Nucleoside Triples from the Group I Intron†

Michael Chastain and Ignacio Tinoco, Jr.*

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

Received July 8, 1993; Revised Manuscript Received October 11, 1993®

ABSTRACT: Oligonucleotides modeled on a proposed base-triple domain in the P4/P6 region of the self-splicing group I intron have been characterized by NMR. The NMR data indicate that single-stranded nucleotides in this domain are in the minor groove of an adjacent helix within hydrogen bonding distance of 2′-hydroxyl groups in an interaction we term a nucleoside triple. Oligonucleotides containing the two most frequently occurring sequences among group I introns in the P4/P6 region formed nucleoside triples in the minor groove, whereas oligonucleotides containing sequences which are not conserved did not form triples. Surprisingly, the structures of the nucleoside triples in the oligonucleotides containing the two most frequently occurring sequences are different. If this difference were maintained in the context of the whole intron, it would suggest that the triples are not directly involved in catalysis, but rather that the nucleoside triples function by aligning the helical domains within the catalytic core of the intron.

The three-dimensional structures of RNA molecules determine their function in catalysis and molecular recognition during translation, RNA splicing, and other processes. Tertiary interactions between secondary structure elements play a large role in determining the three-dimensional structures of RNA molecules, but these interactions are not very well characterized. A prevalent type of tertiary interaction in RNA is the base triple which consists of a base from a loop or single-strand region hydrogen bonded to a base pair in a duplex. Base triples within a biological RNA were first found in the crystal structure of tRNA (Holbrook et al., 1978), and several base triples have been proposed to form in the P4/P6 helices and in the P7 helix of group I introns (Michel & Westhof, 1990).

The simplest model explaining the formation of base triples in tRNA and in the intron is that triples form at the junction between two helices. If the two helices stack coaxially to form a quasi-continuous helix, unpaired nucleotides on the 5′ strand leaving the helix junction can form base triples in the minor groove whereas unpaired nucleotides on the 3′ strand leaving the junction can form base triples in the major groove (Figure 1). The location of the nucleotides on the 5′ strand in the minor groove and on the 3′ strand in the major groove is a result of the right-handed nature of the RNA helix. A rotation between the coaxially stacked helices causes the single-stranded nucleotides at the junction of the helices to be on opposite sides of the base pairs, hence in different grooves. A right-handed rotation causes the nucleotides on the 5′ strand to be in the minor groove and the nucleotides on the 3′ strand to be in the major groove. The location of the nucleotides in the major and minor grooves is analogous to the two loops of a pseudoknot crossing its major and minor grooves (Pleij et al., 1985).

Studying the structures of oligonucleotides designed to form base triples has several advantages over studying the whole molecule in which the base triples are proposed to form. Higher resolution structures can be obtained by studying oligonucleotides by nuclear magnetic resonance (NMR)[1] than by studying large RNA molecules, and the formation of base

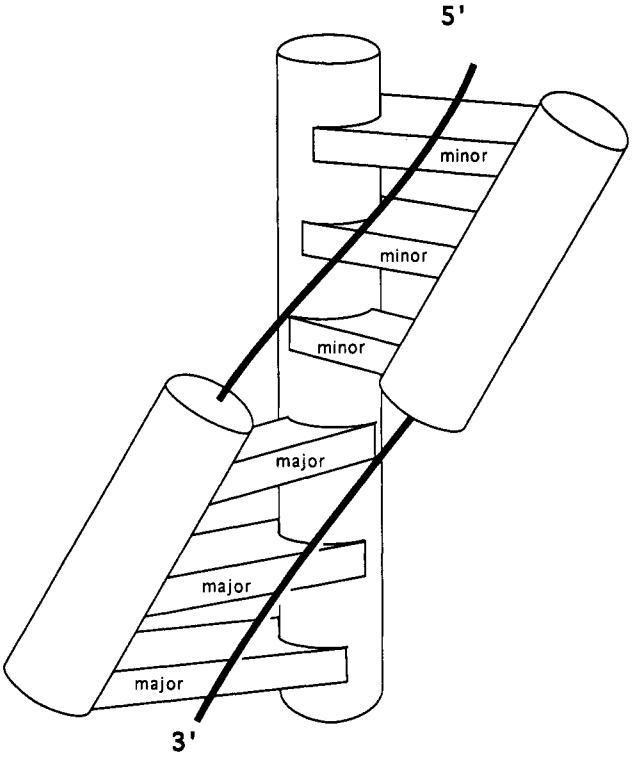

FIGURE 1: Diagram of the junction between two helices. The backbone of each helix is continued at the junction, showing how the two strands of unpaired nucleotides are positioned on opposite sides of the base pairs and therefore are in different grooves of the helices. The 5′ strand goes into the minor groove of one helix, and the 3′ strand goes into the major groove of the other helix. The location of the strands in these grooves is a result of the right-handed rotation between the two helices.

† 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. was a Howard Hughes Medical Institute Doctoral Fellow.

® Abstract published in *Advance ACS Abstracts,* December 15, 1993.

[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, ethylenediaminetetraacetic acid; TSP, 3-(trimethylsilyl)-1-propanesulfonate.

0006-2960/93/0432-14220$04.00/0

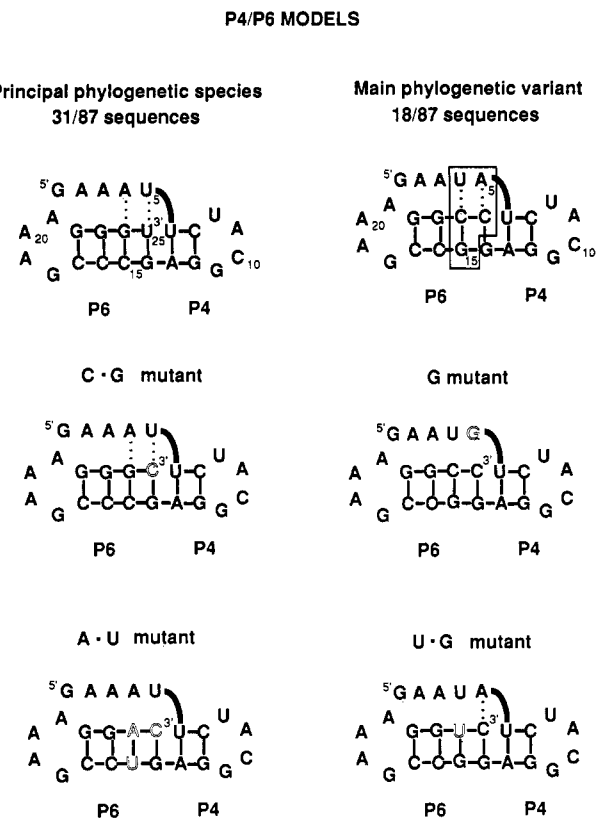

FIGURE 2: Sequences of the oligonucleotides modeled on the P4/P6 region of the group I intron characterized by NMR: the sequence in the principal phylogenetic species occurs in 31/87 aligned introns (Michel & Westhof, 1990). This molecule was characterized previously (Chastain & Tinoco, 1992). The sequence in the main variant occurs in 16/87 aligned introns (Michel & Westhof, 1990). The nucleotides which are changed between the principal species and the main variant are boxed in the main variant. The C·G and A·U mutants are derived from the principal species, and the other two mutants are derived from the main variant. The nucleotides which were changed are shown in shadow font. Tertiary interactions between single-stranded nucleotides and the adjacent helix are indicated by dotted lines.

triples can be studied independently of the formation of other tertiary interactions. By studying different sequences which form base triples, we ultimately hope to determine predictive algorithms for base-triple formation within larger RNA molecules.

