases are approximately coplanar, and A_{21} forms a hydrogen bond between its N_1 and the 2′-hydroxyl of U_8 (Figure 11). The interactions between an adenine and both an A·U pair (tRNA) and a G·C pair (P4/P6 model) indicate that single-stranded adenines can bind to base pairs independent of sequence by forming hydrogen bonds to the 2′-hydroxyl groups. The possible hydrogen bond found in the P4/P6 model between the O_4 of U_5 and the 2′-hydroxyl group of U_{25} suggests that the O_4 position of uracil and the N_1 position of adenine may be equivalent. Single-stranded regions of RNA containing adenine and uracil may bind adjacent duplex regions by forming hydrogen bonds to the 2′-hydroxyl groups.

Tertiary interactions between single-stranded adenines and 2′-hydroxyl groups are important in the structure and function of the group I intron. Specific 2′-hydroxyl groups in the P1 helix are recognized by the intron during the 5′ splicing reaction (Bevilacqua et al., 1991; Pyle & Cech, 1991). Mutagenesis studies have identified a specific adenine, A_{302}, which hydrogen bonds to the 2′-hydroxyl of one nucleotide in the P1 stem

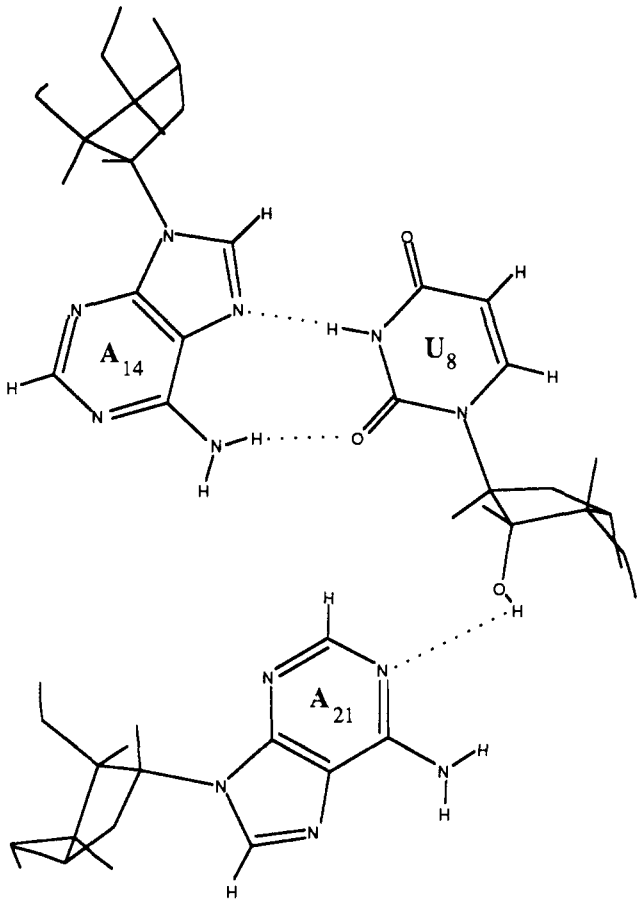

FIGURE 11: Interaction between A_{21} and the tertiary base pair A_{14}·U_8 in the crystal structure of yeast $tRNA^{Phe}$ (Holbrook et al., 1978).

during the 5′ splicing reaction (Pyle et al., 1992). This adenine can be substituted with a uracil and still retain function, but not with a guanine or a cytidine. Additional hydrogen bonds between single-stranded adenines in the intron and 2′-hydroxyl groups in the minor groove of the P1 helix are proposed in the three-dimensional model of the group I intron built by Michel and Westhof (1990). Tertiary interactions involving single-stranded adenines may also occur in ribosomal RNA. Phylogenetic comparison of 16S rRNA sequences showed that conserved single-standed adenines outnumber conserved base-paired adenines by 4:1 whereas the ratio for other nucleotides is close to 1:1 (Moazed et al., 1986). Although there is evidence suggesting that single-stranded adenines are targets of protein binding (Moazed et al., 1986), they may also be involved in tertiary interactions.

The structure of the P4/P6 model indicates that a region containing two adjacent helices flanked by single-stranded nucleotides is the minimum structure needed to form a base-triple domain. The widespread occurrence of such regions in RNA molecules suggests that interactions between single-stranded nucleotides and adjacent helix regions are a repeating feature of RNA structure. The formaton of such interactions will depend on the sequences of the single-stranded nucleotides and the helix regions as well as on the formation of other tertiary interactions in the molecule. The structures of the P4/P6 model and of $tRNA^{Phe}$ show that single-stranded adenines adjacent to helix regions can bind 2′-hydroxyl groups in the minor groove independent of the duplex sequence. The hydrogen bond between a single-stranded adenine and a 2′-hydroxyl emphasizes the importance of 2′-hydroxyl groups in RNA tertiary interactions. The structural characterization of tertiary interactions by NMR from only a few NOE contacts

suggests that NOEs between nucleotides which are far apart in the sequence are more restrictive constraints than NOEs between nucleotides which are adjacent in the sequence. This is very promising for the study of larger RNA molecules by NMR.

ACKNOWLEDGMENT

We thank Mr. David Koh for synthesizing DNA templates, Ms. Barbara Dengler for managing the laboratory, and Dr. Gabriele Varani and Dr. Jacqueline Wyatt for many helpful discussions and for their critical reading of the manuscript.

REFERENCES

Bevilacqua, P. C., & Turner, D. H. (1991) *Biochemistry 30*, 10632–10640.

Brünger, A. T. (1990) *X-PLOR: A System for Crystallography and NMR, Version 2.1*, Yale University, New Haven, CT.

Chastain, M., & Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 315–318.

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.

Doudna, J. A., & Szostak, J. W. (1989) *Nature 339*, 519–522.

Ehrenman, K., Schroeder, R., Chandry, P. S., Hall, D. H., & Belfort, M. (1989) *Nucleic Acids Res. 17*, 9147–9163.

Guthrie, C. (1991) *Science 253*, 157–163.

Heus, H. A., & Pardi, A. (1991) *Science 253*, 191–194.

Holbrook, S. R., Sussman, J. L., Warrant, R. W., & Kim, S.-H. (1978) *J. Mol. Biol. 123*, 631–660.

Jeffrey, G. A., & Saenger, W. (1991) *Hydrogen bonding in biological structures*, Springer-Verlag, New York.

Macura, S., Wüthrich, K., & Ernst, R. R. (1982) *J. Magn. Reson. 46*, 269–282.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585–610.

Michel, F., Hanna, M., Bartel, D. P., & Szostak, J. W. (1989) *Nature 342*, 391–395.

Michel, F., Ellington, A. D., Couture, S., & Szostak, J. W. (1990) *Nature 347*, 578–580.

Moazed, D., Stern, S., & Noller, H. F. (1986) *J. Mol. Biol. 187*, 399–416.

Noller, H. F., Hoffarth, V., & Zimniak, L. (1992) *Science 256*, 1416–1419.

Pleij, C. W. A., Rietveld, K., & Bosch, L. (1985) *Nucleic Acids Res. 13*, 1717–1731.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76–80.

Pyle, A. M., & Cech, T. R. (1991) *Nature 350*, 628–631.

Pyle, A. M., Murphy, F. L., & Cech, T. R. (1992) *Nature 358*, 123–128.

Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.

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.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.

Wyatt, J. R., Chastain, M., & Puglisi, J. D. (1991) *BioTechniques 11*, 764–769.

Yarus, M. (1988) *Science 240*, 1751–1758.