Solution Structure of $(rGCGGACGC)_2$ by Two-Dimensional NMR and the Iterative Relaxation Matrix Approach†

Ming Wu and Douglas H. Turner*

Department of Chemistry, University of Rochester, Rochester, New York 14627-0216

Received January 18, 1996; Revised Manuscript Received May 17, 1996⊗

ABSTRACT: The three-dimensional solution structure of the RNA self-complementary duplex $\begin{smallmatrix}GCG^{GA}CGC\\CGC_{AG}GCG\end{smallmatrix}$ was derived from two-dimensional NMR and the iterative relaxation matrix approach. Each GA mismatch forms two hydrogen bonds: A-NH6 to G-O6 and A-N1 to G-NH1 (imino). This is the first three-dimensional RNA structure with imino hydrogen-bonded tandem GA mismatches. This GA structure is totally different from the sheared tandem GA structure in $\begin{smallmatrix}GGC^{GA}GCC\\CCG_{AG}CGG\end{smallmatrix}$ which also has two hydrogen bonds: A-N7 to G-NH2 and A-NH6 to G-N3 [SantaLucia, J., Jr., & Turner, D. H. (1993) *Biochemistry 32,* 12612−12623]. In particular, the sheared and imino GA mismatches produce a narrowing and widening of the backbone, respectively. The results show that substitutions of Watson−Crick base pairs can have dramatic effects on the three-dimensional structures of adjacent non-Watson−Crick paired regions; i.e., the structure depends on sequence context. Thus compensating substitutions in site-directed mutagenesis experiments may not always restore biological activities.

Predicting secondary and three-dimensional RNA structures requires both thermodynamic and three-dimensional structural information on various motifs of RNA (Tinoco et al., 1971; Turner et al., 1988). As more thermodynamic parameters are becoming available (Turner et al., 1988; Walter et al., 1994a; Wu et al., 1995), structural information becomes more and more important for understanding the thermodynamic stability and structural implications of various RNA motifs. The GA mismatch is an important RNA motif because of its prevalence (Gautheret et al., 1994; Gutell et al., 1992, 1993), structural variations (Walter et al., 1994b; SantaLucia et al., 1990; SantaLucia & Turner, 1993; Ebel et al., 1994), and biological functions. GA mismatches in RNA serve as sites for tertiary interaction (Michel & Westhof, 1990; Pley et al., 1994a,b; Murphy & Cech, 1994; Biou et al., 1994), metal binding (Pley et al., 1994a), and protein recognition (Zwieb, 1992).

Previous NMR studies of the RNA tandem GA mismatch motif $\begin{smallmatrix}5'GA3'\\3'AG5'\end{smallmatrix}$ show that the tandem GA mismatch conformation depends on adjacent base pairs (Walter et al., 1994b). Two-dimensional NMR studies of $(rGGCGAGCC)_2$ (SantaLucia & Turner, 1993) show that the GA mismatches have sheared conformations (Figure 1), with two hydrogen bonds between G and A: G-NH2 to A-N7 and G-N3 to A-H6. On the basis of NMR chemical shifts and an NOE from G4NH1 to A5H2, however, the sequence $(rGCGGACGC)_2$ has GA mismatches with hydrogen bonds involving the imino proton of G (Walter et al., 1994b). This paper reports a full three-dimensional solution structure of $(rGCGGACGC)_2$ determined by two-dimensional NMR and the iterative relaxation matrix approach (IRMA). The structural differences between $(rGCGGACGC)_2$ and $(rGGCGAGCC)_2$ are striking. Thus substitution of Watson−Crick base pairs adjacent to non-Watson−Crick paired regions can have dramatic effects on the three-dimensional shape of an RNA molecule.

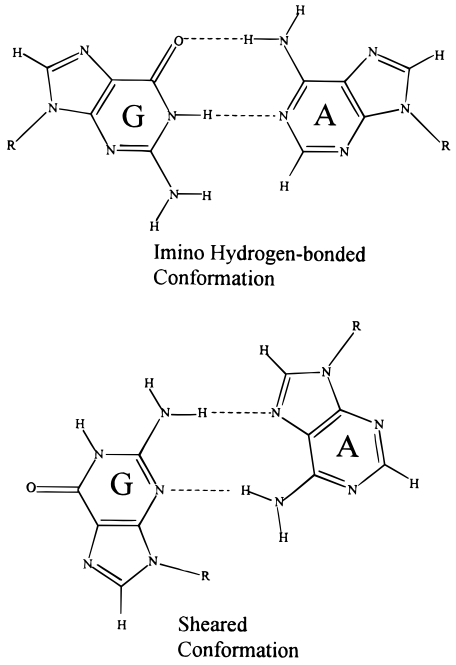

FIGURE 1: Two known hydrogen-bonding patterns of GA mismatches in a GA tandem: (top) imino hydrogen-bonded conformation; (bottom) sheared conformation.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Purification. rGCGGACGC was synthesized on an Applied Biosystems 392 DNA/RNA synthesizer (Usman et al., 1987; Scaringe et al., 1990). RNA monomers were from Applied Biosystems, Inc., with 2′-hydroxyls protected as the *tert*-butyl dimethylsilyl ether (Ogilvie et al., 1977) and 5′-hydroxyls protected as the dimethoxytrityl group. Upon completion of coupling on the synthesizer, the oligomer was removed from solid support and deprotected by treatment with concentrated ammonia in ethanol (3:1 by volume). The 2′-silyl protection was

† This work was supported by NIH Grant GM 22939.
* Author to whom correspondence should be addressed.
⊗ Abstract published in *Advance ACS Abstracts,* July 15, 1996.

S0006-2960(96)00133-X CCC: $12.00

removed by treatment with freshly made 1 M triethylammonium hydrogen fluoride (50 equiv) in pyridine at 55 °C for 48 h. The crude sample was then dried and partitioned between water and diethyl ether. Then the sample was dissolved in 10 mL of 5 mM ammonium acetate (pH 7) and passed through a Sep-pak C-18 column (Waters) to desalt. The sample was purified on a Si500F thin-layer chromatography plate (Baker) and eluted with 1-propanol−ammonia−water (55:35:10 by volume). The least mobile main band was visualized with ultraviolet light, cut out, and eluted with water. The sample was desalted again with a Sep-pak C-18 cartridge. The purity of the rGCGGACGC was checked by HPLC on a C-8 analytical column (Hamilton) and was greater than 99%. The RNA was dialyzed against 0.1 mM EDTA and then doubly distilled water for 24 h each, dried down, and dissolved in 0.5 mM EDTA, 10 mM sodium phosphate, 0.1 M NaCl, and 0.5 mM TSP [3-(trimethylsilyl)-propionate, as an internal reference for proton spectra], pH 7. The sample was lyophilized to dryness and then lyophilized twice from 99.96% D_2O and once from 99.996% D_2O. The sample was dissolved, under dry nitrogen, in 0.6 mL of 99.996% D_2O (or in 0.54 mL of H_2O and 0.06 mL of D_2O, for experiments in 90% H_2O and 10% D_2O). The strand concentration was 3 mM.

NMR Spectroscopy. NMR spectra were acquired on a Varian VXR-500S spectrometer and processed by a Sun 4/260 computer running Varian VNMR software and a Silicon Graphics workstation running Biosym Felix software. All 2-D spectra were collected in the phase-sensitive mode using the hypercomplex methods (States et al., 1982). Proton and phosphorus chemical shifts were referenced to TSP [3-(trimethylsilyl)propionate] and phosphate buffer, respectively.

Exchangeable proton spectra were recorded in 90% H_2O and 10% D_2O by using the 1331 solvent suppression scheme (Hore, 1983) with 12 000 Hz sweep width. One-dimensional NOE experiments were acquired by irradiating for 3 s with a low-power decoupler setting prior to the read pulses. The maximum excitation was adjusted to 12.6 ppm with the first nodes at 20.4 and 4.8 ppm.

A NOESY-JRE (Sklenar et al., 1987; Sklenar & Feigon, 1990) spectrum in 90% H_2O and 10% D_2O at 15 °C was collected with a mixing time of 150 ms. The excitation profile was adjusted to maximize excitation in both imino (12.9 ppm) and amino (7.6 ppm) regions of the proton spectrum, with the delay between the two read pulses set to 188 μs. For each FID, 400 transients were acquired with 2048 complex points over a spectral width of 10 000 Hz using a recycle delay of 2.5 s. A total of 200 FIDs were collected.

NOESY spectra in D_2O were acquired with mixing times of 100 and 400 ms at 35 and 40 °C. Additional NOESY spectra were collected with mixing times of 25, 50, 150, 250, and 400 ms at 35 °C for quantification of the NOE intensities in molecular modeling. Attenuation of the residual HDO resonance was achieved by low-power presaturation during the relaxation delay. For each FID, 32 transients were acquired with 2048 complex points over a spectral width of 4000 Hz using a recycle delay of 5 s. A total of 256 FIDs were collected.

DQF-COSY spectra were collected at 35 and 40 °C. For each FID, 96 transients were acquired with 2048 complex points over a spectral width of 4000 Hz using a recycle delay of 2.5 s. A total of 416 FIDs were collected.

A $^1H-^{31}P$ HETCOR spectrum at 35 °C was collected using the pulse sequence described by Sklenar et al. (1986), with a spectral width of 2000 Hz in the ^1H dimension and 800 Hz in the ^{31}P dimension. For each FID, 400 transients were collected with 2048 complex points using a recycle delay of 2.5 s. A total of 222 FIDs were acquired.

Restraint Generation and Structural Modeling. All nonexchangeable protons were assigned on the basis of careful comparisons among T_1 measurements, DQF-COSY, $^1H-^{31}P$ HETCOR, and TOCSY experiments, along with NOESY spectra with different mixing times and at different temperatures, using strategies described by Varani and Tinoco (1991). Dihedral angle restraints were derived by measuring J coupling constants from DQF-COSY and $^1H-^{31}P$ HETCOR spectra. J coupling constants were measured by fitting the antiphase cross peaks automatically using Felix software. The equations used for converting the J coupling constant to the dihedral angle θ are J_{HCCH} (Hz) $= 10.2 \cos^2 \theta - 0.8 \cos \theta$ (Davies, 1978; Hosur et al., 1988) and J_{HCOP} (Hz) $= 15.3 \cos^2 \theta - 6.1 \cos \theta + 1.6$ (Lankhorst et al., 1984). Generally, one J value will give four possible θ values. Among these four, the one that is closest to the A-form θ value was used in generating dihedral angle restraints. With all the dihedral angle restraints *turned off* during minimization, the dihedral angles of the minimized structures derived using only distance restraints were close to the restrained A-form values. Thus biasing these dihedral angle restraints toward A-form values seems reasonable. The bound widths of dihedral angle restraints were calculated from errors of J measurements. The ϵ and δ angles were derived from H3′−P and H3′−H4′ J coupling constants, respectively. The β and γ angles were determined by P−H5′, P−H5″ and H4′−H5′, H4′−H5″ J coupling constants. The nonstereospecificity of H5′ and H5″ assignments prohibits exact determinations of β and γ angle restraints. This problem was circumvented by biasing β and γ angles toward A-form values and making restraint bounds so generous that they encompassed the range calculated for either possible assignment of H5′ and H5″.

