Solution Structure of (rGGCAGGCC)$_2$ by Two-Dimensional NMR and the Iterative Relaxation Matrix Approach[†,‡]

Ming Wu,[§] John SantaLucia, Jr.,[||] and Douglas H. Turner*[,§]

Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, and Department of Chemistry, Wayne State University, Detroit, Michigan 48202

Received October 15, 1996; Revised Manuscript Received January 30, 1997[⊗]

ABSTRACT: The three-dimensional solution structure of the RNA self-complementary duplex $\begin{smallmatrix}\text{GGCAGGCC}\\ \text{CCGGACGG}\end{smallmatrix}$ 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 GA structure differs from the sheared tandem GA structure in $\begin{smallmatrix}\text{GGCGAGCC}\\ \text{CCGAGCGG}\end{smallmatrix}$ 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], although the only difference between the two sequences is the order of the two GA mismatches. Inspection of three-dimensional structures indicates that substituting $\begin{smallmatrix}5'\text{AG}3'\\ 3'\text{GA}5'\end{smallmatrix}$ for $\begin{smallmatrix}5'\text{GA}3'\\ 3'\text{AG}5'\end{smallmatrix}$ makes GA mismatches unable to have stable sheared conformations. This may explain why the $\begin{smallmatrix}5'\text{AG}3'\\ 3'\text{GA}5'\end{smallmatrix}$ motif is rarely observed in nature, whereas $\begin{smallmatrix}5'\text{GA}3'\\ 3'\text{AG}5'\end{smallmatrix}$ is common.

Little is known about the interactions that determine the structures and stabilities of non-Watson−Crick paired regions in RNA. A particularly interesting non-Watson−Crick motif is the tandem GA mismatch. Inspection of known RNA secondary structures (Gutell et al., 1992, 1993) indicates that the symmetric motif $\begin{smallmatrix}5'\text{CGAG}3'\\ 3'\text{GAGC}5'\end{smallmatrix}$ occurs often naturally, whereas $\begin{smallmatrix}5'\text{GGAC}3'\\ 3'\text{CAGG}5'\end{smallmatrix}$ and $\begin{smallmatrix}5'\text{XAGY}3'\\ 3'\text{YGAX}5'\end{smallmatrix}$, where X and Y form Watson−Crick pairs, are rare. All have similar thermodynamic stabilities, with $\begin{smallmatrix}5'\text{GGAC}3'\\ 3'\text{CAGG}5'\end{smallmatrix}$ being the most stable (Walter et al., 1994). The GA mismatches in $\begin{smallmatrix}5'\text{CGAG}3'\\ 3'\text{GAGC}5'\end{smallmatrix}$ form sheared pairs with hydrogen bonds from A-N7 to G-NH2 and A-NH6 to G-N3, whereas the GA mismatches in the other motifs form pairs with hydrogen bonds from A-NH6 to G-O6 and A-N1 to G-NH1 (imino) (see Figure 1). The NMR structures of (GGCGAGCC)$_2$ (SantaLucia & Turner, 1993) and (GCGGACGC)$_2$ (Wu & Turner, 1996) suggested that overlap of partial charges between mismatches and adjacent base pairs may be responsible for the different conformations. Here we report the NMR structure of (GGCAGGCC)$_2$ for comparison. The results combined with molecular modeling suggest the hydrogen-bonding pattern in the AG motif is determined by steric interactions, helical handedness, and the number of possible hydrogen bonds. Thus several factors must be considered in modeling three-dimensional structures of RNA from sequence.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Purification. rGGCAGGCC was synthesized and purified as described previously for rGCGGACGC (Wu & Turner, 1996). The purity of the rGGCAGGCC 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 by evaporation, and dissolved in 1.5 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. Under dry nitrogen, the sample was dissolved in 0.35 mL of 99.996% D_2O (or in 0.315 mL of H_2O and 0.035 mL of D_2O, for experiments in 90% H_2O and 10% D_2O) and placed in a microvolume NMR tube (Shigemi Inc.). The strand concentration was 4 mM.

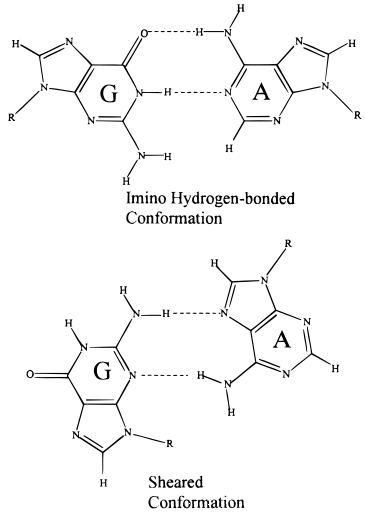

FIGURE 1: Two known hydrogen-bonding patterns of GA mismatches in a GA tandem: (top) imino hydrogen-bonded conformation; (bottom) sheared conformation.

NMR. NMR spectra were acquired on Varian VXR-500S and Unity-500 spectrometers and analyzed on 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 method (States et al., 1982). All proton

[†] This work was supported by NIH Grant GM 22939.
[‡] Coordinates and NMR-derived restraints have been deposited in the Brookhaven Protein Data Bank (PDB ID code 1MWG).
* Author to whom correspondence should be addressed.
[§] University of Rochester.
[||] Wayne State University.
[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0006-2960(96)02591-3 CCC: $14.00

and phosphorus chemical shifts are referenced to TSP [3-(trimethylsilyl)propionate] and phosphate buffer, respectively.

Imino proton spectra were recorded in 90% H_2O and 10% D_2O with the 1331 solvent suppression scheme (Hore, 1983) with 12 000 Hz sweep width. One-dimensional NOE experiments were acquired by irradiating for 1.25 s with a low-power decoupler setting prior to the read pulses. The maximum excitation was adjusted to 12.5 ppm with the first nodes at 20.1 and 4.9 ppm.

NOESY spectra in 100% D_2O were acquired at 35 °C with mixing times of 50, 100, 150, 250, and 400 ms and at 30 °C with a mixing time of 400 ms. NOESY spectra with five different mixing times at 35 °C were used for quantifying NOE cross peaks and calculating distance restraints in structural modeling. Suppression 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. A total of 256 complex FIDs were collected for each NOESY spectrum.

Separate DQF-COSY spectra were collected at 35 °C with spectral widths of 4000 and 700 Hz in both dimensions and at 40 °C with spectral widths of 2000 Hz in both dimensions. A $^1H-^{31}P$ HETCOR spectrum at 35 °C was collected using the pulse sequence described by Sklenar et al. (1986), with a spectrum width of 2000 Hz in the ^1H dimension and 750 Hz in the ^{31}P dimension.

Restraint Generation and Structural Modeling. Dihedral angle restraints were derived by measuring *J*-coupling constants from DQF-COSY and $^1H-^{31}P$ HETCOR spectra at 35 °C (Varani & Tinoco, 1991). *J*-coupling constants were measured by fitting the antiphase cross peaks automatically using Felix software. Dihedral angle (torsion angle) restraints were derived from three-bond *J*-coupling constants using Karplus equations: 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). The bound widths of dihedral angle restraints were calculated from errors of *J* measurements.