The specific region proposed to form base triples that we have studied is the P4/P6 region of the group I intron. On the basis of phylogenetic comparison, two base triples, A·(G·C) and U·(U·G), were proposed to form in the minor groove of the P6 helix (Michel & Westhof, 1990) even though previously characterized base triples occur in the major groove. NMR studies on an oligonucleotide modeled on the P4/P6 region of the group I intron showed that single-stranded nucleotides can form tertiary interactions in the minor groove of the adjacent helix (Chastain & Tinoco, 1992). The sequence contained in this oligonucleotide was the most frequently occurring sequence among group I introns in the P4/P6 region (Michel & Westhof, 1990), and this oligonucleotide will be referred to as the principal species (Figure 2). The NMR data showed that the two helices in the principal species stack coaxially, but the rotation between the two helices is approximately twice as large as the rotation between two base pairs. This rotation allows the single-stranded nucleotides to interact with base pairs in the minor groove. NOEs between the single-stranded nucleotides and sugar protons in the minor groove indicated that the single-stranded nucleotides were within hydrogen bonding distance of 2′-hydroxyl groups.

The minor groove base triples proposed on the basis of phylogeny contained hydrogen bonds between the bases, not between a base and a sugar (Michel & Westhof, 1990). Although we originally described the interactions in the minor groove of the P4/P6 model oligonucleotide as base triples (Chastain & Tinoco, 1992), in this paper we use the term nucleoside triple in order to emphasize the difference from a traditional base triple in which all of the hydrogen bonds are made between bases. Our use of the term nucleoside triple does not preclude hydrogen bonds between the bases, but indicates that the triple occurs in the minor groove and that the base in the minor groove forms a hydrogen bond to the sugar of a base-paired nucleotide. Transfer RNA structures contain interactions which fit this definition of nucleoside triples. The $tRNA^{Phe}$ structure contains a three-nucleotide interaction, $A_{21}\cdot(U_8\cdot A_{14})$, in which the only hydrogen bond between A_{21} and the U·A tertiary base pair is a hydrogen bond to the uridine 2′-hydroxyl (Holbrook et al., 1978). In $tRNA^{Asp}$, the structure of this nucleoside triple changes so that it also includes a hydrogen bond between the A_{21} and A_{14} bases (Westhof et al., 1985).

In the present study, we have used NMR to characterize the structures of five oligonucleotides which have the potential to form nucleoside triples (Figure 2). The most extensively characterized oligonucleotide contains the sequence which occurs second most frequently among group I intron sequences (Michel & Westhof, 1990) and is referred to as the main variant. Phylogenetic comparison predicts that it forms U·(C·G) and A·(C·G) triples. Similar to the principal species, NOEs were measured in the main variant between adenine H_2 protons and sugar protons in the minor groove of the helix analogous to P6. The molecular dynamics program XPLOR (Brünger, 1989) was used to generate structures consistent with the NMR data. The structures generated using molecular dynamics form U·(C·G) and A·(C·G) nucleoside triples stabilized by hydrogen bonds to 2′-hydroxyl groups in the minor groove of the helix. Surprisingly, the structure of these nucleoside triples is very different from the structure of the triples in the principal species. Four mutant oligonucleotides which differ by only a few nucleotides from either the principal species or the main variant (Figure 2) were also characterized by NMR in order to further understand the sequence specificity of nucleoside 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 (Milligan et al., 1987; Wyatt et al., 1991). The yield of the reactions was approximately 100 μg of RNA/mL of reaction for the main variant and the mutants derived from it and approximately 30 μg of RNA/mL of reaction for the mutants derived from the principal species.

NMR Spectroscopy. RNA samples used to measure NMR spectra were dialyzed for 48 h against 10 mM sodium phosphate buffer (pH 6.7)/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. Spectra were referenced by assigning the furthest downfield aromatic resonance to 8.8 ppm. This resonance is from the adenine H_8 of the UACG hairpin loop which was originally

referenced to TSP. NMR spectra were taken on a BRUKER AMX-600 spectrometer or a Nicolet GN-500 spectrometer. Phase-sensitive spectra were collected using the TPPI method (Marion & Wüthrich, 1983). All processing was done with FTNMR (Hare Research, Inc.).

NOESY spectra were collected with 300 FIDs of 2K points and 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 main variant, NOESY spectra with 400-ms mixing times were taken at both 15 °C and 23 °C, and a NOESY spectrum with a 150-ms mixing time was taken at 23 °C. For the G, C·G, and U·G mutant molecules, NOESY spectra with 400-ms mixing times were taken at 23 °C. In addition, NOESY spectra with a 400-ms mixing time were taken on the G, C·G, and A·U mutant molecules at 15 °C. 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.

A phosphorus-decoupled, double quantum filtered COSY experiment was collected for the main variant at 23 °C with a sweep width of 6024 Hz; 400 FIDs of 2K 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 NOESY experiment in water was collected for the main variant 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 three categories: strong, medium, and weak. Strong cross-peaks were the most intense cross-peaks present in the 150-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 in the 150-ms NOESY were assigned a distance range of 2.0–4.0 Å, and weak cross-peaks were assigned a distance range of 3.0–5.0 Å.

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 for the main variant. Molecular dynamics calculations were not done on any of the mutant molecules. 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 45 NMR distance constraints derived from the NOESY spectra as described above, 4 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 20 structures generated by the annealing protocol were then examined for the expected secondary structure (Figure 2) and for close contacts between different regions of the molecule for which no NOEs were seen. Six 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 13 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 (1991). Comparison of the spectra taken of the main variant to previously taken spectra of the principal species (Chastain & Tinoco, 1992) made the assignments much easier. The H_8/H_6 base protons and the sugar $H_{1'}$ protons were assigned from the 400-ms NOESY as shown in Figure 3. The connectivity for the single-stranded nucleotides (5′-GAAUA-) is marked with lines. Once the base H_8/H_6 protons and sugar $H_{1'}$ protons were assigned, further sugar protons were assigned as previously described (Chastain & Tinoco, 1992) from the NOESY data, although whenever possible the assignments were confirmed by the COSY data. The complete list of assignments is given in Table 1.