NOESY-JRE and NOESY spectra indicate all six potential GC pairs form in $(rGCGGACGC)_2$. Therefore, 18 distance restraints (1.8 ± 0.2 Å) were used to maintain Watson−Crick hydrogen bonding between the six GC base pairs. No hydrogen-bonding restraints were imposed between mismatched G and A. Six additional imino distance restraints (4.0 ± 1.0 Å) were used, based on the NOEs in NOESY-JRE spectra.

A total of 192 (per duplex) initial NOE distance restraints between nonexchangeable protons were obtained by measuring the NOESY (the average of 50 ms and 150 ms) cross peak volumes, comparing with the cytidine H5−H6 (2.45 Å) cross peak volumes, and scaling by $1/r^6$, using the isolated two-spin approximation. The bounds were set to ±20% of the distances. In large molecules like $(rGCGGACGC)_2$, however, the two-spin approximation can introduce errors in distances (Boelens et al., 1988; Gorenstein et al., 1990; Nikonowicz et al., 1991), since indirect magnetization transfer via other protons (i.e., spin diffusion) can contribute to NOE intensities. The iterative relaxation matrix approach (IRMA; Boelens et al., 1988) takes spin diffusion into

account in calculating the distances, so it generates more accurate distance restraints. During this iterative process, the distance restraints are refined and differences between the experimental NOE intensities and back-calculated NOE intensities serve as criteria for convergence of the structure. Between IRMA calculations, a simulated annealing protocol including restrained molecular dynamics (rMD) and restrained energy minimization (rEM) was executed to minimize the structure based on 170 refined distance restraints per duplex, while all 46 dihedral angle restraints, 6 imino−imino proton distance restraints, 18 hydrogen-bonding distance restraints, and 16 NOE distance restraints (from overlapped, or weak and noisy cross peaks) were kept constant. IRMA, rMD, and rEM yield refined structures that converge and result in a best fit with the experimentally observed NOE intensities.

Simulated annealing was performed with the Discover package within Biosym InsightII. A total of 210 distance restraints and 46 dihedral angle restraints per duplex were used in the modeling. Although the duplex sequence has 2-fold symmetry, no symmetry constraint was imposed. A recent crystal structure of an RNA internal loop with 2-fold symmetry in sequence did not have 2-fold symmetry in structure (Baeyens et al., 1995). Calculations employed the Amber force-field (Weiner et al., 1986) in vacuum, excluding counterions, solvent molecules, and electrostatic charge interactions. Covalent bonding terms included deformation energies of bond lengths, bond angles, and torsion angles, with force constants depending on the type of bonds and atoms. Nonbonding terms included van der Waals energies with a Leonard-Jones function and an atom pair distance cutoff of 12 Å. Energy terms for distance restraints and dihedral angle restraints were added to the force field, using flat-bottomed quadratic potentials. The force constants for NOE and dihedral angle restraints were set to 10 kcal/(mol Å^2) and 50 kcal/(mol rad^2), respectively. The simulated annealing protocol was (a) 200 iterations of rEM, (b) 6 ps of rMD at 1000 K, (c) 6 ps of rMD with temperature decreased from 1000 to 300 K gradually, (d) 3 ps of rMD at 300 K, (e) 500 iterations of steepest descent rEM, and (f) 2000 iterations of conjugate gradient rEM. The multiplication factors for force constants of covalent and nonbonding terms were increased gradually from 0.1 and 0.001, respectively, to 1 from the beginning of step b to the end of step d. The force constants for distance restraint terms were increased gradually from 5 to 10 kcal/(mol Å^2) [20 kcal/(mol Å^2) before application of IRMA] from the beginning of step b to the end of step c. The multiplication factor for dihedral angle restraint terms was maintained at 0.001 for steps b and c and increased gradually to 1 during step d. The differences among force constant buildup profiles for different force field terms were designed to ensure the best agreement between resulting structure and restraints and to prevent the molecule from being trapped in unreasonably distorted conformations.

Three starting structures (A-form duplex, B-form duplex, and A-form with two separated single strands) were generated with InsightII. First, they were minimized by the initial restraints and a total of 15 structures were generated, 5 from each starting structure with randomized trajectories during the rMD and rEM. Fourteen of 15 structures generated overlapped well and resembled the eventual final structure. The distance restraints were refined using the relaxation matrix approach (RMA): the theoretical relaxation matrix *A* was calculated from the ensemble average of the 14 converged structures. The theoretical NOE intensity matrices calculated from the theoretical relaxation matrix *A* for each mixing time (25, 50, 150, 250, and 400 ms) were merged with the corresponding experimental NOE intensity matrices. The relaxation matrix for each mixing time was back-calculated from the merged NOE intensity matrices, and an average back-calculated relaxation matrix B was derived. The refined distances and restraint bounds are derived from matrix *B* and the standard deviation of back-calculated relaxation matrices, respectively. The simulated annealing protocol was executed again to generate 15 new structures, with the *refined* distance restraints. Then the distance restraints were refined again using RMA. A series of *R*-factors were derived in each cycle, representing differences between experimental and theoretical NOE intensities. This iterative process was repeated until the decrease of *R*-factors leveled off. The most important *R*-factors are defined as $R_1 = \sum|A^{theo} - A^{expt}|/\sum A^{expt}$, $R_2 = \sum\tau_m|A^{theo} - A^{expt}|/\sum\tau_m A^{expt}$, $R_3 = \sum|A^{theo} - A^{expt}|/\sum 0.5(|A^{theo}| + |A^{expt}|)$, and $R_4 = \sum\tau_m|A^{theo} - A^{expt}|/\sum\tau_m 0.5 \times (|A^{theo}| + |A^{expt}|)$, where A^{expt} and A^{theo} are the NOE intensities of experimental and theoretical NOE matrices and τ_m is the mixing time.

RESULTS

Assignment of Exchangeable Protons. The 1-D imino proton spectrum and NOE difference spectra at 15 °C are shown in Figure 2. rGCGGACGC forms a self-complementary duplex, $\begin{smallmatrix}GCG^{GA}CGC\\CGC_{AG}GCG\end{smallmatrix}$. The four resonances in the 12−14 ppm region belong to G1, G7, G3, and G4 imino protons (see Figure 2). When the temperature is raised from 0 to 40 °C, the resonance at 13.01 ppm first sharpens due to reduced aggregation and then broadens due to duplex melting. Its line width is always broader than that of the other three imino resonances. This is characteristic of the imino resonance of a terminal nucleotide. Thus the resonance at 13.01 ppm is assigned to G1H1. From the NOE difference spectra in Figure 2, it is clear that NOEs exist between resonances at 12.52 and 12.81 ppm and between resonances at 13.23 and 12.81 ppm but not between resonances at 12.52 and 13.23 ppm. Thus the resonance at 12.81 ppm is assigned to G3H1 because G3H1 is between G4H1 and G7H1 in space and therefore should exhibit two NOEs. Irradiating the resonance at 12.51 ppm gives rise to an NOE at 7.76 ppm. An NOE between resonances at 12.51 and 7.76 ppm is also observed in a NOESY-JRE spectrum (Figure 3). The resonance at 7.76 ppm is assigned to A5H2 (see below). G7H1 cannot be close to A5H2 in space, but the imino hydrogen-bonded GA mismatch shown in Figure 1 places A5H2 2.3 Å from G4H1. Thus the resonance at 12.51 ppm is assigned to G4H1. By elimination, the imino resonance at 13.23 ppm is assigned to G7H1. As expected, the resonances for the terminal G1H1 and mismatched G4H1 are broader than those for the internal G3H1 and G7H1 because of base pair fraying and dynamics. Also, no NOE is observed to G1H1 due to its rapid exchange with water.

Figure 3 shows a NOESY-JRE spectrum that helps to assign the amino proton resonances and confirm the base-pair patterns of the secondary structure. The G4H1−G3H1, G3H1−G7H1, and G4H1−A5H2 cross peaks agree with 1-D imino proton NOE spectra. The NOE connectivities G3H1−

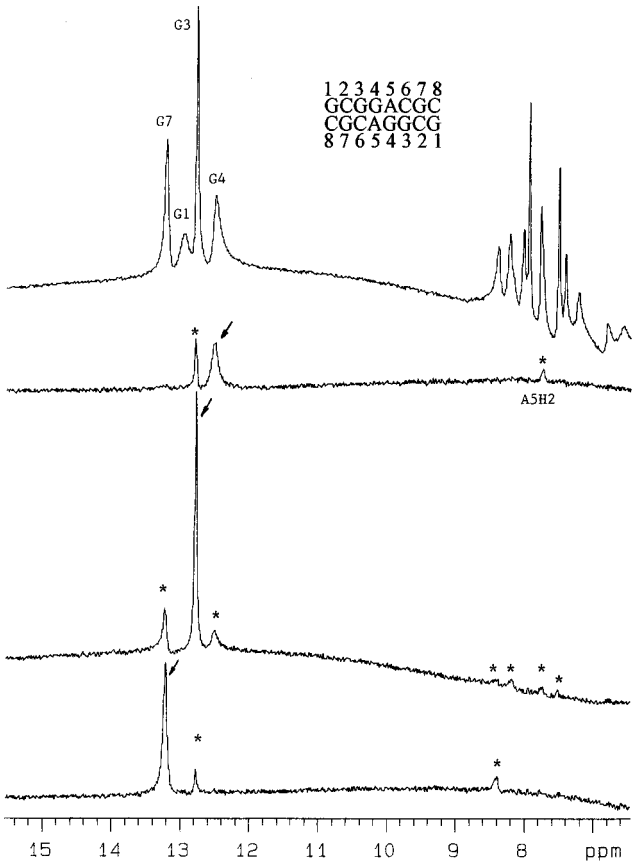

FIGURE 2: 500 MHz proton NMR and 1-D NOE difference spectra of 3.0 mM $(rGCGGACGC)_2$ at 25 °C in 0.1 M NaCl, 10 mM sodium phosphate, and 0.5 mM EDTA in 90% H_2O and 10% D_2O, pH = 7. From top to bottom: off-resonance spectrum; difference spectra between off-resonance and on-resonance spectra acquired with 1 s saturation at 12.51, 12.79, and 13.23 ppm, respectively. The saturated resonances are indicated by arrows while the observed NOEs are designated by asterisks. The secondary structure of $(rGCGGACGC)_2$ is shown at the top of the figure.