Initial distance restraints between nonexchangeable protons were obtained by measuring the 100 ms NOESY 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 (rGGCAGGCC)$_2$, however, the two-spin approximation can introduce errors in distances, especially long 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 by numerically solving for the entire spin-relaxation network, thus providing more accurate distance restraints than the isolated two-spin approximation. During this iterative process, the distance restraints are refined, and differences (*R*-factors) between the experimental NOE intensities and back-calculated NOE intensities serve as the criteria for convergence of the structure.

A-form and B-form duplexes were used as starting structures and refined with a simulated annealing protocol using 35 dihedral angle restraints and 114 initial distance restraints per strand. The converged structures were submitted to an RMA (relaxation matrix approach) calculation, and more accurate distance restraints were derived and used for the further refinement of the structures with the simulated annealing protocol. This process was repeated until *R*-factors no longer decreased, indicating the best match between experimental NOESY spectra and theoretical NOESY spectra back-calculated from the refined structure. Some representative *R*-factors are

$$R_1 = \frac{\sum|A^{\text{theo}} - A^{\text{expt}}|}{\sum A^{\text{expt}}}$$

$$R_2 = \frac{\sum\tau_m|A^{\text{theo}} - A^{\text{expt}}|}{\sum\tau_m A^{\text{expt}}}$$

$$R_3 = \frac{\sum|A^{\text{theo}} - A^{\text{expt}}|}{\sum 0.5(|A^{\text{theo}}| + |A^{\text{expt}}|)}$$

$$R_4 = \frac{\sum\tau_m|A^{\text{theo}} - A^{\text{expt}}|}{\sum\tau_m \times 0.5(|A^{\text{theo}}| + |A^{\text{expt}}|)}$$

$$R_5 = \frac{\sqrt{\sum(|A^{\text{theo}}|^{1/6} - |A^{\text{expt}}|^{1/6})^2}}{\sum|A^{\text{expt}}|^{1/6}}$$

$$R_6 = \frac{\sqrt{\sum[\tau_m(|A^{\text{theo}}|^{1/6} - |A^{\text{expt}}|^{1/6})]^2}}{\sum\tau_m|A^{\text{expt}}|^{1/6}}$$

where A^{expt} and A^{theo} are the NOE intensities of experimental and theoretical NOE matrices and τ_m is the mixing time.

Each RMA calculation started with building a theoretical relaxation matrix based on isotropic tumbling with a correlation time of 2.5 ns (determined by optimizing the match between experimental and theoretical NOEs) and interatomic distances in the converged structures from the previous simulated annealing refinement. Then theoretical NOE intensity matrices for each mixing time (50, 100, 150, 250, 400 ms) were computed from the eigenvalues and eigenvectors of the theoretical relaxation matrix. The theoretical and experimental NOE matrices were merged, and the merged NOE matrices were back-transformed to new relaxation matrices. The average back-calculated relaxation matrix elements were used to derive refined distance restraints. In all cases, these refined distance restraints were given error limits of ±20%.

Simulated annealing and IRMA calculations were performed using Discover software within Biosym Insight II on a Silicon Graphics workstation. Calculations employed distance and dihedral angle restraints and 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, angles, and dihedral angles. Nonbonding terms included van der Waals energies with a Lennard-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,

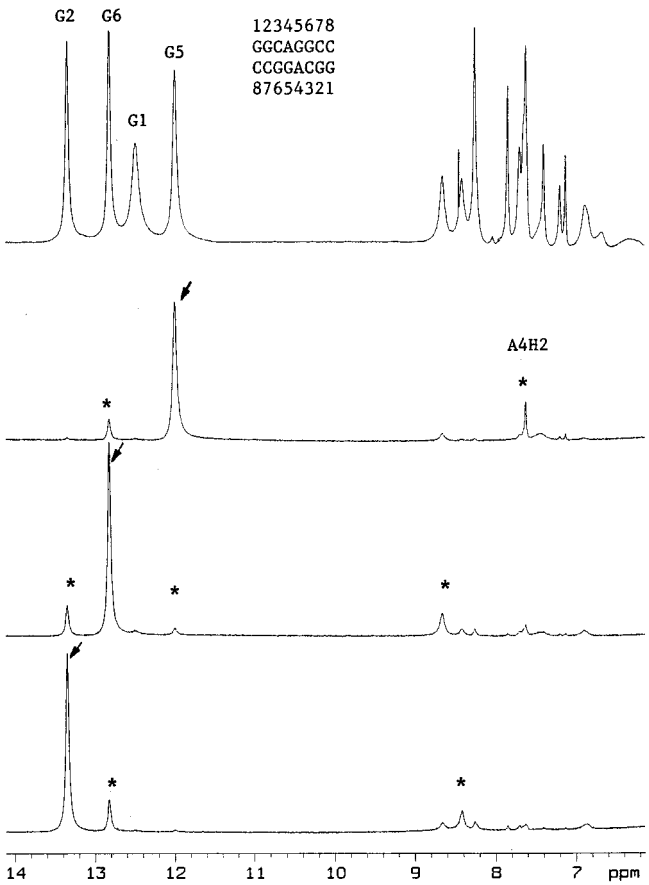

FIGURE 2: 500 MHz proton NMR and 1-D NOE difference spectra of 4.0 mM $(rGGCAGGCC)_2$ at 20 °C in 0.1 M NaCl, 10 mM sodium phosphate, and 1.5 mM EDTA in 90% H_2O and 10% D_2O, pH = 7. The off-resonance spectrum is on the top. Difference spectra between off-resonance and on-resonance spectra acquired with 1.25 s saturation at 12.00, 12.83, and 13.35 ppm are on the bottom. The saturated resonances are indicated by arrows while the observed NOEs are designated by asterisks. The secondary structure of $(rGGCAGGCC)_2$ is shown at the top of the figure.

using flat-bottomed quadratic potentials. The force constants for distance and dihedral angle restraints were set to 50 kcal/(mol Å^2) and 50 kcal/(mol rad^2). The simulated annealing protocol consists of a combination of restrained molecular dynamics (rMD) and restrained energy minimization (rEM): (a) 100 iterations of rEM; (b) 6 ps of rMD at 1000 K, with restraint terms increasing from 0.05 to 1 (full value), nonbonding terms increasing from 0.001 to 0.1, and covalent bonding terms increasing from 0.1 to 1 (full value); (c) 7 ps of rMD with temperature decreased from 1000 to 300 K gradually and nonbonding terms increasing from 0.1 to 1 (full value); (d) 3 ps of rMD at 300 K; (e) 500 iterations of rEM; and (f) 2 ps of rMD at 300 K to prevent the molecule from being trapped in a local energy minimum; and (g) 2000 iterations of rEM.