Structural Features Determined by NMR. The NMR data measured for the stem and loop regions of the main variant show these regions are in conformations similar to those of the principal species. The structures of the two tetraloops, 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 a cross-strand NOE to

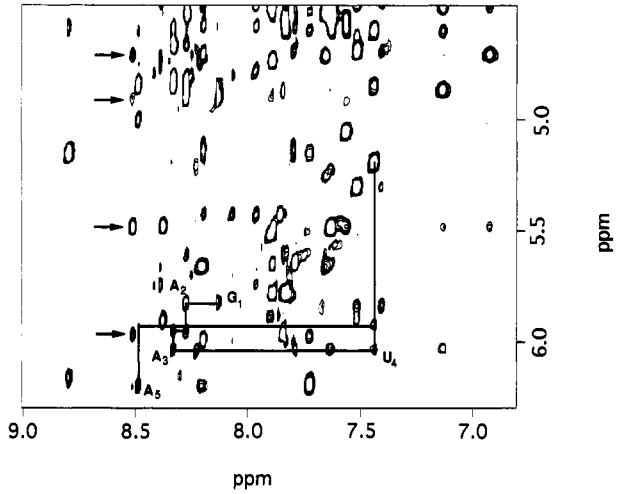

FIGURE 3: NOESY spectra of the main variant showing the $H_8/H_6/H_2$ base proton to $H_{1'}/H_5$ sugar and base proton regions. The cross-peaks resulting from the single-stranded nucleotides are connected by lines, and the cross-peaks between the A_5 H_2 base proton and the G_{14} $H_{1'}$, G_{14} $H_{2'}$, and U_6 $H_{1'}$ sugar protons are marked with arrows. These cross-peaks were crucial in determining the structure. The spectrum was taken at 23 °C in 10 mM sodium phosphate (pH 6.7)/0.1 mM EDTA.

Table 1: Chemical Shifts (ppm) of Assigned Protons in the Main Variant Relative to TSP[a]

	H_8/H_6	H_2/H_5	$H_{1'}$	$H_{2'}$	$H_{3'}$	$H_{4'}$
G_1	8.13	na[b]	5.83	4.91	4.85	4.30
A_2	8.28		5.96	4.66		
A_3	8.34	8.38	6.04	4.60	4.86	
U_4	7.44	5.20	5.93	4.25		
A_5	8.48	8.52	6.20	4.83	5.01	
U_6	8.21	5.67	5.98	4.75		
C_7	7.89	5.78	5.65	4.53	4.35	
U_8	7.83	5.78	5.60	3.86		
A_9	8.80	8.30	6.16	5.15	4.15	
C_{10}	7.73	6.19	5.98	4.13		
G_{11}	7.84	na	5.96	4.87	5.63	4.42
G_{12}	8.28	na		4.54		
A_{13}	7.64	7.91	6.04	4.87		
G_{14}	7.14	na	5.49	4.71	4.50	
G_{15}	6.93	na	4.93	4.33		
C_{16}	7.56	5.06	5.48	4.55		
C_{17}	7.63	5.49	5.65	4.52		
G_{18}	7.66	na	5.67			
A_{19}	8.39	7.86	5.74	4.78		
A_{20}	7.96	7.81	5.42	4.33		
A_{21}	8.20	8.10	6.00	4.68		
G_{22}	7.80	na		4.35		
G_{23}	7.41	na	5.85	4.71	4.48	
C_{24}	7.52	5.31	5.91	4.52		
C_{25}	7.89	5.52	5.88	4.17	4.37	4.90

[a] Spectra were taken in 10 mM sodium phosphate, pH 6.7, at 23 °C. [b] na = not applicable.

the $H_{1'}$ proton of C_7 in addition to the same-strand NOE to the $H_{1'}$ proton of G_{14}. The main variant forms a kinetically more stable structure than the principal species since fewer of the sugar puckers show evidence of interconversion between the $C_{2'}$-*endo* and the $C_{3'}$-*endo* conformation. All of the sugar puckers in the helices are in the $C_{3'}$-*endo* conformation whereas in the principal species the sugars corresponding to U_6 and C_{24} were intermediate between $C_{2'}$-*endo* and $C_{3'}$-*endo*.

As shown previously for the principal species, there is a rotation at the junction between the helices analogous to P4 and P6 (Chastain & Tinoco, 1992). The normal NOE from the $H_{2'}$ sugar proton to the H_6 base proton is missing between C_{25} and U_6 in the short mixing time NOESY, although there is a normal NOE pathway on the other strand, between A_{13} and G_{14}. An NOE between the C_{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. Instead of C_{25} stacking on U_6 so that the two helices form a continuous helix through the junction, the two helices are rotated so that A_5 is positioned in the minor groove where it can interact with G_{14}·C_{25}.

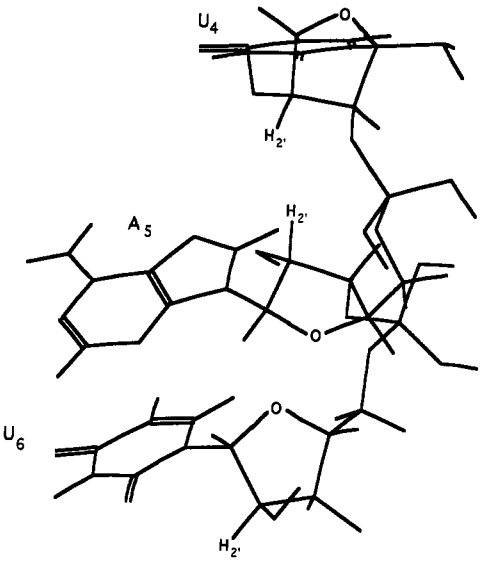

FIGURE 4: Positions of the three nucleotides, U_4, A_5, and U_6 are shown in order to illustrate the upside-down sugar of A_5. The orientation of the A_5 sugar is reversed relative to the sugar above and below it. Unusual NOEs from the A_5 sugar to both the U_4 and U_6 sugars showed that the sugar of A_5 is upside-down (Table 2). A similar upside-down nucleotide has been characterized in oligonucleotide models of internal loops from 5S rRNA (Wimberly et al., 1993) and 28S rRNA (Szewczak et al., 1993).

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 first four single-stranded nucleotides (5′-GAAU-) are 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 two nucleotides (5′-GA-) are 4–6 Hz, intermediate between $C_{2'}$-*endo* and $C_{3'}$-*endo* coupling constants. These coupling constants indicate that the first two sugars interconvert between the two sugar conformations, with approximately 50% of each conformer present. The sugar of the third nucleotide is in the $C_{3'}$-*endo* conformation. In the principal species, the first three single-stranded nucleotides were intermediate between $C_{2'}$-*endo* and $C_{3'}$-*endo*. The reduced number of sugars in the main variant interconverting between the two sugar conformations indicates that the main variant is kinetically more stable.