$C6NH_2$−C6H5 and G7H1−$C2NH_2$−C2H5, the characteristic pattern of Watson−Crick GC base pairs, proves the assignments of G7H1 and G3H1 and the secondary structure: $^{GCG^{GA}CGC}_{CGC_{AG}GCG}$. C2H5, C6H5, and C8H5 resonances were assigned using standard procedures from 2-D NMR in 100% D_2O. Due to exchange with water, the resonance for G1H1 does not appear in the diagonal of the NOESY-JRE spectrum. The exchange cross peaks at the lower right corner of the NOESY-JRE spectrum directly represent the proton exchange rate between these imino protons and water protons (4.91 ppm) (Dobson et al., 1986; Sklenar et al., 1987). The relative exchange rates with water protons are G1H1 > G4H1 > G7H1 ∼ G3H1. This is consistent with the line width distribution of imino proton resonances and suggests dynamics at the end of the helix and in the mismatch region.

Assignment of Nonexchangeable Protons and Phosphorus. Chemical shift assignments of protons and phosphorus at 35 °C are shown in Table 1. All the protons were assigned using standard procedures described by Varani and Tinoco (1991).

Figure 4 shows the base H8/H6/H2 to H1′/H5 region of a 400 ms NOESY spectrum. Three cytidine H5−H6 cross peaks were identified because of their existence in the same regions of both DQF-COSY and TOCSY spectra (data not shown). All the base H8/H6 and sugar H1′ protons are thus assigned from the straightforward sequential NOE connectivity pathway H8/H6(n)−H1′(n)−H8/H6(n+1) (Petersheim & Turner, 1983; Hare et al., 1983; Wuthrich, 1986). The resonance left is at 7.76 ppm, which is assigned to A5H2. The assignment of A5H2 was confirmed by its long T_1 relaxation time of 5.3 s (see Supporting Information for a table of T_1s) and by its NOE to A5H1′ (interstrand) and C6H1′ (Wuthrich, 1986; Varani & Tinoco, 1991). In 400 ms NOESY spectra, weak sequential base−base NOEs (G1H8−C2H6−G3H8, G4H8−A5H8, C6H6−G7H8−C8H6) confirm the assignments of base protons. The extremely broad C6H1′ resonance caused some difficulties in finding the pathway and indicates dynamic behavior at the junction between the tandem mismatch and stem base pairs. The cross peak between G4H1′ and G4H8 is weak and broad but can clearly be seen if the contour level is lowered. The pattern and intensity distribution of sequential NOE cross peaks reveal that $(rGCGGACGC)_2$ forms basically an A-form duplex and all the bases are in anti conformation relative to their sugars (Wuthrich, 1986).

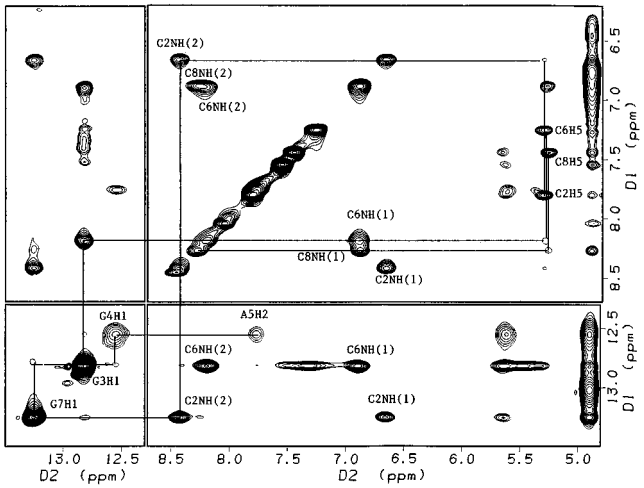

FIGURE 3: Tile contour plot of the NOESY-JRE spectrum of $(rGCGGACGC)_2$ in 90% H_2O and 10% D_2O at 15 °C. The imino proton region and amino/base proton region are both maximized in the excitation profile and brought together by cutting the region between them. The assignments are for the D_2 dimension. Intrabase-pair GH1−CNH_2−CH5 connectivities for G7−C2 and G3−C6 pairs and sequential G7H1−G3H1−G4H1 connectivities are shown in solid lines.

H2′ proton resonances are assigned on the basis of short mixing time (25, 50, and 100 ms) NOESY spectra where spin diffusion does not significantly contribute to NOE intensities (Neuhaus & Williamson, 1989). In the H1′/H5 to sugar region of the 100 ms NOESY spectrum (Figure 5), intranucleotide H1′−H2′ cross peaks are the strongest for each H1′ resonance, since H1′ is closer in space to H2′ of the same sugar ring than to any other sugar protons for both C3′-endo and C2′-endo conformations. Eight H2′ resonances are thus assigned and confirmed by sequential H8/H6(n)−H2′(n)−H8/H6(n+1) NOE connectivities delineated in the sugar−H8/H6/H2 region of the 400 ms NOESY spectrum in Figure 6. The sugar−H1′/H5 region of a DQF-COSY spectrum (data not shown) helps to determine the sugar pucker conformation through intranucleotide H1′−H2′ scalar coupling constants (Table 2). Scalar coupling between H1′ and H2′ of a sugar ring is 0−2 Hz for C3′-endo and 10 Hz for C2′-endo (Wuthrich, 1986; Varani & Tinoco, 1991). In DQF-COSY, nonexistence of H1′−H2′ scalar coupling cross peaks for the internal nucleotides suggests C3′-endo sugar

Table 1: Proton and Phosphorus Chemical Shift Assignments of (rGCGGACGC)$_2$ at 35 °C (ppm)[a]

base	H8/H6	H2/H5	H1′	H2′	H3′	H4′	H5′/H5″ [b]	imino[c]	amino[c]	phosphorus
G1	7.99	na[d]	5.68	4.76	4.55	4.32	4.02/3.89	12.96		na
C2	7.79	5.31	5.62	4.50	4.55	4.48	4.54/4.20	na	8.41/6.65	−2.46
G3	7.55	na	5.62	3.74	4.75	4.37	4.57/4.14	12.81		−2.16
G4	8.02	na	5.89	4.93	4.85	4.52	4.50/4.18	12.52		−3.08
A5	7.93	7.76	5.97	4.74	4.29	4.62	4.43/4.37	na		−1.91
C6	7.29	5.34	5.35	4.23	4.42	4.40	4.48/4.07	na	8.19/6.89	−2.55
G7	7.53	na	5.66	4.36	4.60	4.40	4.48/4.07	13.23		−2.52
C8	7.46	5.24	5.69	3.97	4.11	4.17	4.47/4.02	na	8.25/6.88	−2.58

[a] Chemical shifts are relative to TSP for proton and to phosphate buffer for phosphorus resonances. [b] The H5′ and H5″ assignments are not stereospecific. [c] Exchangeable proton chemical shifts are given for 15 °C. [d] Not applicable.

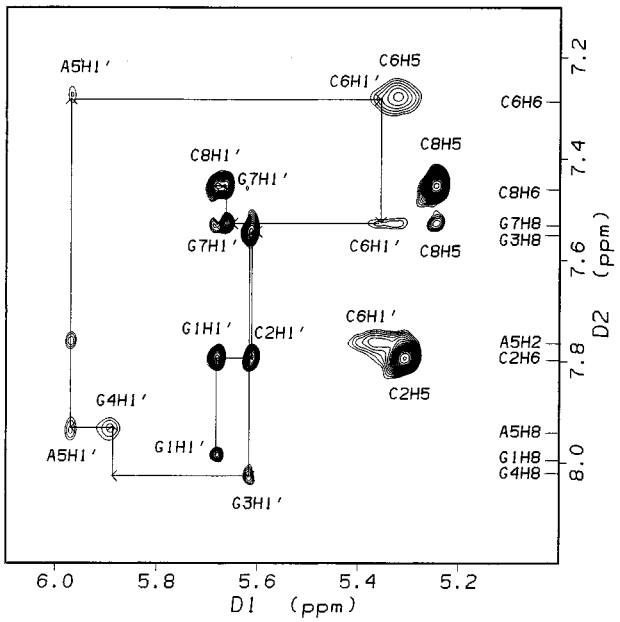

FIGURE 4: H8/H6/H2−H1′/H5 region of 400 ms NOESY of (rGCGGACGC)$_2$ at 35 °C in 100% D_2O. Assignments for the D_2 dimension are on the right axis. The rest of the assignments are for the D_1 dimension. Sequential −H8/H6(n)−H1′(n)−H8/H6(n+1)− connectivities are shown in solid lines.

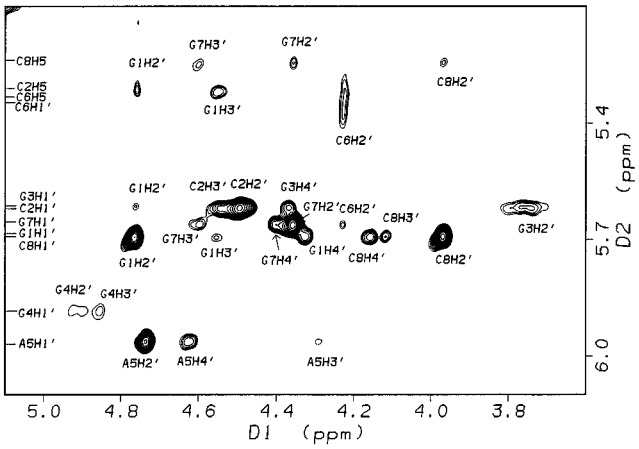

FIGURE 5: Sugar−H1′/H5 region of the 100 ms NOESY of (rGCGGACGC)$_2$ at 35 °C in 100% D_2O. Assignments for the D_2 dimension are on the left axis. The rest of the assignments are for the D_1 dimension.

pucker for these nucleotides. Weak H1′−H2′ crosspeaks for G1 and C8 indicate that the sugar pucker of terminal nucleotides is a mixture of C3′-endo and C2′-endo, with C3′-endo predominating.