RESULTS

Resonance Assignments. The 1-D imino proton spectrum and NOE difference spectra at 20 °C are shown in Figure 2. Single imino resonances are observed for each of the four guanine imino protons of rGGCAGGCC. Since rGGCAGGCC can form a self-complementary duplex, $\overset{GGCAGGCC}{CCGGACGG}$, each imino resonance is attributed to two imino protons which have identical chemical shifts. This is consistent with 2-fold symmetry in the average three-dimensional structure. Imino proton assignments were based on sequential NOEs, G2H1−G6H1−G5H1, and on NMR melting experiments. Imino proton resonances of the terminal G1 and mismatched G5 are broader than those of the internal G2 and G6, indicating dynamics of terminal and mismatched base pairs. Both the chemical shift of G5H1 and the strong NOE from G5H1 to A4H2 indicate the GA mismatches have an imino hydrogen-bonded conformation (Figures 1 and 2).

All the nonexchangeable protons were assigned on the basis of careful comparisons among T_1 measurements, DQF-COSY, $^1H-^{31}P$ HETCOR, and NOESY spectra with different mixing times and at different temperatures, using strategies described by Varani and Tinoco (1991) (see Supporting Information for T_1 relaxation times and the 1-D proton spectrum at 35 °C; see paragraph at end of paper regarding Supporting Information). Chemical shift assignments of protons and phosphorus at 35 °C are summarized in Table 1.

The base H8/H6/H2 to H1′/H5 region of a 150 ms NOESY spectrum is shown in Figure 3. H5−H6 cross peaks of C3, C7, and C8 were identified because of their existence in the same regions of DQF-COSY spectra (data not shown). All the base H8/H6 and sugar H1′ protons are 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) as delineated in Figure 3. The resonance left at 7.62 ppm is assigned to A4H2 on the basis of its long T_1 relaxation time of 6.0 s. The assignment of A4H2 was confirmed by its strong sequential NOE to G5H1′ and medium interstrand NOE to G6H1′ (Wuthrich, 1986; Varani & Tinoco, 1991). Since H5−H6 cross peaks are the strongest in this region, there is no syn glycosyl conformation. Sequential H8/H6(n)−H8/H6(n+1) connectivity in 400 ms NOESY spectra confirms the assignments of base protons (see Supporting Information). Observation of weak sequential H1′(n)−H1′(n+1) connectivity in 400 ms NOESY spectra confirms the assignments of H1′ protons, although the connectivity is interrupted at the G5H1′−G6H1′ step (see Supporting Information).

H2′ proton resonances are assigned on the basis of short mixing time (50, 100, and 150 ms) NOESY spectra where spin diffusion does not significantly contribute to NOE intensities (Neuhaus & Williamson, 1989). The sugar to H1′/H5 region of a 150 ms NOESY spectrum is shown in Figure 4. Each sugar H1′ proton is closer to the H2′ proton of the same sugar than any other sugar proton regardless of sugar pucker. Thus the eight strongest peaks in this region are the H1′−H2′ cross peaks (Figure 4). The chemical shift degeneracy of G5H1′/C8H1′ and C3H1′/C7H1′ was partially resolved in the 400 ms NOESY spectrum at 30 °C, which facilitated the assignment of these H2′ protons. 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 150 ms NOESY spectrum in Figure 5. The alternating weak−strong NOE intensity pattern indicates an overall A-form geometry for the stem portion of the duplex. The weak−strong pattern is reversed at the mismatch region, however, where the A4H8−A4H2′ cross peak is bigger than the A4H2′−G5H8 cross peak, and both of them are of medium intensity. This suggests deviation from A-form geometry at the A4 nucleotide and A4−G5 juncture. H2′ assignments of G1, A4, and C8 were confirmed by intrasugar H1′−H2′ scalar cross peaks

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

base	H8/H6	H2/H5	H1′	H2′	H3′	H4′	H5′/H5″ [b]	imino[c]	phosphorus
G1	7.93	na[d]	5.67	4.88	4.52	4.28	3.97/3.88	12.50	na
G2	7.46	na	5.86	4.51	4.59	4.52	4.54/4.14	13.35	−2.16
C3	7.67	5.27	5.45	4.06	4.66	4.38	4.63/4.13	na	−2.61
A4	8.27	7.62	6.10	4.88	4.84	4.52	4.52/4.19	na	−2.39
G5	7.22	na	5.71	4.67	4.23	4.59	4.30/4.22	12.00	−1.77
G6	7.16	na	5.81	4.48	4.55	4.47	4.52/4.11	12.83	−2.33
C7	7.72	5.25	5.44	4.20	4.46	4.38	4.54/4.06	na	−2.82
C8	7.63	5.48	5.70	4.03	4.16	4.16[e]	4.46/4.03	na	−2.33

[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 20 °C. [d] Not applicable. [e] For the related sequence, (rGGCGAGCC)$_2$, C8H4′ was originally assigned to 4.55 ppm (SantaLucia & Turner, 1993). The correct assignment is 4.16 ppm, consistent with that observed for (rGGCAGGCC)$_2$.

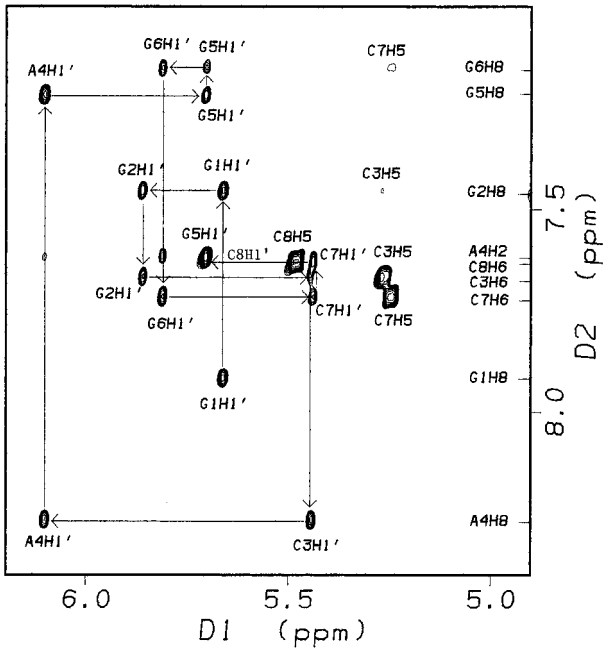

FIGURE 3: H8/H6/H2−H1′/H5 region of the 150 ms NOESY of (rGGCAGGCC)$_2$ at 35 °C in 100% D_2O. Assignments for the D_2 dimension are on the right axis. The rest of the assignments are for the D_1 dimension. Sequential −H8/H6(n)−H1′(n)−H8/H6/(n+1)− connectivities are shown in solid lines.