The final single-stranded nucleotide, A_5, is in an unusual conformation with its sugar turned upside-down relative to the nucleotides above and below it (U_4 and U_6) (Figure 4). The NMR evidence for this consists of several unusual NOEs from the sugar protons of A_5 to the sugar protons of U_4 and U_6. The sugar protons for A_5 were all assigned (except for the $H_{5'}$ and $H_{5''}$) from the COSY since the sugar is in the $C_{2'}$-*endo* conformation. The normal NOE from U_4 $H_{2'}$ to A_5 H_8 is missing in the short mixing time NOESY. Instead, there are NOEs from the A_5 $H_{2'}$ to the U_4 H_6 and $H_{2'}$ protons. There is also an NOE from the A_5 $H_{1'}$ proton to the U_6 $H_{1'}$ proton. No upside-down sugars occurred in the principal species. A nucleotide containing an upside-down sugar is found in internal loops from eukaryotic 5S rRNA (Wimberly et al., 1993) and 28S rRNA (Szewczak et al., 1993).

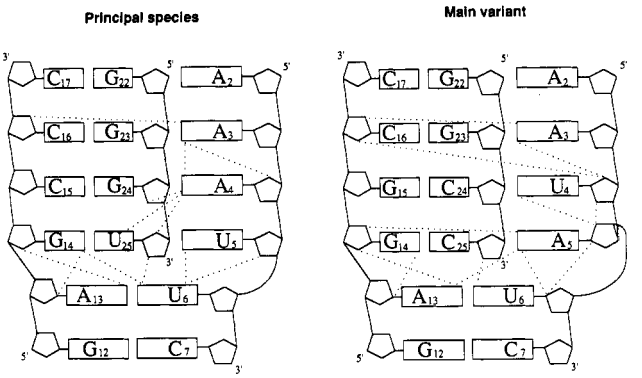

FIGURE 5: Dotted lines represent NOEs seen between protons on different nucleotides in the principal species and the main variant. The dotted lines do not indicate which protons on the sugar or base are involved in NOEs, but simply whether an NOE was seen to a base or sugar proton. A complete list of the unusual NOEs measured in the main variant is given in Table 2. Note that the NOEs from the adenine in the principal species, A_4, are to G_{24} whereas the NOEs from the adenine in the main variant, A_5, are to G_{14} (on the other side of the P6 helix from G_{24}).

Table 2: Unusual NOEs Measured

NOE measured	distance constraint range (Å)	NOE measured	distance constraint range (Å)
A_5 $H_{1'}$–U_6 $H_{1'}$	2.0–4.0	A_5 H_2–U_6 $H_{1'}$	3.0–5.0
A_5 $H_{2'}$–U_4 H_6	2.0–4.0	A_5 H_2–A_{13} H_2	not used
C_{25} $H_{2'}$–U_6 H_5	2.0–4.0	U_4 $H_{1'}$–C_{16} $H_{1'}$	3.0–5.0
A_5 H_2–G_{14} $H_{1'}$	2.0–4.0	A_3 H_2–C_{16} $H_{1'}$	3.0–5.0
A_5 H_2–G_{14} $H_{2'}$	3.0–5.0		

The pattern of NOEs between the single-stranded nucleotides and the duplex is very different in the main variant from the pattern in the previously characterized principal species (Figure 5). In both molecules, there is an NOE from the A_3 H_2 proton to the C_{16} $H_{1'}$ proton. This NOE suggests that the sugar phosphate backbone of the single-stranded nucleotides leaves the helix at a similar point in both structures. In the main variant, there is an NOE between the U_4 $H_{1'}$ sugar proton and the C_{16} $H_{1'}$ sugar proton. The U_4 imino exchanges too quickly to be seen in the imino spectrum, suggesting that it is not hydrogen bonded. The A_5 H_2 in the main variant has NOEs to G_{14} $H_{1'}$, G_{14} $H_{2'}$, A_{13} H_2, and U_6 $H_{1'}$ (Table 2, Figure 5). In the principal species, the NOEs occur between A_4 and one side of the helix (G_{24}), whereas in the main variant the NOEs occur between A_5 and the other side of the helix (G_{14}) (Figure 5). If A_5 is stacked in normal A-form geometry above U_6 in the main variant, it would be closer to C_{25} than to G_{14}. The $C_{2'}$-*endo* conformation and the kink in the backbone that turns the sugar of A_5 upside-down allow A_5 to interact with G_{14} on the other side of the helix. In both the principal species and the main variant, most of the NOEs between the single-stranded nucleotides and the duplex occur between adenine H_2 protons and sugar $H_{1'}$ protons in the minor groove of the helix analogous to P6. These NOEs show that the single-stranded nucleotides are in the minor groove close to the sugars. To determine how well the structure of the oligonucleotide is determined by these NMR constraints, molecular dynamics were performed.

Molecular Dynamics. Restrained molecular dynamics were done using distance and torsion angle constraints derived from the NMR data in order to find structures of the main variant consistent with the NMR data. Thirteen folded structures were generated from random starting structures by molecular dynamics using XPLOR (Brünger, 1990). None of the 13 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 13 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 based on all the atoms in the structure is 3.1 Å. If G_1 and A_2 are removed from the calculation, then the RMSD is lowered to 1.8 Å. The positions of these two nucleotides (G_1 and A_2) are not determined by the NMR data; however, the positions of the single-stranded nucleotides proposed to form triples, U_4 and A_5, are fairly well determined by the NMR data.

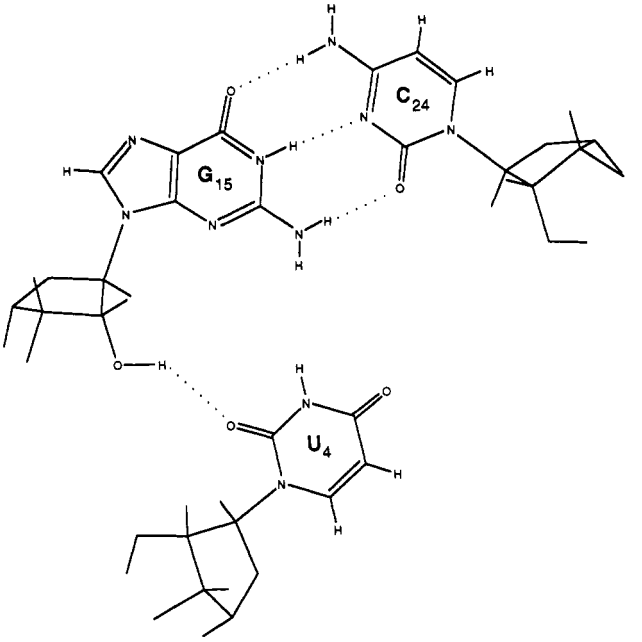

FIGURE 6: Positions of U_4, G_{15}, and C_{24} in the average structure calculated during molecular dynamics. The O_2 position of U_4 is within hydrogen bond distance of the 2′-hydroxyl oxygen of G_{15}.