H3′, H4′, H5′, and H5″ resonances were assigned from DQF-COSY (Figure 7) and ^{1}H−^{31}P HETCOR spectra

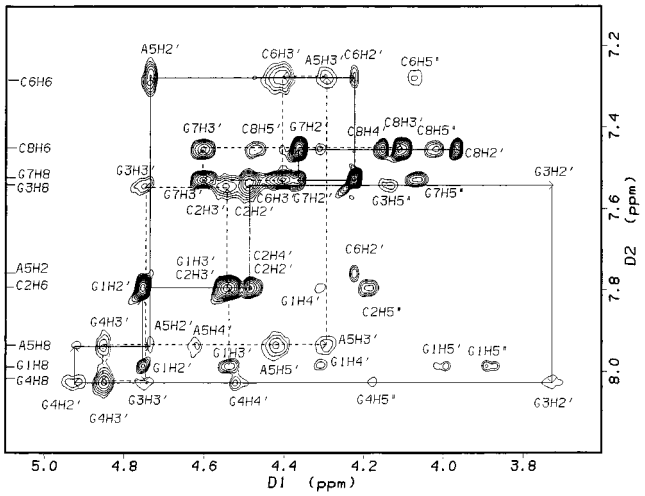

FIGURE 6: Sugar−H8/H6/H2 region of the 400 ms NOESY of (rGCGGACGC)$_2$ at 35 °C in 100% D_2O. Assignments for the D_2 dimension are on the left axis. The rest of the assignments are for the D_1 dimension. The sequential −H8/H6(n)−H2′(n+1)−H8/H6(n+1)− and −H8/H6(n)−H3′(n+1)−H8/H6(n+1)− connectivities are shown in solid lines and dashed lines, respectively.

(Figure 8). The scalar connectivity H2′−H3′−H4′−H5′/H5″ in the DQF-COSY spectrum helps to identify sugar spin systems, though the connectivity is sometimes interrupted by missing peaks due to broad line width, overlap, and small coupling constants. H2′−H3′ scalar cross peaks for G3 and G4 are missing because of the broad line width of G3H2′ and G4H2′. The H2′−H3′ scalar cross peak for C2 cannot be seen due to overlap. Resonances for C2H3′, G3H3′, and G4H3′ were identified in the ^{1}H−^{31}P HETCOR from their strong scalar coupling to the phosphorus of the adjacent nucleotide on the 3′ side. Assignment of H3′ resonances was confirmed by sequential NOE connectivities H8/H6(n)−H3′(n)−H8/H6(n+1) (Figure 6). H4′ resonances were therefore assigned from H3′−H4′ scalar cross peaks in the DQF-COSY spectrum (Figure 7). Large H3′−H4′ scalar couplings once again suggest predominant C3′-endo conformation for all the sugar puckers. In the DQF-COSY spectrum, eight H5′−H5″ scalar cross peaks were found to have large coupling constants. It is hard to assign each H5′/H5″ pair to a particular sugar, however, since most of the H4′−H5′/H5″ scalar cross peaks are too weak to be seen due to the small amplitude of the coupling constants (less than 3 Hz). The ^{1}H−^{31}P HETCOR experiment solves this problem through the intranucleotide scalar connectivities P−H5′/H5″. Assignments of H5′/H5″ resonances do not have stereospecificity, and H5′ is assumed to resonate downfield from H5″ (Varani & Tinoco, 1991). Intranucleotide H8/H6−H5′/H5″ NOEs confirm the assignments of H5′/

Table 2: *J* Coupling Constants (Hz) for r(GCGGACGC)$_2$[a]

base	H1′−H2′	H2′−H3′	H3′−H4′	H4′−H5′	H4′−H5″	H5−H6	H5′−H5″
G1	4.1	5.0	8.5	7.6	3.7	na	10.2
C2	<2	overlap	9.5	overlap	overlap	6.9	9.6
G3	broad	broad	10.3 ± 2.0	<3	<3	na	9.8
G4	<2	broad	7.6	<3	<3	na	9.9
A5	<2	5.3	10.0	overlap	overlap	na	8.9±2.0
C6	broad	4.8	<3	<3	<3	7.1	9.8
G7	<2	5.7	9.0	<3	<3	na	10.2
C8	3.9	5.1	8.4	<3	<3	6.8	9.8

5′-base	H3′−P	P−H5′	P−H5″	P−H4′	3′-base
G1	9.6	5.4	5.7	5.0	C2
C2	9.2	overlap	5.4	5.6	G3
G3	11.0	4.9	6.0	overlap	G4
G4	9.3	6.2	<3	<3	A5
A5	9.5	overlap	overlap	overlap	C6
C6	8.7	overlap	5.0	overlap	G7
G7	9.1	overlap	5.6	4.7	C8

[a] Proton−proton and proton−phosphorus *J* coupling constants have error bars of ±1 and ±2 Hz, respectively, unless otherwise stated.

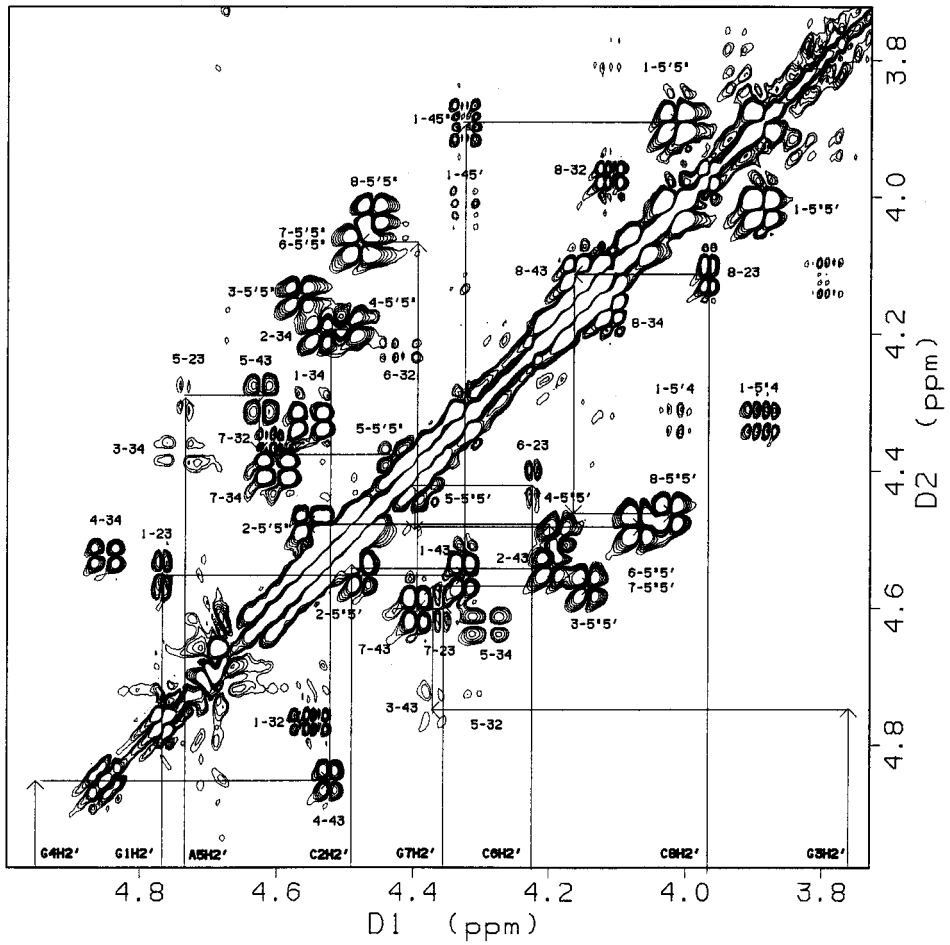

FIGURE 7: Sugar−sugar region of the DQF-COSY of (rGCGGACGC)$_2$ at 35 °C in 100% D_2O. The intrasugar scalar H2′−H3′−H4′−H5′/H5″ connectivities are shown in solid lines for all eight bases. On the connectivity pathways, some of the cross peaks are missing due to broad line width, overlap and small coupling constants.

H5″ resonances (Figure 6). No H5′/H5″ protons were used in distance restraints for simulated annealing due to the nonstereospecificity of H5′/H5″ assignments. The completeness of the assignments of all the cross peaks in 400 ms NOESY, DQF-COSY, and ^{1}H−^{31}P HETCOR spectra further confirms the accuracy of the assignments in Table 1.

^{1}H−^{31}P HETCOR reveals that all seven phosphorus resonances are within 1.2 ppm, indicating that there is no major backbone distortion. The upfield-shifted G4P and downfield-shifted A5P and G3P, however, suggest some distortion at the mismatch region compared to Watson−Crick base-paired regions.

Structural Modeling. A total of 105 distance and 23 dihedral angle restraints per strand (Tables 3 and 4) were used in structural modeling. The mismatched G4 and A5 have more distance restraints than any other nucleotide, so the structure of the GA tandem is well defined. G3 and C6 have the least number of restraints, partially because the resonances for G3H2′ and C6H1′ are broad, presumably due to dynamic behavior at the mismatch−stem junctures.

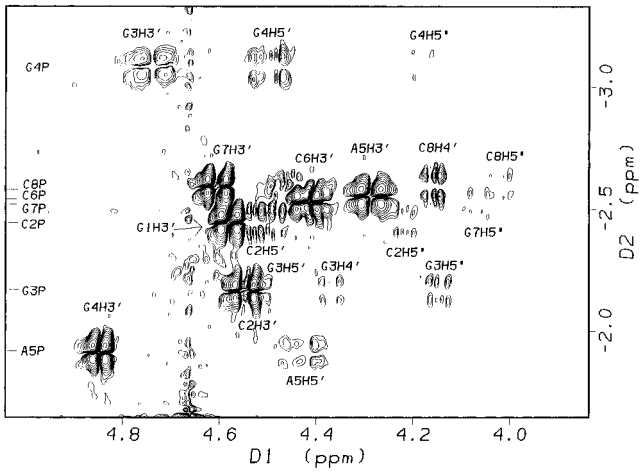

FIGURE 8: ^{1}H−^{31}P HETCOR of (rGCGGACGC)$_2$ at 35 °C in 100% D_2O. Proton chemical shift (D_1 dimension) assignments are shown at the peaks, while phosphorus chemical shift (D_2 dimension) assignments are shown on the left axis.

For NOEs that are close to the dyad axis, it can be problematic to distingish interstrand contacts from intrastrand contacts. The only NOE that could cause confusion is the NOE between A5H2 and A5H1′. On the basis of the two-spin approxmiation, the NOE corresponds to 4.1 Å in distance. The intrastrand A5H2−A5H1′ distance is at least 4.8 Å regardless of glycosyl angle. Thus, the observed NOE is attributed to the interstrand A5H2−A5H1′ distance. In the final structure, inter- and intrastrand A5H2−A5H1′ distances are 4.5 and 4.9 Å, respectively.