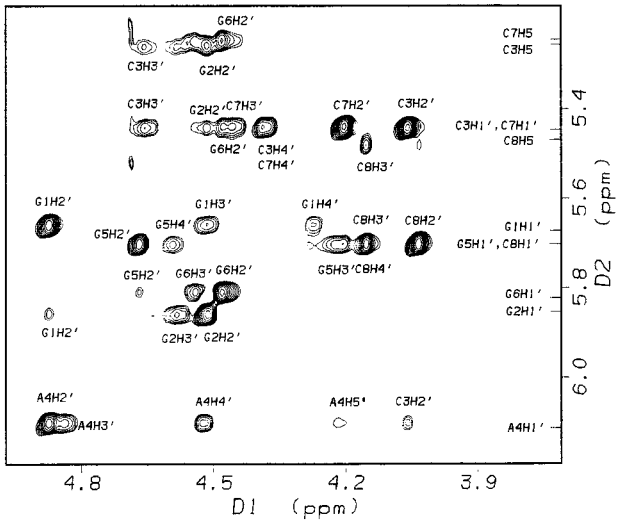

FIGURE 4: Sugar−H1′/H5 region of the 150 ms NOESY of (rGGCAGGCC)$_2$ at 35 °C in 100% D_2O. Assignments for the D_2 dimension are on the right axis. The rest of the assignments are for the D_1 dimension.

in the sugar−H1′/H5 region of DQF-COSY spectra. In DQF-COSY spectra, weak H1′−H2′ scalar peaks are observable only for G1, A4, and C8, suggesting G1, A4, and C8 have mixed C3′-endo and C2′-endo sugar puckers, with relative populations of 70% and 30% for C3′-endo and C2′-endo, respectively. All the other nucleotides have predominantly C3′-endo sugar puckers (over 90% of the population). Observation of mixed sugar pucker for G1, C8, and A4 suggests dynamics at the ends of the helix and in the mismatch region. With a 5° increase of the temperature to 40 °C, most H1′−H2′ scalar cross peaks are faintly visible, indicating the C2′-endo population increases with temperature and the molecule becomes dynamic above 35 °C. Thus structure modeling was based on data at 35 °C.

H3′, H4′, H5′, and H5″ resonances were assigned from DQF-COSY (Figure 6) and ^{1}H−^{31}P HETCOR spectra (Figure 7). 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 overlap and small coupling constants. H2′−H3′ cross peaks can be easily identified from their characteristic thin pattern, since the H2′−H3′ J-coupling constant is about 5 Hz regardless of sugar pucker. H2′−H3′ scalar cross peaks for G2, A4, and G6 can hardly be seen, since they are very close to diagonal peaks and overlapped with other peaks. H3′s of G2, A4, and G6 were, however, identified in the ^{1}H−^{31}P HETCOR spectrum from their strong scalar coupling to the phosphorus of the adjacent nucleotide on the 3′ side. In the ^{1}H−^{31}P HETCOR spectrum, H3′(n−1)−P(n) cross peaks are the strongest due to the large coupling (about 10 Hz) between them, and some P(n)−H5′/H5″/H4(n) cross peaks can be seen. Assignment of H3′ resonances can be confirmed by their medium NOEs to H8/H6 of their own bases and the next base in the 3′ direction in a 400 ms NOESY spectrum (Figure 5). H4′ resonances were therefore assigned from H3′−H4′ scalar cross peaks in the DQF-COSY spectrum (Figure 6). 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 either overlapped or too weak to be seen due to the small amplitude of the coupling constants (less than 3 Hz). Observation of P(n)−H5′/H5″ scalar cross peaks in a ^{1}H−^{31}P HETCOR spectrum and H1′(n)−H5′/H5″(n and n+1), H8/H6(n)−H5′/H5″(n) NOE cross peaks in 400 ms NOESY spectra helped to assign H5′ and H5″ for all eight nucleotides, although assignments are not stereospecific so H5′ is assumed to resonate downfield from H5″ (Varani & Tinoco, 1991). No distance

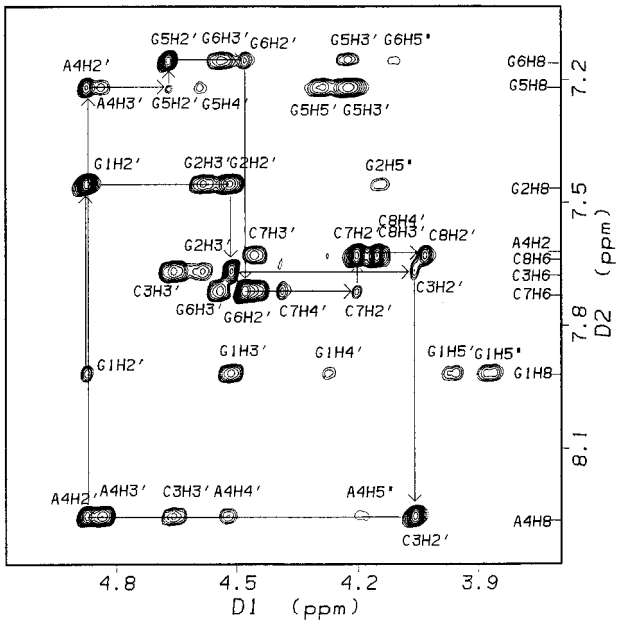

FIGURE 5: Sugar−H8/H6/H2 region of the 400 ms NOESY of (rGGCAGGCC)$_2$ at 35 °C in 100% D_2O. Assignments for the D_2 dimension are on the right axis. The rest of the assignments are for the D_1 dimension. The sequential H8/H6(n)−H2′(n+1)−H8/H6(n+1) connectivities are shown in solid lines.

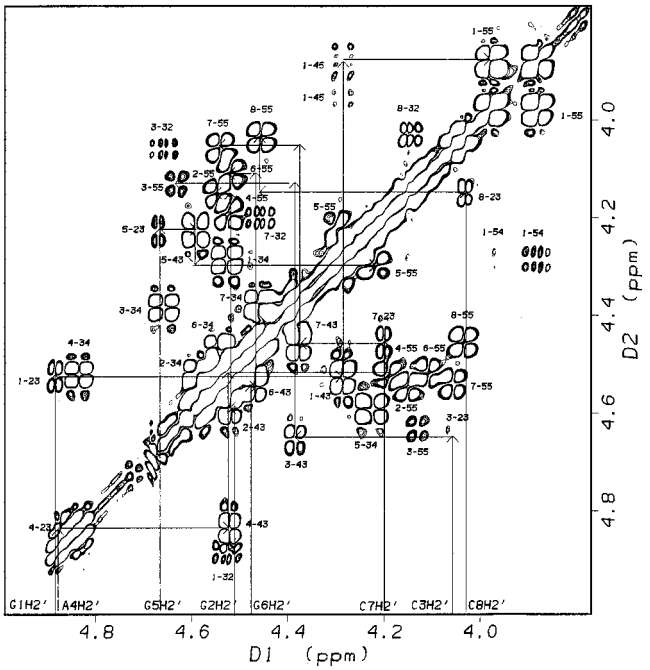

FIGURE 6: Sugar−sugar region of the DQF-COSY of (rGGCAGGCC)$_2$ at 35 °C in 100% D_2O. The intrasugar scalar H2′−H3′−H4′−H5′/H5″ connectivities are shown in solid lines for all eight bases. On the connectivity pathways, some of the cross peaks are missing due to overlap and small coupling constants. Assignments are adjacent to the antiphase cross peaks. In the assignment notation a−bc, a represents the sequence number of the nucleotide, and b and c denote the protons for the cross peak. For example, 8−32 represents the C8H3′−C8H2′ cross peak.

involving H5′/H5″ protons was 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 provides confidence in their accuracy and reliability.