The positions of U_4, G_{15}, and C_{24} in the average structure are shown in Figure 6. The three bases are approximately coplanar. The O_2 of U_4 can form hydrogen bonds with either the 2′-hydroxyl or the amino, or both groups of G_{15}. The average distance in the 13 different structures between the O_2 and the 2′-hydroxyl oxygen is 3.5 Å; the average distance between the O_2 and the amino nitrogen is 4.9 Å. Superimposing the U·(C·G) nucleoside triples in the 13 different structures based on the positions of U_4 and G_{15} shows that the position of U_4 relative to G_{15} is well-defined (Figure 7); the RMSD for these 3 nucleotides is 1.2 Å. The only direct NMR data constraining U_4 to be near G_{15} are the NOE from the U_4 sugar $H_{1'}$ proton to the C_{16} sugar $H_{1'}$ proton (Table 2); however, there are also NOEs between A_3 and C_{16} and between A_5 and the helix so the backbone is heavily constrained as well. In addition to the hydrogen bond between the O_2 and the 2′-hydroxyl, the structure shown in Figure 6 could be stabilized by a hydrogen bond from the U_4 imino proton to the N_3 position of G_{15}. The average distance between the U_4 imino nitrogen and the G_{15} N_3 is 4.2 Å. No evidence for the involvement of the uracil imino proton in a hydrogen bond could be obtained from the NMR spectrum since the uracil imino proton exchanged with the solvent too rapidly to be seen.

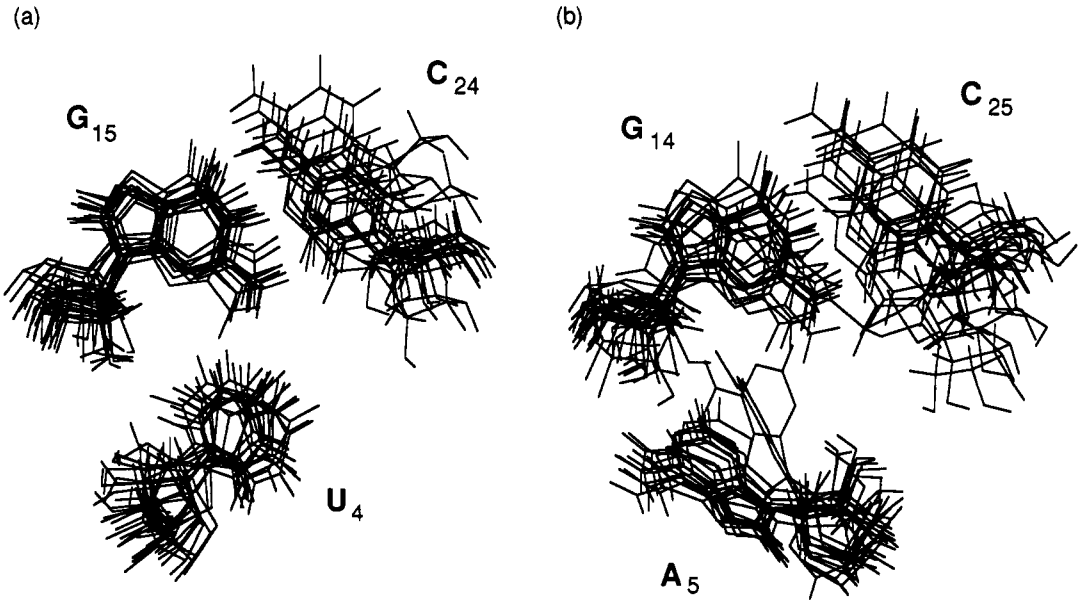

FIGURE 7: Positions of U_4, G_{15}, and C_{24} in the 13 structures generated during molecular dynamics are superimposed. The structures are superimposed based only on the positions of U_4 and G_{15} in order to show that the relative orientation of these nucleotides is well determined by the NMR data. (b) The positions of A_5, G_{14}, and C_{25} in the 13 structures generated during molecular dynamics are superimposed. The structures are superimposed based only on the positions of A_5 and G_{14}.

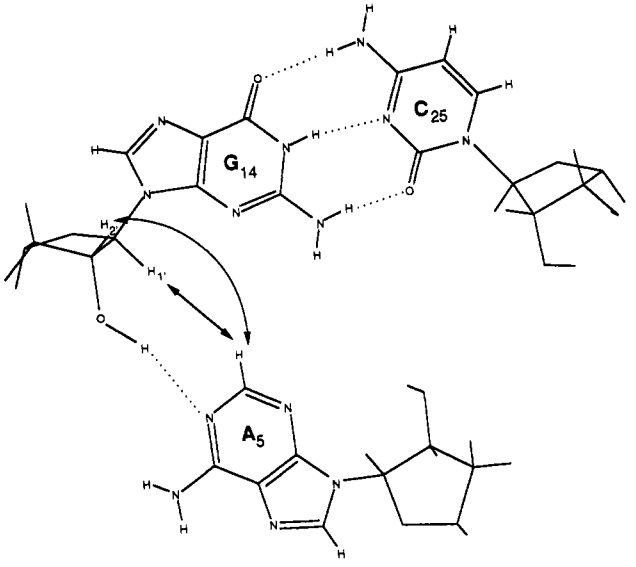

FIGURE 8: Positions of A_5, G_{14}, and C_{25} in the average structure calculated during molecular dynamics. The adenine is shown coplanar with the G·C base pair even though it is not completely coplanar in most of the structures. The N_1 position of the adenine is always within hydrogen bonding distance of the 2′-hydroxyl oxygen of G_{14}. Arrows are drawn between the protons for which NOEs were measured.

The A·(C·G) nucleoside triple characterized here is very different from the A·(G·C) nucleoside triple found in the principal species. Instead of being in position to hydrogen bond to the 2′-hydroxyl of C_{25}, the adenine interacts with G_{15} on the other side of the helix as shown in Figure 8. The N_1 position of the adenine is within hydrogen bonding distance of the 2′-hydroxyl of G_{14}. The average distance between them is 3.0 Å, and the largest distance is only 3.9 Å. The adenine is constrained to be near the sugar of G_{14} since there are NOEs from the A_5 H_2 to the G_{14} $H_{1'}$ and $H_{2'}$ (Table 2). In most of the structures, the adenine is not completely coplanar with the C·G base pair as can be seen in the superposition of the A·(C·G) nucleoside triple in the 13 different structures (Figure 7). The RMSD for these three nucleotides is 1.3 Å. The nucleoside triple could also be stabilized by a hydrogen bond between the G_{14} amino and the N_3 position of the adenine. The average distance between these groups in the 13 structures is 5.9 Å.