Over the course of four IRMA iterations, convergence was indicated by the leveling off of the *R*-factors (Figure 9). IRMA was used to refine 85 of 105 distance restraints per strand. The rest of the distance restraints are either hydrogen-bonding restraints or qualitative NOE distances derived from very weak cross peaks. IRMA not only refines the distances but also narrows the restraint bounds, so that the average bound width is only 0.6 Å or ±9%. In the final average structure, most of the restrained distances satisfy the distance restraints within 0.5 Å, with six distances violating the restraints by 0.5−1.0 Å. As shown in Figure 10, there is good agreement between final structure and experimental distance restraints. The great majority of distances satisfy the restraints within ±20% of the center of distance restraints. This is less than or equal to the bound width used by most other methods. The great majority of restrained dihedral angles of the final structure (Table 5) satisfy the dihedral angle restraints within 3°. On the basis of the final structure, a 400 ms NOESY spectrum was back-calculated. The H8/H6/H2−H5/sugar regions of experimental and back-calculated 400 ms NOESY spectra are shown in Figure 11. The match is quite good. The cross peak at 5.31−7.99 ppm

Table 3: Final Distance Restraints (Å) for (rGCGGACGC)$_2$ at 35 °C after IRMA (for Each Strand)[a]

intra	H1′−H8/H6	H2′−H8/H6	H3′−H8/H6	H4′−H8/H6	H1′−H5	H2′−H5	H3′−H5
G1	4.0 ± 0.5	3.7 ± 0.5	3.8 ± 0.8	4.2 ± 0.8			
C2	3.6 ± 0.1				4.5 ± 0.2		
G3	3.1 ± 0.1	4.2 ± 0.2	3.6 ± 0.1				
G4	4.1 ± 0.1	3.4 ± 0.2	2.9 ± 0.1	4.6 ± 0.3			
A5	3.8 ± 0.1	4.9 ± 0.4	3.3 ± 0.1	4.1 ± 0.1			
C6		3.8 ± 0.3	2.7 ± 0.1				
G7	4.0 ± 0.3	3.1 ± 0.3	2.7 ± 0.1				
C8	3.2 ± 0.1	3.1 ± 0.1	2.5 ± 0.1		3.1 ± 0.8		4.5 ± 0.1
inter	H1′−H8/H6	H2′−H8/H6	H3′−H8/H6	H4′−H8/H6	H1′−H5	H2′−H5	H3′−H5
G1−C2	3.3 ± 0.7	2.1 ± 0.2		5.5 ± 0.5	4.2 ± 0.5	3.0 ± 0.6	2.9 ± 0.4
C2−G3		2.6 ± 0.1	3.0 ± 0.1				
G3−G4	4.6 ± 0.3	3.1 ± 0.1	3.8 ± 0.1				
G4−A5	3.4 ± 0.1	3.8 ± 0.3	3.4 ± 0.1	4.4 ± 0.9			
A5−C6	4.6 ± 0.1	2.8 ± 0.1	3.1 ± 0.1	4.8 ± 0.6			3.6 ± 0.1
C6−G7	3.7 ± 0.7	2.2 ± 0.1					
G7−C8	3.9 ± 0.1	2.2 ± 0.1	3.1 ± 0.1		4.0 ± 0.3	3.7 ± 0.6	3.5 ± 0.1
intra	H2′−H1′	H3′−H1′	H4′−H1′	H2′−H3′	H3′−H4′		
G1	3.3 ± 0.4	3.6 ± 0.2	3.2 ± 0.2	2.0 ± 0.2			
C2	2.6 ± 0.1	2.9 ± 0.1					
G3	2.9 ± 0.1		3.9 ± 0.2				
G4	3.0 ± 0.3	3.3 ± 0.3	3.9 ± 0.8		2.4 ± 0.2		
A5	2.9 ± 0.1	3.6 ± 0.1	3.7 ± 0.1				
C6	2.6 ± 0.2						
G7	2.3 ± 0.1	3.2 ± 0.2	3.2 ± 0.2				
C8	2.5 ± 0.1	3.4 ± 0.1	3.2 ± 0.1				
inter	H2′−H1′	H3′−H1′	H4′−H1′	H8/H6−H8/H6	H8/H6−H5	H2−H1′	H2−H2′
G1−C2	3.1 ± 0.4			4.6 ± 0.8[b]			
C2−G3				3.8 ± 0.8[b]			
G3−G4	> 4.5[b]						
G4−A5	4.0 ± 0.1	4.2 ± 0.5	4.7 ± 0.8[b]	4.1 ± 0.8[b]			
A5−C6	4.0 ± 1.0[b]					3.4 ± 0.7	4.7 ± 0.8
C6−G7	3.5 ± 0.3			3.9 ± 0.8[b]			
G7−C8				3.7 ± 0.7[b]	3.8 ± 0.2		

[a] (1) One interstrand restraint A5H2−A5′H1′ = 4.1 ± 0.1 Å. (2) Imino restraints: G1H1−G7′H1= 4.0 ± 1.0 Å; G7′H1−G3H1 = 4.0 ± 1.0 Å; G3H1−G4H1= 4.0 ± 1.0 Å. (3) Nonproton restraints: nine hydrogen-bonding restraints (generic restraints), 1.8 ± 0.2 Å. Total 105 distance restraints = 49 (intraresidue) + 43 (interresidue) + 1 (interstrand) + 3 (imino) + 9 (H-bond) = 85 (IRMA) + 20 (generic) = 85 (NOE) + 11 (qualitative NOE) + 9 (H-bond). [b] Qualitative generic restraints that are not refined by IRMA.

Table 4: Torsion Angle Restraints for (rGCGGACGC)$_2$ (deg)

residue	α	β	γ	δ	ε	ζ	χ
G1	na[a]	na	nr[b]	nr	−138 ± 20	nr	nr
C2	nr	178 ± 28	nr	83 ± 15	−141 ± 20	nr	nr
G3	nr	178 ± 26	58 ± 30	83 ± 25	−120 ± 34	nr	nr
G4	nr	178 ± 29	58 ± 30	94 ± 15	−141 ± 20	nr	nr
A5	nr	178 ± 30	nr	83 ± 15	−139 ± 20	nr	nr
C6	nr	nr	58 ± 30	nr	−145 ± 20	nr	nr
G7	nr	178 ± 34	58 ± 30	86 ± 15	−142 ± 20	nr	nr
C8	nr	178 ± 37	58 ± 30	nr	na	na	nr
A-form	−68	178	54	82	−153	−71	−158

[a] Not applicable. [b] Not restrained.

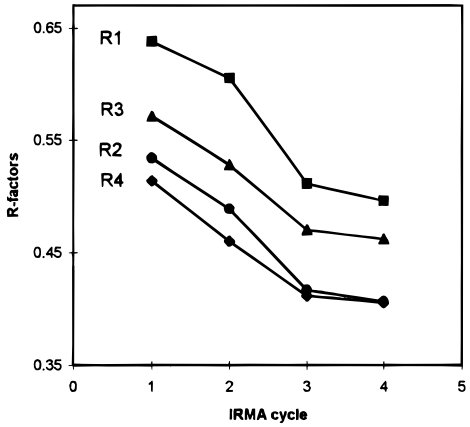

FIGURE 9: *R*-factors decreased and leveled off over the course of IRMA cycles. See Materials and Methods for definitions of R_1, R_2, R_3, and R_4. IRMA was terminated after four iterations.

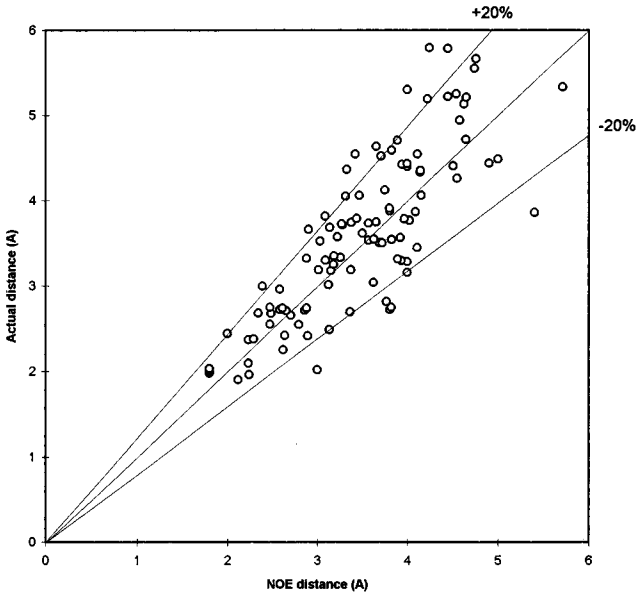

FIGURE 10: Correlation between the center of distance restraints (from Table 2) and the actual distances from the final averaged structure of (rGCGGACGC)$_2$ for all the distances restrained.

(C2H5-G1H8) can not be seen in the experimental 400 ms NOESY of Figure 11. It is seen, however, in a lower threshold contour plot. At lower temperatures (30 and 25 °C), this peak is much bigger, comparable to the one in the back-calculated 400 ms NOESY, probably because of slower chemical exchange at lower temperatures. The other peak at 5.89−8.02 ppm (G4H1′−G4H8) is also observed in experimental 400 ms NOESY coutour plots with lower threshold. Since spin diffusion is very strong at a mixing time of 400 ms, there is a wide variety of peak intensities.

Table 5: Average Dihedral Angles for the 14 Final Structures of (rGCGGACGC)$_2$ (deg)

residue	α	β	γ	δ	ε	ζ	χ	P[b]
G1	na[a]	na	176.8	79.6	−160.4	−63.2	−159.9	10.0
C2	−72.6	177.4	56.7	75.6	−161.2	−63.7	−165.0	13.6
G3	−73.8	179.7	61.0	79.9	−167.7	−72.8	−163.5	17.7
G4	−74.5	179.4	62.5	78.5	−171.8	−56.9	−144.1	11.1
A5	141.5	190.6	−176.4	85.5	−149.8	−60.0	−180.4	0.3
C6	−72.0	184.2	55.25	78.1	−164.5	−63.2	−163.0	18.6
G7	−75.2	177.8	56.8	79.0	−164.0	−61.4	−154.2	17.6
C8	−76.8	180.5	54.9	77.8	na	na	−158.8	22.4
A-form	−68	178	54	82	−153	−71	−158	18

[a] Not applicable. [b] Pseudorotation angles.