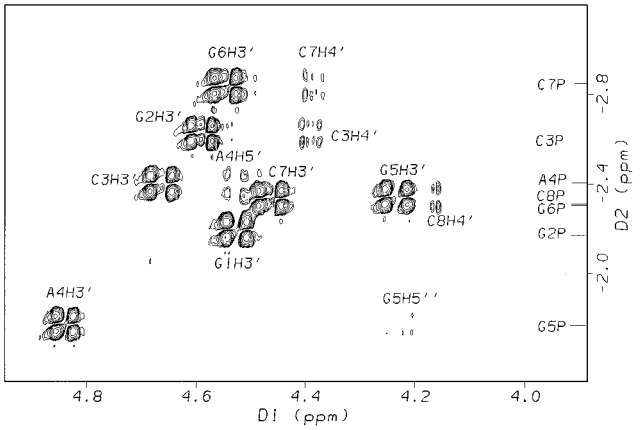

FIGURE 7: ^{1}H−^{31}P HETCOR of (rGGCAGGCC)$_2$ at 35 °C in 100% D_2O. Proton chemical shift (D_1 dimension) assignments are shown at the peaks, while phosphorus chemical shift (D_2 dimension) assignments are shown on the right axis.

A ^{1}H−^{31}P HETCOR spectrum reveals that all seven phosphorus resonances are within 1.1 ppm, indicating that double AG mismatches are incorporated into an A-form helix without major backbone distortion. Assignments of ^{31}P chemical shifts were confirmed by a ^{1}H−^{31}P hetero-TOCSY NOESY experiment (data not shown), as described by Kellogg et al. (1992).

Structural Modeling. Table 2 summarizes proton−proton and proton−phosphorus *J*-coupling constants measured in DQF-COSY and ^{1}H−^{31}P HETCOR spectra, respectively. A total of 35 dihedral angle restraints per strand (Table 3) were derived from these *J*-coupling constants by solving Karplus equations (see Materials and Methods).

Dihedral angles ν_1, ν_2, and δ ($=\nu_3 + 125°$) were derived from *J*-coupling of H1′−H2′, H2′−H3′, and H3′−H4′, respectively. The errors of *J*-coupling measurements were used to derive upper and lower bounds for these dihedral angle restraints. On the basis of Karplus equations, usually one to four dihedral angles are consistent with a particular *J* value. Thus dihedral angle restraints for ν_1, ν_2, and δ all have more than one range. Dihedral angle restraints for β were derived from P−H5′/H5″ couplings. Since none of the P−H5′ or P−H5″ couplings is more than 5.5 Hz, β angles have to be trans for all seven nucleotides (from G2 to C8), based on the plot of *J*-coupling constants of P−H5′ and P−H5″ with β (Wijmenga et al., 1993). The γ angles were derived from H4′−H5′ and H4′−H5″ couplings. Since the sum of H4′−H5′ and H4′−H5″ *J*-coupling constants is less than 6 Hz for G2, C3, G5, G6, C7, and C8, g^+ will be the predominant conformation for γ (Wijmenga et al., 1993). The ϵ angles were derived from H3′(n)−P(n+1) *J*-couplings. Out of two ranges of ϵ, trans and g^+, possible from the Karplus equation, g^+ is sterically disallowed (Wijmenga et al., 1993) and ϵ is restrained to be trans for all seven nucleotides from G2 to C8. The α and ζ angles are related to phosphorus chemical shifts (Gorenstein, 1984). No quantitative relationship has been established, however, so they are not restrained. The glycosidic angle, χ, is not restrained.

A total of 105 NOE distance restraints (53 interresidue, 50 intraresidue, and 2 interstrand) were derived for each strand from a 100 ms NOESY spectrum. Eighteen hydrogen-

Table 2: *J*-Coupling Constants (Hz) for (rGGCAGGCC)$_2$ at 35 °C[a]

residue	H1′−H2′	H2′−H3′	H3′−H4′	H4′−H5′	H4′−H5″	H5−H6	H5′−H5″
G1	4.3 ± 2.0	5.2 ± 2.0	9.3 ± 3.0	6.0 ± 2.0	4.3 ± 2.0	na	10.3 ± 3.0
G2	<2	overlap	9.1 ± 3.0	<3	<3	na	11.2 ± 3.0
C3	<2	5.0 ± 2.0	9.7 ± 3.0	<3	<3	7.2 ± 2.0	10.2 ± 3.0
A4	4.8 ± 2.0	overlap	8.2 ± 3.0	overlap	overlap	na	12.5 ± 3.0
G5	<2	4.9 ± 2.0	14.7 ± 3.0	<3	<3	na	10.3 ± 3.0
G6	<2	overlap	10.4 ± 3.0	<3	<3	na	11.1 ± 3.0
C7	<2	5.2 ± 2.0	9.1 ± 3.0	<3	<3	6.8 ± 2.0	10.5 ± 3.0
C8	4.2 ± 2.0	5.9 ± 2.0		<3	<3	7.0 ± 2.0	10.5 ± 3.0

5′-base	H3′−P	P−H5′	P−H5″	P−H4′	3′-base
G1	11.1 ± 3.0	<5	<3	overlap	G2
G2	10.5 ± 3.0	<3	<3	5.3 ± 2.0	C3
C3	11.8 ± 3.0	<5	<3	overlap	A4
A4	10.5 ± 3.0	3.5 ± 2.0	3.0 ± 2.0	<3	G5
G5	10.7 ± 3.0	<5	<3	overlap	G6
G6	11.5 ± 3.0	<5	<3	5.5 ± 2.0	C7
C7	11.1 ± 3.0	<5	<3	5.7 ± 2.0	C8

[a] Proton−proton and proton−phosphorus coupling constants were measured from DQF-COSY and ^{1}H−^{31}P HETCOR spectra, respectively, at 35 °C.