Mutant Oligonucleotides. In an attempt to understand the sequence specificity of nucleoside triples, NOESY spectra were taken of the four mutant oligonucleotides shown in Figure 2. The spectra of these molecules were not fully assigned, but the spectra were compared to the spectra of both the principal species and the main variant in order to determine whether the tertiary contacts between the single-stranded nucleotides and the duplex seen for those molecules were present in the mutants. If a similar pattern of tertiary NOEs was seen for a mutant, it was concluded that it formed nucleoside triples similar either to the principal species or to the main variant. The absence of tertiary NOEs led to the conclusion that nucleoside triples did not form, although in the absence of complete assignments there is a chance that the tertiary NOEs were missed due to overlap, or to a large change in chemical shift in the mutant relative to either the principal species or the main variant. The results of the NMR studies on the mutant oligonucleotides are summarized in Figure 2 where dotted lines between single-stranded nucleotides and the P6 helix indicate the formation of nucleoside triples.

The C·G mutant contains a single nucleotide change from the principal species. The U_{25}·G_{14} base pair of the U·(U·G) nucleoside triple was changed to a C·G base pair in order to determine if the U·G mismatch was required for triple formation. The NOESY spectrum of the C·G mutant is shown in Figure 9a. An arrow marks the cross-peak from the A_4 H_2 to the G_{24} $H_{1'}$ proton. In a second NOESY taken at 23 °C, a cross-peak was also seen to the C_{25} $H_{1'}$ proton (data not shown). A cross-peak to the G_{24} $H_{2'}$ proton was not seen at either temperature. This pattern of NOEs is similar to the pattern seen for the principal species (Chastain & Tinoco, 1992), indicating that an A·(G·C) nucleoside triple similar to the one found in the principal species forms in the C·G mutant. The formation of the A·(G·C) nucleoside triple in the C·G mutant suggests that the structures of the U·(U·G) and the U·(C·G) triples do not differ very much. In most of the structures generated for the principal species, U_5 was within

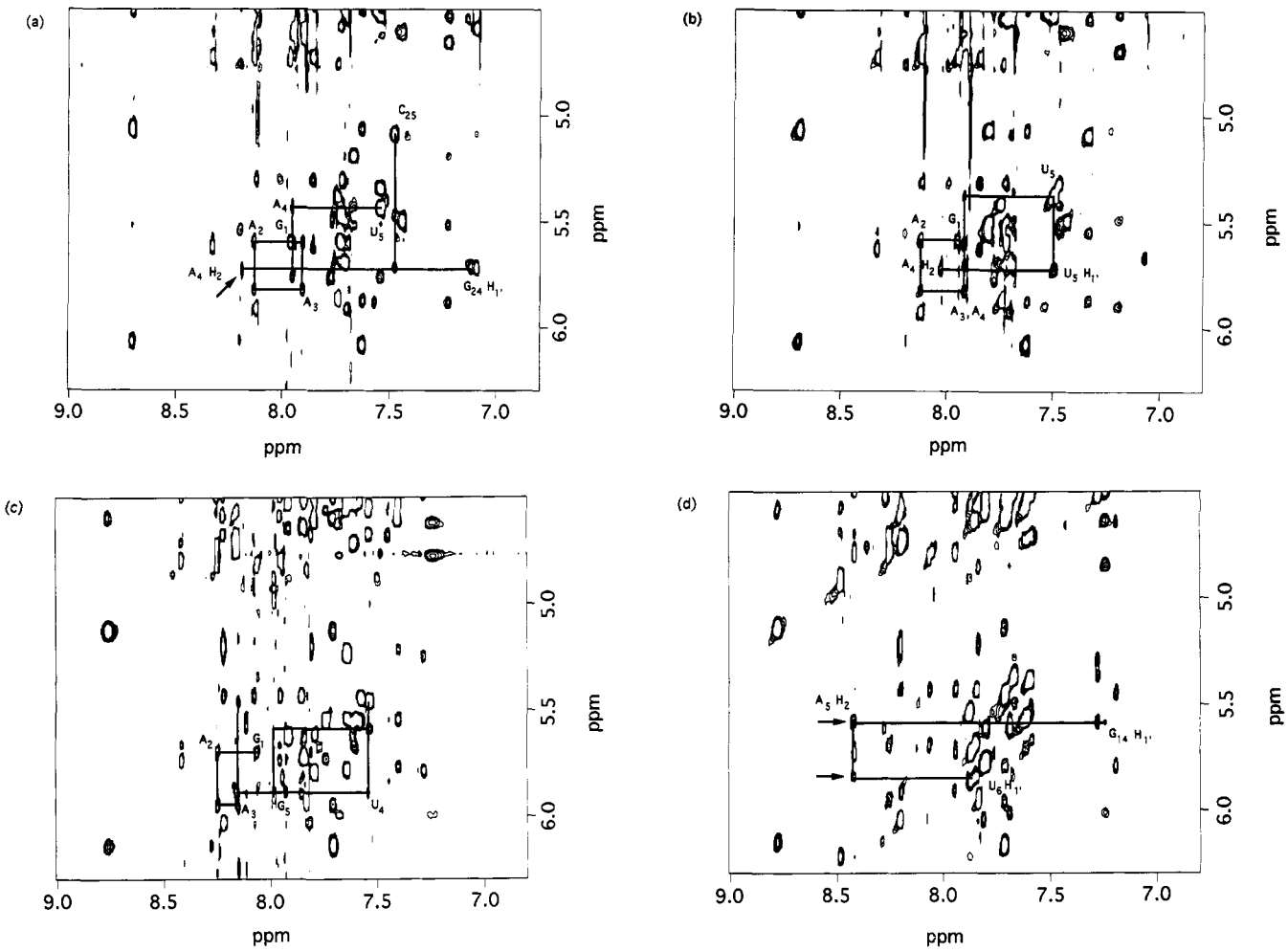

FIGURE 9: NOESY spectra showing the $H_8/H_6/H_2$ to $H_{1'}/H_5$ region taken on the mutant oligonucleotides. Spectra were taken in 10 mM sodium phosphate (pH 6.7)/0.1 mM EDTA. The spectra of the C·G and A·U mutant samples were taken at 15 °C, whereas the other two spectra were taken at 23 °C. (a) C·G mutant: The NOE pathway for the single-stranded nucleotides is shown connected by lines. In addition, the cross-peak from the A_4 H_2 to the G_{24} $H_{1'}$ is marked by an arrow. This cross-peak shows that nucleoside triples form in this oligonucleotide. (b) A·U mutant: The NOE pathway for the single-stranded nucleotides is shown connected by lines. The cross-peak from the A_4 H_2 to the U_5 $H_{1'}$ is also shown. No NOEs are seen from the single-stranded nucleotides to the helix. (c) G mutant: The NOE pathway for the single-stranded nucleotides is shown connected by lines. No NOEs are seen from the single-stranded nucleotides to the helix. (d) U·G mutant: The cross-peaks marked by arrows from the A_5 H_2 to the G_{14} and U_6 $H_{1'}$ protons show that a nucleoside triple forms in this nucleotide.

hydrogen bonding distance of either the U_{25} 2′-hydroxyl group or the G_{14} amino group (Chastain & Tinoco, 1992). The formation of the A·(G·C) triple in the C·G mutant suggests that the G_{14} amino is not required to stabilize the U·(U·G) or U·(C·G) triples since the G_{14} amino is presumably hydrogen bonded to C_{25} in the C·G mutant.