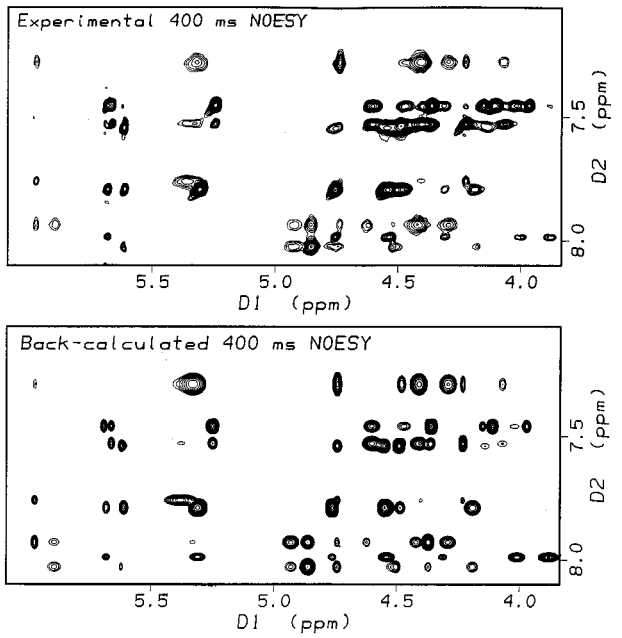

FIGURE 11: H8/H6/H2−sugar/H5 regions of the experimental spectrum (upper) and back-calculated 400 ms NOESY spectrum (lower) based on coordinates averaged over 14 final structures of (rGCGGACGC)$_2$ at 35 °C.

Both short and long proton−proton distances of the final structure reproduce the experimental NOEs. Thus the final structure is consistent with the large body of experimental data.

Stereoviews of 14 structures generated are shown from two different angles in Figure 12. Averaged pairwise rmsds are 0.81, 0.64, 0.40, 0.45, 0.48, 0.28, 0.30, and 0.40 Å for G1, C2, G3, G4, A5, C6, G7, and C8 residues, respectively. The good overlap shown in Figure 12 does not necessarily imply the accuracy of the final structures. It reflects high precision, probably due to IRMA decreasing the average bound width of distance restraints from ±20% to ±9%. These tightened bounds are compensated by a decrease of force constants from 20 to 10 kcal/mol Å^2 to avoid freezing the molecule during the simulated annealing. Starting from 3 different starting structures, 14 structures were minimized using the initial distance restraints (comparable to the strong/medium/weak approach), giving an averaged pairwise rmsd of 0.73 Å. The 14 final structures (Figure 12) minimized using final IRMA-refined distance restraints have an averaged pairwise rmsd of 0.52 Å. These two sets of structures overlap reasonably well with an average pairwise rmsd of 1.00 Å between them, indicating that IRMA did not change the global features of the structures. The drop of *R*-factors,

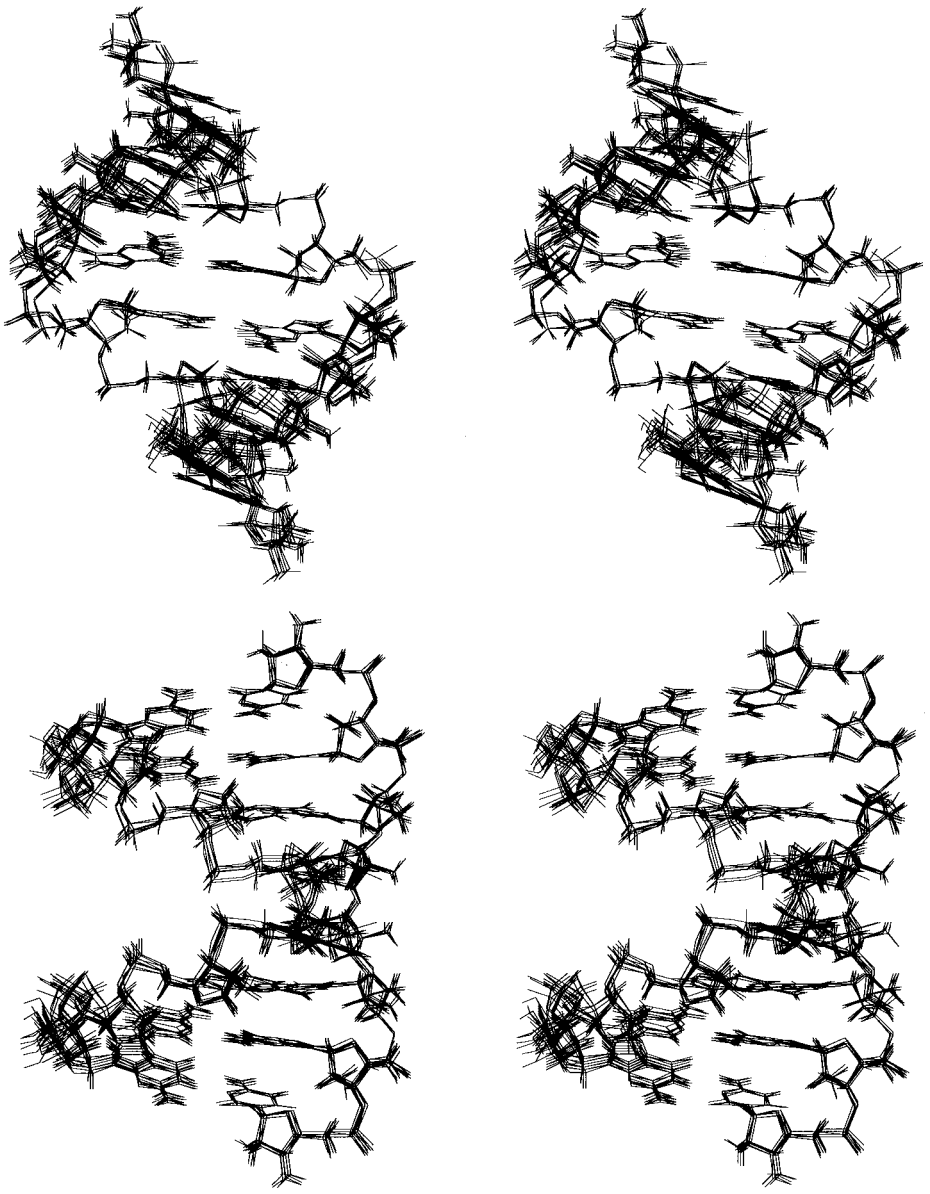

FIGURE 12: Stereoviews of superposition of 14 refined structures of (rGCGGACGC)$_2$ at two different angles.

which is independent of rmsd of structures, shows that IRMA refined the structures quantitatively so that the final refined structures satisfy the experimental NOESY spectra better than the structures generated with the initial constraints. The average energy of the final structure is 545 ± 4 kcal/mol.

Tables 6 and 7 list helical parameters for the final structure according to the Cambridge convention (Dickerson, 1989) calculated with the NEWHEL92 program, distributed by the Protein Data Bank. The helix parameters indicate that (rGCGGACGC)$_2$ is basically A-form with minor deviations at the mismatch region. All the sugar puckers are in C3′-endo conformation and all the bases are in anticonformation relative to their sugars. In the center of the helix, each of the GA mismatches is in the imino hydrogen-bonded conformation (Figure 1) with two potential hydrogen bonds: G-O6 to A-NH6 (2.15 Å) and G-NH1 to A-N1 (2.09 Å) (see Figure 13). There is substantial intrastrand stacking between GA mismatches as illustrated in Figure 13.

The helix has 11.2 base pairs per turn on average, compared to 10.9 for A-form RNA. Overall average helical twist is 32.2°, very close to 32.7° for A-form. The helical twists are 33.2° and 34.3° for G1C8−C2G7, C2G7−G3C6 steps, but the helix is slightly overwound at the stem−mismatch junction G3C6−G4A5 with 36.0° twist and is underwound between the two GA mismatches G4A5−A5G4 with 21.0° twist. The helical rise between two GA mis-

Table 6: Helical Parameters[a] for Base Steps[b] in the Final Structure of (rGCGGACGC)$_2$

base steps	helical twist (deg)	rise (Å)	slide (deg)	roll (deg)	tilt (deg)	interstrand P−P distance[c] (Å)
1	33.2	3.0	−1.2	7.5	−3.7	17.8
2	33.9	3.0	−1.5	5.8	1.4	17.6
3	36.0	2.8	−1.3	−11.8	1.2	17.8
4	21.0	2.5	−1.6	14.7	0.3	20.4
5	36.1	2.8	−1.3	−12.5	−0.4	17.8
6	34.7	3.0	−1.5	4.0	−1.7	17.6
7	33.1	3.0	−1.2	6.6	2.6	17.8
average	32.3	2.9	−1.4	2.0	0.0	18.1
A-form	32.7	2.8	−1.5	−0.4	13.0	17.5

[a] The parameters were calculated using NEWHEL92 from Brookhaven Protein Data Bank. [b] Base steps are defined as $\frac{G}{C}\frac{1}{G}\frac{C}{C}\frac{2}{G}\frac{G}{C}\frac{3}{A}\frac{G}{G}\frac{4}{A}\frac{A}{G}\frac{5}{G}\frac{C}{C}\frac{6}{G}\frac{G}{C}\frac{7}{G}\frac{C}{G}$. [c] Interstrand phosphorus−phosphorus distance is defined as the distance between nP and $(10 - n)'$P; n = 2, 3, 4, ..., 8, e.g., C2P−C8′P.

Table 7: Helical Parameters[a] for Base Pairs in the Final Structure of (rGCGGACGC)$_2$

base pairs	propeller twist (deg)	inclination (deg)	buckle (deg)	X-DSP (Å)	interstrand C1′−C1′ distance[b] (Å)	λ_1[c] (deg)	λ_2[c] (deg)
G1−C8′	−12.5	11.0	−4.9	−3.2	10.8	49.7	55.4
C2−G7′	−15.9	7.4	4.7	−3.4	10.9	50.0	55.3
G3−C6′	−13.4	8.8	−0.7	−3.1	10.7	49.5	55.2
G4−A5′	−21.0	10.0	−7.2	−3.1	12.8	39.4	51.8
A5−G4′	−21.2	10.2	7.5	−3.1	12.8	52.0	39.3
C6−G3′	−14.0	9.9	1.0	−3.0	10.7	55.1	49.3
G7−C2′	−14.3	8.3	−4.7	−3.4	10.9	55.5	50.0
C8−G1′	−10.6	10.8	3.7	−3.3	10.8	55.5	50.3
average	−15.4	9.5	−0.1	−3.2	11.3	50.8	50.8
A-form	−13.8	16.9	0.2	−4.4	10.9	52.2	52.2

[a] The parameters were calculated using NEWHEL92 from Brookhaven Protein Data Bank. [b] The interstrand C1′−C1′ distance is between two C1′ atoms of two paired bases. [c] λ_1 and λ_2 are defined as the angles between N9/N1−C1′ bonds and the C1′−C1′ vector for each base pair, with the subscripts 1 and 2 designating left and right base, respectively, in column 1.