Table 3: Dihedral Angle Restraints for (rGGCAGGCC)$_2$ (deg)[a]

residue	α	β	γ	ν_1	ν_2	δ (ν_3 + 125°)	ε	ζ	χ[b]
G1	na	na	nr	nr	40 ± 13, −40 ± 13, 132 ± 12, −132 ± 12	60 ± 42, −120 ± 34	−120 ± 50	nr	nr
G2	nr	180 ± 40	58 ± 40	−32 ± 26, 152 ± 26	nr	60 ± 43, −120 ± 36	−120 ± 50	nr	nr
C3	nr	180 ± 40	58 ± 40	−32 ± 26, 152 ± 26	42 ± 12, −42 ± 12, 131 ± 11, −131 ± 11	60 ± 39, −120 ± 32	−120 ± 50	nr	nr
A4	nr	180 ± 40	nr	nr	nr	nr	−120 ± 50	nr	nr
G5	nr	180 ± 40	58 ± 40	−32 ± 26, 152 ± 26	42 ± 12, −42 ± 12, 131 ± 11, −131 ± 11	nr	−120 ± 50	nr	nr
G6	nr	180 ± 40	58 ± 40	−32 ± 26, 152 ± 26	nr	60 ± 36, −120 ± 27	−120 ± 50	nr	nr
C7	nr	180 ± 40	58 ± 40	−32 ± 26, 152 ± 26	40 ± 13, −40 ± 13, 132 ± 12, −132 ± 12	60 ± 43, −120 ± 36	−120 ± 50	nr	nr
C8	nr	180 ± 40	58 ± 40	nr	36 ± 13, −36 ± 13, 136 ± 11, −136 ± 11	nr	na	na	nr
A-form[c]	−68	178	54	−25	38	82	−153	−71	−158

[a] Definition of dihedral angles: α = O3′−P−O5′−C5′, β = P−O5′−C5′−C4′, γ = O5′−C5′−C4′−C3′, ν_1 = O4′−C1′−C2′−C3′, ν_2 = C1′−C2′−C3′−C4′, δ = C5′−C4′−C3′−O3′, ν_3 = O4′−C4′−C3′−C2′, ε = C4′−C3′−O3′−P, and ζ = C3′−O3′−P−O5′. Abbreviations: na = not applicable; nr = not restrained. [b] Glycosidic angle. [c] A-form values are from Saenger (1984).

bonding distance restraints (1.8 ± 0.2 Å) per *duplex* were used to maintain six Watson−Crick GC base pairs. No hydrogen-bonding restraints were imposed between mismatched G and A.

Both A-form and B-form starting structures were submitted to simulated annealing with initial restraints. For each starting structure, 15 structures were generated with the simulated annealing protocol and randomized trajectories. The resulting 30 structures overlapped well with a pairwise rmsd of 0.81 Å. IRMA was then performed on the ensemble average of these 30 structures. Only 75 of 105 NOE distance restraints were submitted to IRMA for refinement, since 30 distance restraints were derived from weak and partially overlapped cross peaks. The resulting structure was then taken through the simulated annealing protocol 10 separate times. At least 8 of 10 resulting structures overlapped well. IRMA was performed again on the ensemble average of well-overlapped structures. Over the course of five IRMA iterations, convergence of structure was indicated by the leveling off of the *R*-factors (Figure 8), which means a best fit between the final structure and experimental NOESY spectra. For the fifth refinement, the structure was taken through the simulated annealing protocol 15 separate times. The final structure is the average of these 15 final overlapped structures (Figure 9) which have a pairwise rmsd of 0.41 Å. Stereoviews of the final NMR structure are shown in two different angles in Figure 10. A major groove view is shown in space-filling representation in Figure 11. The final IRMA refined distance restraints are summarized in Table 4.

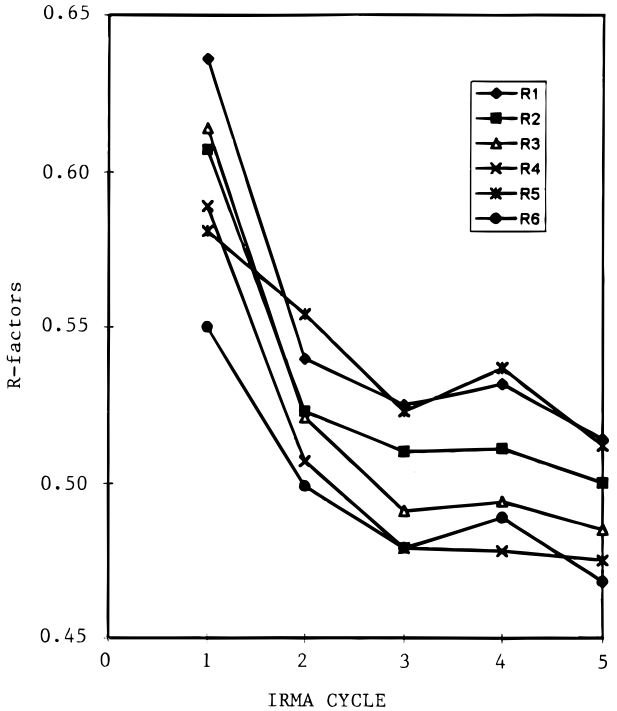

FIGURE 8: *R*-factors decreased and leveled off over the course of IRMA cycles. See Materials and Methods for definitions of R_1, R_2, R_3, R_4, R_5, and R_6. IRMA was terminated after five iterations.

In the final structure, all the distances and dihedral angle restraints were satisfied within 0.4 Å and 4°, respectively.

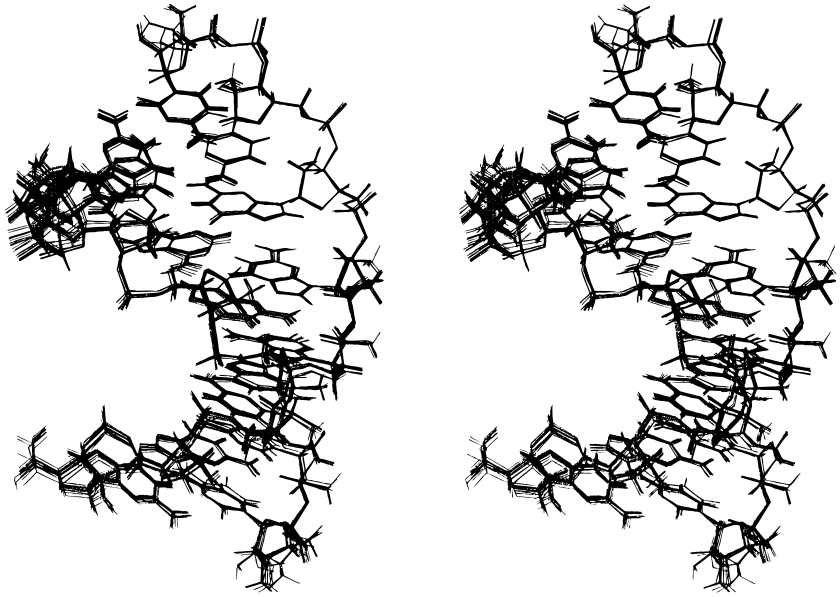

FIGURE 9: Superposition of 15 overlapped final structures of (rGGCAGGCC)$_2$ after IRMA.