The A·U mutant contains three base changes from the principal species. The G_{15}·C_{24} base pair involved in the A·(G·C) nucleoside triple was changed to an A·U base pair to see if an A·(A·U) nucleoside triple could form. The U·G base pair in the U·(U·G) nucleoside triple was also changed to a C·G base pair to avoid alternative base pairing. As shown above for the C·G mutant, nucleoside triples still form when this change is made. The expected effect of changing the G_{15}·C_{24} base pair to an A·U base pair was that a nucleoside triple would still form since the nucleoside triple was thought to be stabilized by a hydrogen to the 2′-hydroxyl. The NOESY spectrum of the A·U mutant is shown in Figure 9b. There is an NOE from an H_2 proton, presumably on A_4, to the U_5 $H_{1'}$ proton, but the NOEs to $H_{1'}$ protons in the minor groove of the helix are not seen. The lack of NOEs to $H_{1'}$ protons in the minor groove implies that changing the G_{15}·C_{24} base pair to A·U abolishes the formation of nucleoside triples. The cross-peak which we have tentatively assigned to A_4 H_2 and the U_5 $H_{1'}$ suggests that the NOEs to $H_{1'}$ protons of the helix were not missed simply due to spectral overlap or to a large change in chemical shift of the A_4 H_2 proton. The apparent lack of nucleoside triple formation in the A·U mutant suggests that the A·(G·C) nucleoside triple in the principal species is stabilized by additional interactions between the adenine and the guanine besides a hydrogen bond to the 2′-hydroxyl. The stabilization of the A_4·(G_{15}·C_{24}) nucleoside triple could be due either to the stacking of A_3 (the single-stranded nucleotide following the nucleoside triple) over the G_{15} amino or to hydrogen bonding between G_{15} and A_4. The hydrogen bond that is easiest to make without violating the NOE constraints is a hydrogen bond between the guanine amino proton and the N_3 position of the adenine.

The G mutant of the main variant changes the single-stranded adenine that forms the A·(C·G) nucleoside triple to a guanine. Since guanine, unlike adenine, has a hydrogen bond donor at the N_1 position, the expected effect of this mutant was that it would no longer form nucleoside triples. Since the NOEs seen between the single-stranded adenine and the duplex were to the adenine H_2 proton which the guanine mutant does not have, other NOEs in the spectrum were examined to see whether the G mutant formed nucleoside triples. The NOEs between the A_3 H_2 and the duplex were

missing in the G mutant (Figure 9c). The NOEs between the U_4 $H_{1'}$ and the duplex and the G_5 $H_{1'}$ and the duplex were missing as well. In addition, the chemical shift of G_{15} in the main variant is unusually upfield at 6.9 ppm. In the G mutant, the G_{15} chemical shift changes to 7.3 ppm. All of these changes indicate that the G mutant oligonucleotide does not form nucleoside triples; this is consistent with the A·(C·G) nucleoside triple containing a hydrogen bond to the N_1 position of the adenine.

The last mutant oligonucleotide characterized is the U·G mutant of the main variant. In this mutant, the C·G base pair of the U_4·(C_{24}·G_{15}) nucleoside triple was changed to a U·G base pair. This mutant was expected to form nucleoside triples since the NOE measured between U_4 and C_{16} (Table 2) constrains U_4 to be near G_{15}, not C_{24}. This mutation was made in order to make the main variant more similar to the principal species. The nucleoside triples in the main variant are U_4·(C_{24}·G_{15}) and A_5·(C_{25}·G_{14}); both of these triples are different from the ones in the principal species, A_4·(G_{24}·C_{15}) and U_5·(U_{25}·G_{14}). The U·G mutant makes one of the nucleoside triples the same in order to see the effect on the other triple. The NOESY spectrum in Figure 9d shows that the NOEs between the A_5 H_2 proton and the duplex are still present in the U·G mutant. However, the NOEs between U_4 and the duplex and A_3 and the duplex are absent. In addition, the chemical shift of G_{15} moves from 6.9 to 7.3 ppm. These facts suggest that the U_4·(U_{24}·G_{15}) nucleoside triple does not form in the U·G mutant although the A·(C·G) nucleoside triple still forms. The lack of a U·(U·G) nucleoside triple suggests either that U_4 interacts with C_{24} as well as G_{15} in the main variant or that the structure of the U·G mismatch is different enough from the C·G base pair that the nucleoside triple is destabilized.

DISCUSSION

Minor groove base triples were predicted to form in the group I intron on the basis of phylogenetic comparison and model building (Michel & Westhof, 1990). Previously, we have shown that an oligonucleotide containing the most frequently occurring sequence at the junction of the P4 and P6 helices formed A·(G·C) and U·(U·G) nucleoside triples (Chastain & Tinoco, 1992). Here we have shown that an oligonucleotide containing the second most frequently occurring sequence at the P4 and P6 helices forms U·(C·G) and A·(C·G) nucleoside triples. The structures we have determined for these two oligonucleotides are different than the structures proposed by Michel and Westhof, although the formation of tertiary interactions in the minor groove at the P4/P6 junction is consistent with their proposal. Surprisingly, the structures of the nucleoside triples in these two oligonucleotides are very different. Mutant oligonucleotides with slight changes in sequence from the first two oligonucleotides show that nucleoside triple formation is sequence-specific. Sequences which contain G·C and/or G·U base pairs in the P6 helix and adenine or uracil in the single-stranded region tend to form nucleoside triples whereas A·U base pairs in the P6 helix or guanine in the single-stranded region prevents nucleoside triple formation (Figure 2).

Although the hydrogen bonding and NMR constraints used are insufficient to completely determine the structure of the oligonucleotides containing sequences from the P4/P6 junction, several aspects of the structures are well-defined. The two helices in these oligonucleotides are coaxially stacked, but there is a large rotation between the two helices (approximately twice the rotation between normal base pairs) which allows the single-stranded nucleotides to interact with base pairs in the minor groove of the P6 helix. The NOEs measured show that the single-stranded nucleotides are close to sugars in the minor groove. Structures generated by molecular dynamics calculations constrained by the NMR data show that the single-stranded nucleotides are within hydrogen bonding distance of 2′-hydroxyl groups. There is no direct NMR evidence for these hydrogen bonds, but the limited variation in the structures generated by molecular dynamics suggests that the NMR data are sufficient to constrain the molecule to structures in which these hydrogen bonds form. The limited variation in the nucleoside triples in the generated structures also suggests that the unusual, nonsequential NOEs measured between adenine H_2 protons and sugar protons in the minor groove are more restrictive than more typically measured NOEs between adjacent nucleotides. The power of nonsequential NOEs to define RNA structure shown here together with the development of methods for labeling RNA molecules with ^{13}C (Batey et al., 1992; Nikonowicz et al., 1992) implies that NMR will determine the important structural features of larger labeled RNA molecules that contain tertiary interactions.