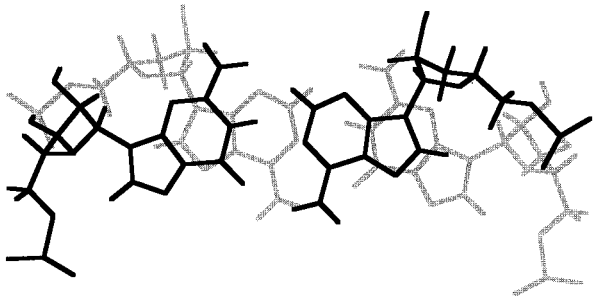

FIGURE 13: Intrastrand stacking between the two GA mismatches in the averaged structure for (rGCGGACGC)$_2$. The GA pair in dark lines is closer to the reader than the GA pair in shaded lines.

matches is 2.5 Å, clearly shorter than the A-form value of 2.8 Å. Also, α (141.5°) and γ (−176.4°) for A5 are completely different from the A-form values of −68° and 54°, respectively. Interstrand C1′−C1′ and P−P distances for both of the GA mismatches are about 12.8 and 20.4 Å, respectively, compared to 10.6 and 17.5 Å for A-form Watson−Crick base pairs. Sequential P−P distances are 5.8, 6.0, 6.1, 6.8, 5.7, and 5.8 Å from the 5′ to 3′ end of each strand. The A5P−C6P distance of 6.8 Å is significantly longer than the usual A-form distance of 5.9 Å. All of these indicate a local deviation from A-form.

Examination of the final structure reveals that all of the deviations are required to accommodate the conformation of the GA mismatch shown in Figure 13. In the GA mismatch region, the bulkiness of two opposing purines forces the phosphodiester backbone to stretch outward in order to accommodate the imino hydrogen-bonded conformation. The unusually long C1′−C1′ and P−P distances are the direct results of this stretching. The phosphodiester backbones cannot accommodate this stretching and maintain A-form torsion angles. The unusual α and γ angles for A5 push the A5 phosphorus atoms outward so that the interstrand P−P distance (20.4 Å) is long enough to accommodate the GA mismatch, and the outward stretch of these two phosphorus atoms at the G4−A5 step draws the two GA pairs closer to each other, resulting in a shorter rise (2.5 Å) between the GA mismatches. The unusual α and γ angles twist the backbone at the G4A5−A5G4 step against the direction of helical twist, resulting in lower twist angle (21.0°) and longer distance (6.8 Å) between C6P and A5P. The underwinding of the helix at the mismatch step favors the intrastrand stacking of the GA mismatches. The unusual α angle of A5 is presumably one of the reasons for the downfield shift of the A5 phosphorus resonance (Figure 8).

DISCUSSION

A wealth of RNA sequence information is rapidly becoming available. Optimal use of this data requires methods for modeling both secondary and three-dimensional structure from sequence. While this is relatively straightforward for regions of Watson−Crick base pairs, little is known about the interactions that determine secondary and three-dimensional structures in non-Watson−Crick regions. Thus it is important to develop model systems for studying these interactions.

GA mismatches are a common non-Watson−Crick motif in RNA. Studies in both RNA and DNA have shown that the thermodynamic stabilities and structures of tandem GA mismatches are dependent on the adjacent base pairs (Walter et al., 1994b; Morse & Draper, 1995; Li et al., 1991a,b; Cheng et al., 1992; Li & Agrawal., 1995). For example, measurements of the thermal stabilities of (CAGGACUG)$_2$, (GAGGACUC)$_2$, and (GCGGACGC)$_2$ (Walter et al., 1994b) indicate that the GGAC motif in these duplexes is in excess of 2 kcal/mol more stable than the CGAG motif in (GGCGAGCC)$_2$ (SantaLucia et al., 1990). Moreover, NMR chemical shifts for the imino protons indicate that the structure of the GA mismatches in the former duplexes differs from that of (GGCGAGCC)$_2$. These previous results suggest the base pairs directly flanking the tandem GA mismatch determine the properties of the motif. To provide further insight into the structural implications of base pair substitutions, the structure of (GCGGACGC)$_2$ was determined for comparison with the previously determined structure of (GGCGAGCC)$_2$ (SantaLucia & Turner, 1993).

The three-dimensional structures determined by NMR for (GCGGACGC)$_2$ and (GGCGAGCC)$_2$ (SantaLucia & Turner, 1993) are shown in Figure 14. The difference in overall shape is striking. Clearly, simple substitutions of Watson−Crick base pairs can induce dramatic changes in structure. The helix at the GA mismatches is much wider for the GGAC than for the CGAG motif. For example, the distances separating the phosphates between G and A on opposite strands are 20.4 and 12.5 Å, respectively. This is due to the difference in space required to accommodate the imino hydrogen-bonded and sheared GA mismatches in (GCGGACGC)$_2$ and (GGCGAGCC)$_2$, respectively (Figure 1). The GGAC motif also has a much wider and deeper major groove than CGAG. The GGAC helix is underwound (21.0°) at the mismatch region, which facilitates intrastrand stacking between mismatches (Figure 13), while the CGAG motif is overwound (80.8°), which facilitates interstrand stacking of the mismatches (SantaLucia & Turner, 1993).

It is likely that the differences observed in Figure 14 have biological consequences. The CGAG motif is present in known secondary structures of RNA (Gutell et al., 1992, 1993), but the GGAC motif has not been observed even though it is thermodynamically more favorable by 2 kcal/mol. It has been suggested that the CGAG and related motifs are important because they allow formation of tertiary

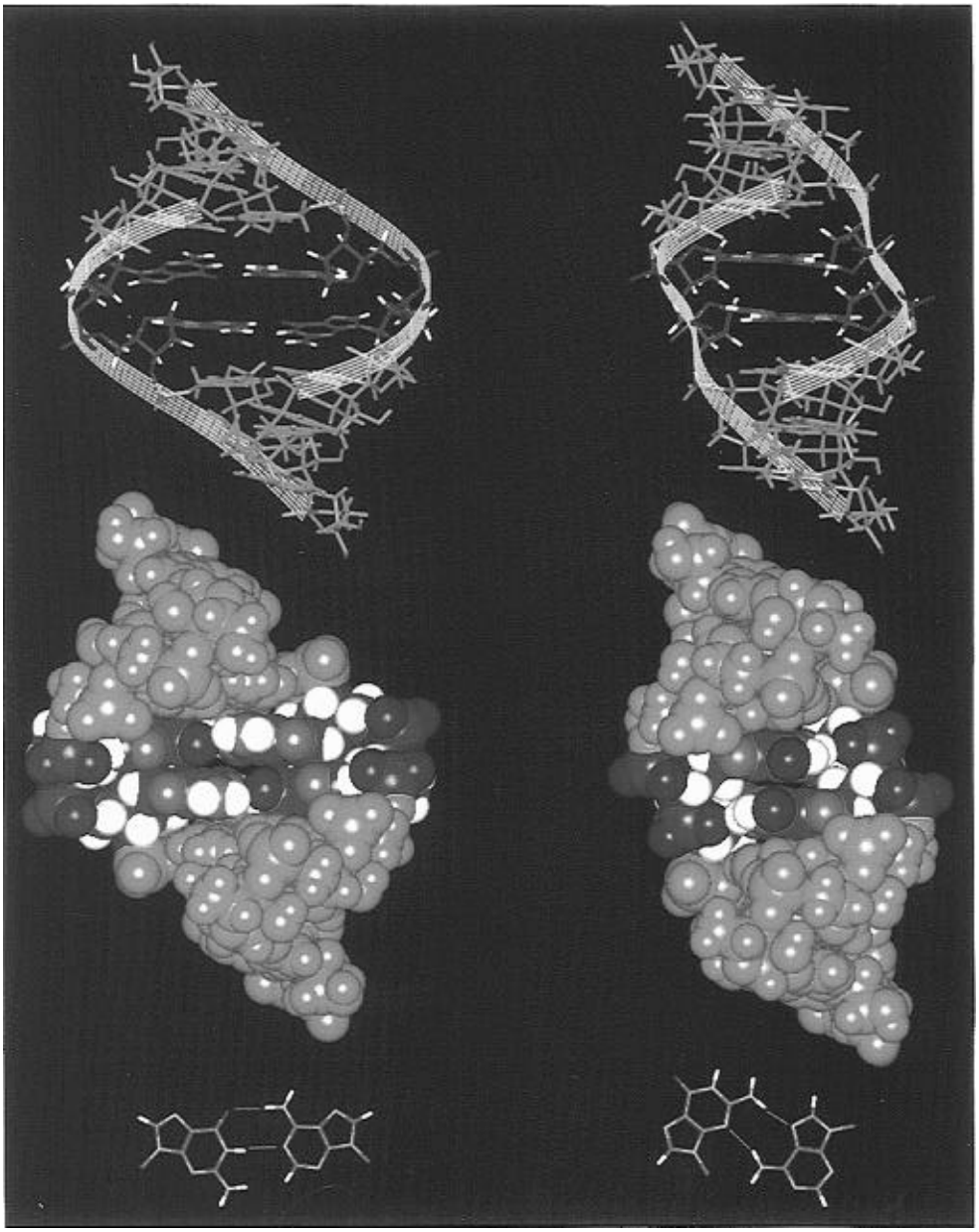

FIGURE 14: Solution NMR structures of $\frac{GCG^{GA}CGC}{CGC_{AG}GCG}$ (left half, this work) and $\frac{GGC^{GA}GCC}{CCG_{AG}CGG}$ (right half; SantaLucia & Turner, 1993) in stick (top two) and space-filling (center two) representations, viewed from the major grooves. Respective imino hydrogen bonded (bottom left) and sheared (bottom right) GA mismatches are shown in stick representations.

contacts with the base functional groups that are not involved in hydrogen bonding in the GA mismatch (SantaLucia & Turner, 1993; Walter et al., 1994b; Wimberly, 1994; Gautheret et al., 1994; Ebel et al., 1994). Equivalent tertiary interactions are not possible with the imino hydrogen-bonded GA mismatches in the GGAC motif. In addition, the large global difference observed in Figure 14 makes it unlikely that a GGAC motif would even fit into a space designed for a CGAG motif. This suggests that site-directed mutagenesis to substitute a GGAC for a CGAG motif could lead to large changes in global folding and therefore function of a large RNA. Such site-directed mutagenesis experiments may provide a useful way to detect the biological functions of tandem GA motifs. In particular, the prediction is that the GGAC sequence will not be able to form sheared GA pairs unless the tertiary interactions dependent on the sheared GA pairs contribute more than 2 kcal/mol of favorable free energy. For cases involving such strong tertiary interactions, the temperature dependence of folding and function should still be affected. Reversal of CG to GC pairs adjacent to GA or GG mismatches in junctions or hairpins has been found to affect function in the hammerhead ribozyme (Ruffner et al., 1990), hepatitis delta ribozyme (Been & Perrotta, 1995), and the ∝-sarcin/ricin domain of 28S rRNA (Glück et al., 1994).