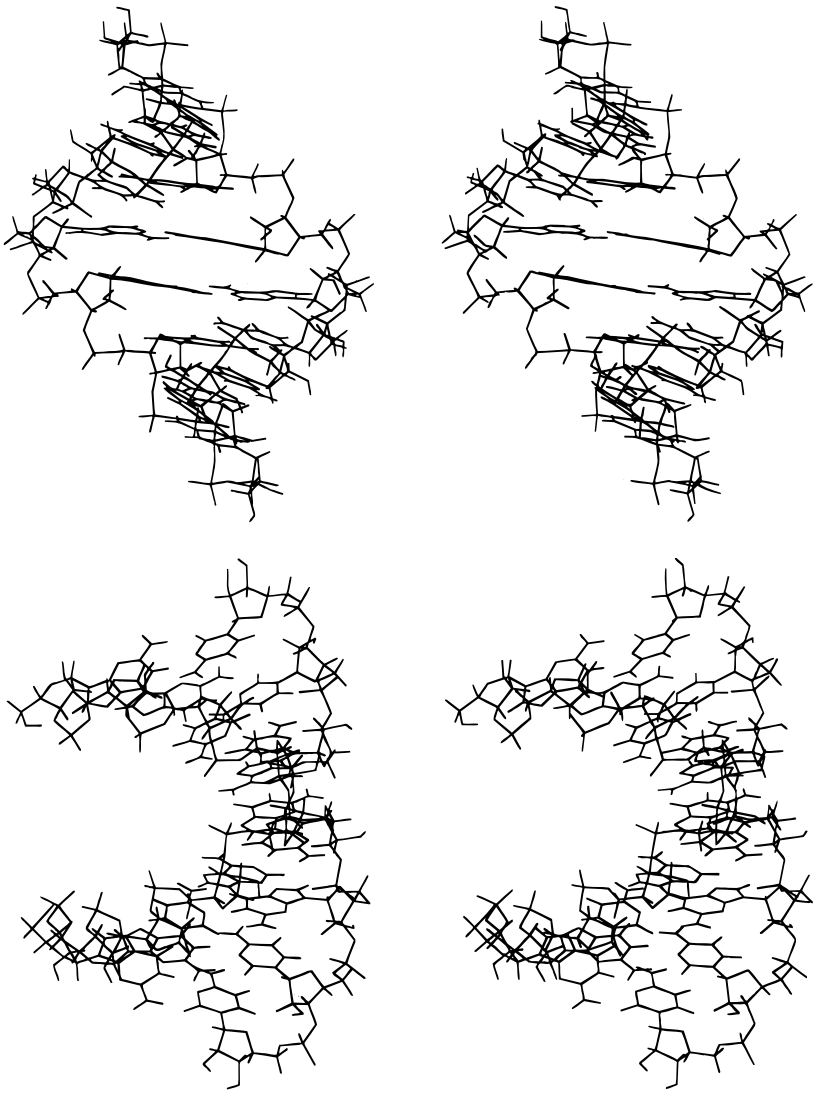

FIGURE 10: Stereoviews of the final average structure of (rGGCAGGCC)$_2$ at two different angles.

Figure 12 shows the plot of restrained distances and actual distances of the final structure. The great majority of distances satisfy the restraints within ±20% of the center of distance restraints.

A much more rigorous test of structural accuracy is back-calculation of the NOESY spectra based on the final structure. Figures 13 and 14 and Supporting Information show comparisons of representative regions of experimental

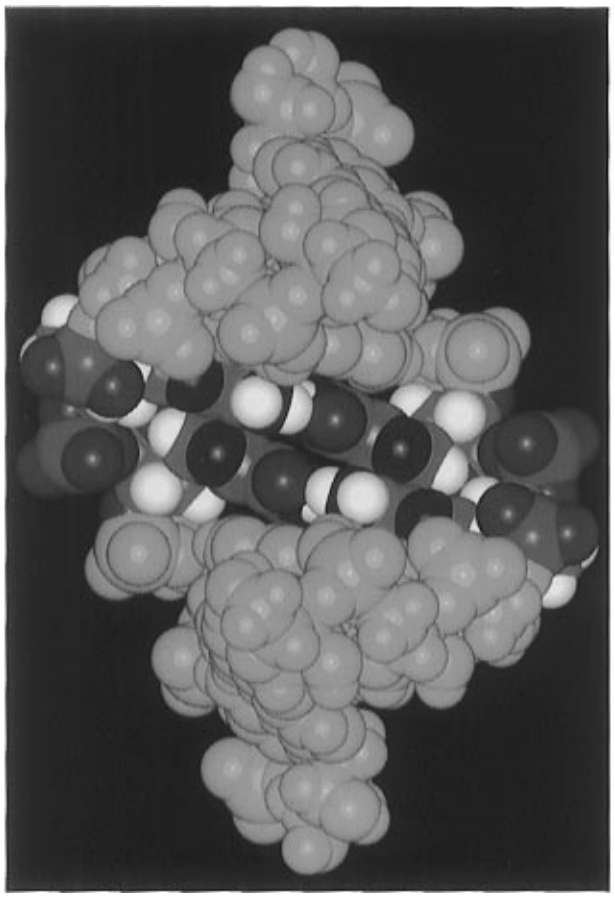

FIGURE 11: Space-filling view from the major groove of the final structure of $(rGGCAGGCC)_2$. The AG tandem mismatches are colored according to atoms.

and back-calculated 400 ms NOESY spectra. In general, there is a good match between experimental and calculated spectra. Most discrepancies can be attributed to lack of restraints, potential dynamics, or the difficulty of comparing three-dimensional data when presented as two-dimensional contour plots. For example, in Figure 14, peaks for G1H8 to G1H4′, H5′, and H5″ are larger in the measured spectrum than in the calculated one. G1H4′, H5′, and H5″ are not well defined since they have only 2, 0, and 0 restraints, respectively. Moreover, since G1 is a terminal nucleotide, it probably has dynamic motion which is not included in the analysis. Dynamics in the GA mismatch may be responsible for discrepancies involving G5 protons, G5H1′−G6H8 (Figure 13) and G5H8−G5H4′ (Figure 14). The peaks for G2H8−C3H5 and G2H8−G2H5″ are not seen in the calculated spectra of Figures 13 and 14, respectively, but are seen when the threshold is lowered. In general, backbone and terminal nucleotides are the least well determined parts of the structure. Nevertheless, with few exceptions, the good match between experimental and back-calculated spectra indicates that the final structure not only satisfies the proton−proton distance restraints but also is consistent with the entire proton relaxation network that generates the time-dependent NOESY spectra. The latter criterion is much more rigorous and difficult to satisfy than simple distance agreement.