Surprisingly, the structure of the main variant oligonucleotide characterized here is very different from the structure of the principal species oligonucleotide characterized previously (Chastain & Tinoco, 1992). In the main variant, the first single-stranded nucleotide at the junction of the two helices, A_5, is in an unusual conformation in which its sugar (in the $C_{2'}$-*endo* conformation) is turned upside-down. This unusual backbone conformation allows A_5 to interact with G_{14} rather than with C_{25} as would be expected from the structure of the principal species. The difference between the two structures could be explained by stacking between the helix analogous to P4 and the adjacent single-stranded nucleotide. In the principal species, this nucleotide, U_5, is a pyrimidine whereas in the main variant this nucleotide, A_5, is a purine. In both molecules, the two helices are coaxially stacked with a large rotation between the two helices. This rotation allows the single-stranded nucleotides to stack onto the P4 helix and interact with the P6 helix in the minor groove. Since the purine in the main variant A_5 is larger than the pyrimidine U_5 in the principal species, there may not be room, even with the rotation between the helices, for the two helices to stack coaxially and for the adenine to stack normally above the P4 helix. Instead, there is a kink in the backbone at A_5 which turns the sugar of A_5 upside-down, allowing A_5 to interact with G_{14} on the other side of the helix. Another possible explanation for the difference in structures is that the guanine amino group stabilizes the nucleoside triple. This implies that changing from the A·(G·C) nucleoside triple in the principal species to the A·(C·G) triple in the main variant would require the adenine to move to the other side of the helix to maintain the contact between the adenine and the guanine amino.

Formation of nucleoside triples in two of the four mutant oligonucleotides shows that there is sequence specificity to nucleoside triple formation. Since we propose that the nucleoside triples are stabilized by hydrogen bonds to 2′-hydroxyl groups, sequence specificity of the nucleoside triples is somewhat surprising. However, sequence specificity, as discussed below, is required to explain the phylogenetic results if nucleoside triples form in the context of the whole group I intron. There are several possible explanations for the sequence specificity. One possibility is that changing the helix sequence changes the helix structure enough that the 2′-

hydroxyl groups are no longer positioned so that they can form nucleoside triples. Another explanation is that stacking between the single-stranded nucleotides in the minor groove and the base pairs in the helix analogous to P6 accounts for the sequence specificity. Finally, the specificity could be a result of additional hydrogen bonds between the single-stranded nucleotides and base pairs in the P6 helix. Guanine amino groups in the minor groove of the helix analogous to P6 could form hydrogen bonds to the N_3 position of the single-stranded adenines to stabilize the nucleoside triples.

The sequence-specific formation of nucleoside triples in the oligonucleotides containing sequences from the P4/P6 junction does not prove that these interactions form within group I introns. Other tertiary interactions in the intron, such as the proposed base triples in the major groove of helix P4 (Michel et al., 1989), could prevent, or change, the formation of nucleoside triples. In addition, the length of the P6 helix in the model oligonucleotides has been lengthened to four base pairs from two base pairs in the intron, and the length of the P4 helix has been shortened to two base pairs from six base pairs in the intron. Two lines of evidence suggest that nucleoside triples similar to the ones characterized here form within the intron. First, mutational studies on the intron show that changing the sequence of base pairs within the P6 helix greatly reduces the activity of the intron (Ehrenman et al., 1989; Couture et al., 1990). This strongly suggests that the P6 helix is involved in tertiary interactions. Second, oligonucleotides which contain phylogenetically conserved sequences at the P4/P6 junction form nucleoside triples, but oligonucleotides containing sequences which are not conserved do not form triples. An intron containing the sequence at the P4/P6 region corresponding to the C·G mutant oligonucleotide was shown to have 40% of the activity of the wild-type intron (Couture et al., 1990). This is consistent with our finding that the C·G mutant forms nucleoside triples. The correlation between sequences in the P4/P6 region which form nucleoside triples in oligonucleotides and which have activity in the context of an entire intron implies that the intron contains nucleoside triples similar to the ones described here.

If nucleoside triples form in the intron, how do they contribute to its function? None of the biochemical evidence concerning the mechanism of intron function has suggested that the P6 helix is directly involved in catalysis. This is consistent with the fact that the oligonucleotides containing the two most frequently occurring sequences at the P4/P6 junction form different structures. If the nucleoside triples were directly involved in the catalysis, one would expect that their structures would be very similar. Instead, the triples must serve an indirect functional role by orienting other parts of the intron which actually catalyze the reaction. The formation of nucleoside triples orients the P4 helix relative to the P6 helix. In addition, the formation of triples constrains the orientation of the helical domain consisting of P4 and P6 relative to the domain formed by P3, P7, and P8 (Michel & Westhof, 1990). The orientation of P4 relative to P6 could be important during the 5′ splicing step since the internal loop P6a is proposed to interact with the P1 helix during this step (Michel & Westhof, 1990). The large rotation between the P4 and P6 helices caused by the formation of nucleoside triples would orient the P6a loop/P1 domain relative to the P7 helix which contains the guanosine binding site (Michel et al., 1990). Studies on tRNA synthetase binding to the core region of the intron also led to the proposal that the P4/P6 triples oriented the P7 helix relative to the P4/P6 helices (Guo & Lambowitz, 1992). The nucleoside triples described here presumably play a large role in this orientation.

ACKNOWLEDGMENT

We thank Mr. David Koh for synthesizing DNA templates and Ms. Barbara Dengler for managing the laboratory. We thank Dr. Jennifer Hines, Mr. Marco Molinaro, and Ms. Ling Shen for their help in synthesizing the C·G mutant oligonucleotide.

REFERENCES

Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D., & Williamson, J. R. (1992) *Nucleic Acids Res. 20*, 4515–4523.
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) *Biochemistry 31*, 12733–12741.
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.
Ehrenman, K., Schroeder, R., Chandry, P. S., Hall, D. H., & Belfort, M. (1989) *Nucleic Acids Res. 17*, 9147–9163.
Guo, Q., & Lambowitz, A. M. (1992) *Genes Dev. 6*, 1357–1372.
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.
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., 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.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585–610.
Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.
Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M., & Pardi, A. (1992) *Nucleic Acids Res. 20*, 4507–4513.
Pleij, C. W. A., Rietveld, K., & Bosch, L. (1985) *Nucleic Acids Res. 13*, 1717–1731.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
Szewczak, A. A., Moore, P. B., Chan, Y. L., & Wool, I. G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Varani, G., & Tinoco, I., Jr. (1991) *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.
Wimberly, B., Varani, G., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 1078–1087.
Wyatt, J. R., Chastain, M., & Puglisi, J. D. (1991) *BioTechniques 11*, 764–769.