The large global difference in the structures shown in Figure 14 suggests another potential biological function of tandem GA mismatches. In principle, binding of a protein could supply sufficient free energy to switch sheared GA mismataches to imino hydrogen-bonded mismatches, or vice versa. This in turn could propagate a conformational change.

Comparison of the structures of the GGAC and CGAG motifs provides a working hypothesis for the interactions that determine the hydrogen-bonding pattern for the GA mismatch. In Figure 15, the observed GC to GA interfaces for GGAC and CGAG motifs are presented along with structures obtained by inverting the Watson−Crick GC pairs for each case. The latter structures are easy to generate

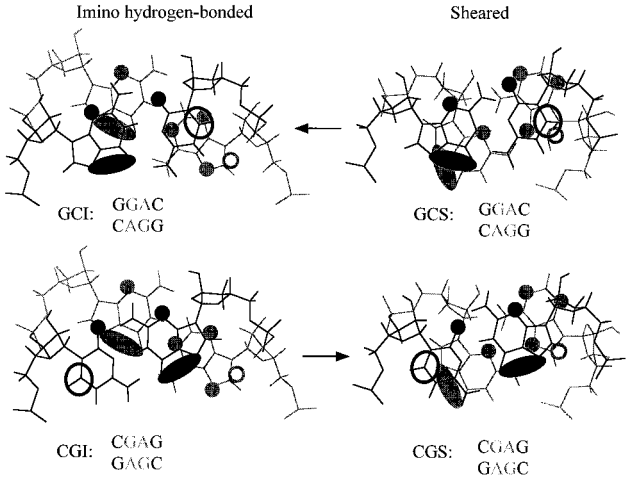

FIGURE 15: Schematic representation of electrostatic interactions between GA mismatch and adjacent base pair for $\frac{GGAC}{CAGG}$ and $\frac{CGAG}{GAGC}$ sequences in imino hydrogen bonded and sheared conformations. GCI and GCS denote stacking for the $\frac{GGAC}{CAGG}$ sequence with GA tandem in imino hydrogen-bonded and sheared conformations, respectively. CGI and CGS denote stacking for the $\frac{CGAG}{GAGC}$ sequence with GA tandem in imino hydrogen-bonded and sheared conformations, respectively. Arrows point toward structures determined by NMR. The other structures were generated by reversing the bases of the GC pairs in the NMR structures. The electrostatic potential distributions are those calculated for AT and GC pairs in DNA (Hunter, 1993) at a distance of 1.7 Å above and below each base-pair plane. Hollow and solid circles or ovals represent positive and negative potentials, respectively. The overlaps of potentials suggest a speculative working model for the interactions determining the structures of the GA mismatches.

because GC and CG pairs are isosteric. On the basis of calculations by Hunter (1993) for Watson−Crick base pairs in DNA, regions of high negative and positive electrostatic field are indicated, respectively, by filled in and open circles and ovals. Evidently, the structures that are not observed have fewer favorable and more unfavorable interactions than those that are observed. It has been suggested, however, that buried surface area rather than electrostatic interactions determine folding stabilities for nucleic acids (Friedman & Honig, 1995). The structure presented here should facilitate testing of these and other theories for the interactions that determine local structure and stability for nucleic acids.

In symmetric tandem GA mismatches, the structure presumably depends on the interaction of the GA mismatch with the base pair 5′ of G because the interactions 3′ of G are always with another GA. This is not a general rule for GA mismatches. The structures of several single GA mismatches in RNA have been determined by X-ray crystallography and NMR. The sequence motifs $\frac{5'GGA3'}{3'CAU5'}$ in which the AU is a reversed Hoogsteen pair (Wimberly et al., 1993; Szewczak et al., 1993), $\frac{5'GGC^{G}N}{3'CCG_{A}A}$ where N in the hairpin is C or A (Heus & Pardi, 1991), and $\frac{5'CGA3'}{3'GAA5'}$ (Biou et al., 1994) have sheared GA mismatches. $\frac{5'CGA3'}{3'GAU5'}$ has an imino hydrogen bonded GA (Leonard et al., 1994), and $\frac{5'GGG3'}{3'CAG5'}$ has either an imino hydrogen-bonded GA (Peterson et al., 1994) or no hydrogen bonding between G and A (Battiste et al., 1994), depending on the surrounding context. Thus the structure of single GA mismatches is not determined solely by the base pairs 5′ of the G.

ACKNOWLEDGMENT

We thank Jeffrey A. McDowell for help with the acquisition of $^{1}H-^{31}P$ HETCOR and structural modeling. We also thank Dr. Matthew A. Fountain for assistance with the acquisition of NOESY-JRE and data analysis with Felix and Dr. Charlie Bailey for discussions about IRMA. We thank Dr. Xiaoqi Jiao for help with the generation of helical parameters and Dr. Dan Gottschling for suggesting possible protein interactions.

SUPPORTING INFORMATION AVAILABLE

Two figures showing a (1) 1-D nonexchangeable proton 500 MHz NMR spectrum of $(rGCGGACGC)_2$ at 35 °C and (2) 1-D ^{31}P 202 MHz spectrum of $(rGCGGACGC)_2$ at 35 °C and one table of T_1s for base and H1′ protons of $(rGCGGACGC)_2$ (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Baeyens, K. J., DeBondt, H. L., & Holbrook, S. R. (1995) *Nat. Struct. Biol. 2*, 56−62.
Battiste, J. L., Tan, R., Frankel, A. D., & Williamson, J. R. (1994) *Biochemistry 33*, 2741−2747.
Been, M. D., & Perrotta, A. T. (1995) *RNA 1*, 1061−1070.
Biou, V., Yaremchuk, A., Tukalo, M., & Cusack, S. (1994) *Science 263*, 1404−1410.
Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct. 173*, 229−311.
Cheng, J. W., Chou, S. H., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 1037−1041.
Davies, D. B. (1978) *Prog. NMR Spectrosc. 12*, 135−225.
Dickerson, R. E. (1989) *Nucleic Acids Res. 17*, 1797−1803.
Dobson, C. M., Lian, L.-Y., Redfield, C., & Topping, K. D. (1986) *J. Magn. Reson. 69*, 201−209.
Ebel, S., Brown, T., & Lane, A. N. (1994) *Eur. J. Biochem. 220*, 703−715.
Friedman, R. A., & Honig, B. (1995) *Biophys. J. 69*, 1528−1535.
Gautheret, D., Konings, D., & Gutell, R. R. (1994) *J. Mol. Biol. 242*, 1−8.
Glück, A., Endo, Y., & Wool, I. G. (1994) *Nucleic Acids Res. 22*, 321−324.
Gorenstein, D. G., Meadows, R. P., Metz, J. T., Nikonowicz E., & Post, C. B. (1990) *Adv. Biophys. Chem. 1*, 47−124.
Greene, K. L., Jones, R. L., Li, Y., Robinson, H., Wang, A. H.-J., Zon, G., & Wilson, W. D. (1994) *Biochemistry 33*, 1053−1062.
Gutell, R. R., Schnare, M. N., & Gray, M. W. (1992) *Nucleic Acids Res. 20 (Suppl.)*, 2095−2109.
Gutell, R. R., Gray, M. W., & Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319−336.
Heus, H. A., & Pardi, A. (1991) *Science 253*, 191−194.
Hore, P. J. (1983) *J. Magn. Reson.* 55, 283−300.
Hosur, R. V., Govil, G., & Miles, H. J. (1988) *Magn. Reson. Chem. 26*, 927−944.
Hunter, C. A. (1993) *J. Mol. Biol. 230*, 1025−1054.
Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C., & Altona, C. J. (1984) *J. Biomol. Struct. Dyn. 1*, 1387−1405.
Leonard, G. A., McAuley-Hecht, K. E., Ebel, S., Lough, D. M., Brown, T., & Hunter, W. N. (1994) *Structure 2*, 483−494.
Li, Y., Zon, G., & Wilson, W. D. (1991a) *Biochemistry 30*, 7566−7572.
Li, Y., Zon, G., & Wilson, W. D. (1991b) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26−30.
Li, Y., & Agrawal, S. (1995) *Biochemistry 34*, 10056−10062.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
Morse, S. E., & Draper, D. E. (1995) *Nucleic Acids Res. 23*, 302−306.
Murphy, F. L., & Cech, T. R. (1994) *J. Mol. Biol. 236*, 49−63.

Neuhaus, D., & Williamson, M. P. (1989*) The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, Inc., New York.
Nikonowicz, E. P., Meadows, R. P., Fagan, P., & Gorenstein, D. G. (1991) *Biochemistry 30*, 1323−1334.
Ogilvie, K. K., Theriault, N., & Sadana, K. L. (1977) *J. Am. Chem. Soc. 99*, 7741−7743.
Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 264−268.
Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J., & Feigon, J. (1994) *Biochemistry 33*, 5358−5365.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994a) *Nature (London) 372*, 68−74.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994b) *Nature (London) 372*, 111−113.
Ruffner, D. E., Stormo, G. D., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 10695−10702.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813−8819.
SantaLucia, J., Jr., & Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
Scaringe, S. A., Franklyn, C., & Usman, N. (1990) *Nucleic Acids Res. 18*, 5433−5441.
Sklenar, V., & Feigon, J. (1990) *Nature (London) 345*, 836−838.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94−98.
Sklenar, V., Brooks, B. R., Zon, G., & Bax, A. (1987) *FEBS Lett. 216*, 249−252.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Szewczak, A. A., Moore, P. B., Chan, Y. L., & Wool, I. G. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 9581−9585.
Tinoco, I., Jr., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature (London) 230*, 363−367.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
Usman, N., Ogilvie, K. K., Jiang, M.-Y., & Cedergren, R. J. (1987) *J. Am. Chem. Soc. 109*, 7845−7854.
Varani, G., & Tinoco, I., Jr. (1991) *Q. Rev. Biophys. 24*, 479−532.
Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Muller, P., Mathews, D. H., & Zuker, M. (1994a) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
Walter, A. E., Wu, M., & Turner, D. H. (1994b) *Biochemistry 33*, 11349−11354.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
Wimberly, B. (1994) *Nat. Struct. Biol. 1*, 820−827.
Wimberly, B., Varani, G., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 1078−1087.
Wu, M., McDowell, J. A., & Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley Interscience, New York.
Zwieb, C. (1992) *J. Biol. Chem. 267*, 15650−15656.

BI960133Q