The average pairwise rmsd of 0.41 Å reflects good overlap and quality of the final structures (Figure 9). Since NOEs can be observed only for proton pairs within 6 Å, NOE distance restraints are short range and the errors accumulate

Table 4: Final Distance Restraints (Å) for $(rGGCAGGCC)_2$ at 35 °C (Each Strand)[a]

intra	H1′−H8/H6	H2′−H8/H6	H3′−H8/H6	H4′−H8/H6	H1′−H5	H2′−H5	H3′−H5	H2′−H1′	H3′−H1′	H4′−H1′
G1	**3.3 ± 0.5**	**3.3 ± 0.6**	**3.0 ± 0.5**	**3.6 ± 0.6**				**2.5 ± 0.4**	**3.0 ± 0.7**	**3.2 ± 0.8**
G2	**3.3 ± 0.5**		**2.7 ± 0.4**					**2.4 ± 0.4**	**3.0 ± 0.5**	
C3	4.0 ± 1.0	**3.2 ± 0.6**	**2.4 ± 0.4**	4.5 ± 1.5	5.5 ± 1.5	4.5 ± 1.0	**3.2 ± 0.6**	**2.4 ± 0.4**	**2.9 ± 0.6**	**3.1 ± 0.6**
A4	**3.4 ± 0.6**	**2.5 ± 0.4**		**2.9 ± 0.6**				**2.5 ± 0.4**	**3.2 ± 0.5**	**3.6 ± 0.6**
G5	**3.6 ± 0.6**	4.5 ± 1.0	2.7 ± 0.5	**3.4 ± 0.8**				**2.8 ± 0.5**	**3.0 ± 0.5**	
G6	**3.7 ± 0.6**	**3.0 ± 0.6**	**2.9 ± 0.4**					**2.3 ± 0.4**	3.5 ± 1.0	
C7	**3.4 ± 0.5**	4.5 ± 1.5	**2.5 ± 0.5**	**3.1 ± 0.5**		4.5 ± 1.0		**2.4 ± 0.4**		
C8		**3.0 ± 0.5**	**3.0 ± 1.0**		5.5 ± 1.5	4.5 ± 1.0		**2.3 ± 0.4**	3.0 ± 0.6	

inter	H1′−H8/H6	H2′−H8/H6	H3′−H8/H6	H4′−H8/H6	H1′−H5	H2′−H5	H1′−H1′	H2′−H1′	H8/H6−H8/H6	H8/H6−H5	H5−H8/H6	H2−H1′	H2−H8	H5−H5
G1−G2	**3.7 ± 0.6**	**2.3 ± 0.4**		**4.4 ± 0.7**			5.5 ± 1.5	**3.8 ± 0.7**	**4.1 ± 0.7**					
G2−C3	**3.9 ± 0.8**	**2.5 ± 0.4**	**2.9 ± 0.5**		5.5 ± 1.5	**2.7 ± 0.6**	5.5 ± 1.5	**4.0 ± 0.8**	5.5 ± 1.5	**4.0 ± 0.8**				
C3−A4	**3.8 ± 0.7**	**2.3 ± 0.4**	**2.8 ± 0.4**	4.0 ± 0.5			5.5 ± 1.5	**3.7 ± 0.6**	**4.1 ± 1.0**		4.5 ± 1.0			
A4−G5	**3.1 ± 0.5**	**2.8 ± 0.4**	**3.4 ± 0.6**	5.5 ± 1.5			5.5 ± 1.5		**4.3 ± 0.8**			**2.6 ± 0.4**	4.5 ± 1.0	
G5−G6	**3.9 ± 0.8**	**2.5 ± 0.4**	**3.4 ± 0.6**				5.5 ± 1.5	**3.9 ± 0.7**	5.5 ± 1.5					
G6−C7	**4.1 ± 0.7**	**2.1 ± 0.4**	**2.8 ± 0.5**		4.0 ± 1.0	**2.8 ± 0.6**	5.5 ± 1.5		**4.5 ± 1.6**	**4.5 ± 1.0**				
C7−C8	**3.3 ± 0.6**	**2.3 ± 0.4**	**2.8 ± 0.5**			4.0 ± 1.5	5.5 ± 1.5		5.5 ± 1.5	4.5 ± 1.0				**3.8 ± 0.7**

[a] Plus (1) two interstrand restraints, G6H1′−A4H2 = 3.9 ± 1.4 Å and G5H1−A4H2 = 2.5 ± 1.5 Å, and (2) nine hydrogen-bonding restraints: 1.8 ± 0.2 Å. The total 114 distance restraints = 50 (intraresidue) + 53 (interresidue) + 2 (interstrand) + 9 (H-bond). The 75 IRMA-refined distance restraints are shown in bold letters. The other restraints were not refined by IRMA.

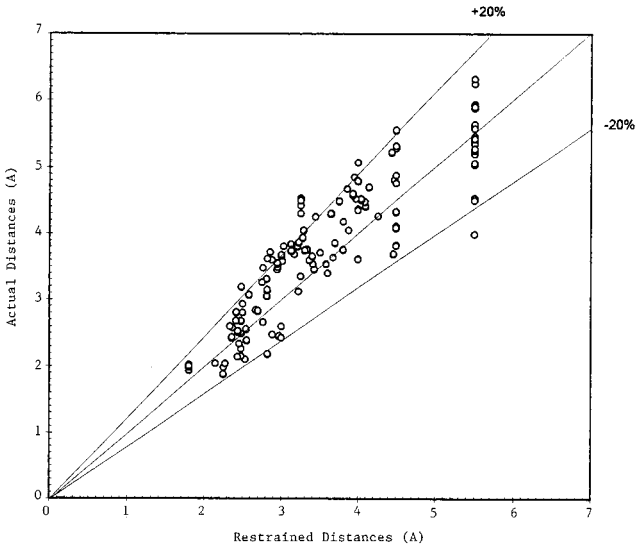

FIGURE 12: Correlation between the center of distance restraints (from Table 4) and the actual distances from the final averaged structure of (rGGCAGGCC)$_2$ for all the distances restrained. Note that experimental error limits for some distances are greater than ±20%.

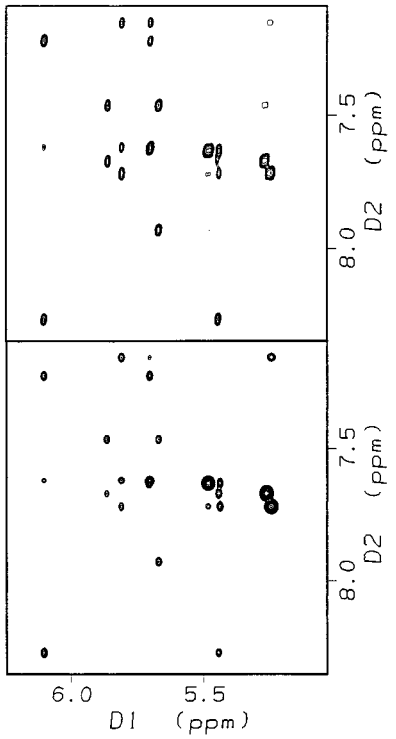

FIGURE 13: Experimental (upper) and back-calculated (lower) 400 ms NOESY spectra of the base H8/H6/H2 to H1′/H5 region.

along the helix. That is one reason the central base pairs overlap better than the terminal base pairs. A look at the mismatch region of the final average structure in Figure 15 (also see Supporting Information) indicates both GA mismatches form imino hydrogen-bonded conformations, and each GA mismatch has two hydrogen bonds: A-N1 to G-NH1 and A-NH6 to G-O6. In the average final structure, these two pairs of atoms are both within hydrogen-bonding range: 1.97 Å for AN1-GNH1 and 2.09 Å for ANH6-GO6. Figure 15 also shows the stacking between the adjacent CG pair and GA mismatch and between the two GA mismatches. Helical underwinding (27°) between the two AG mismatches

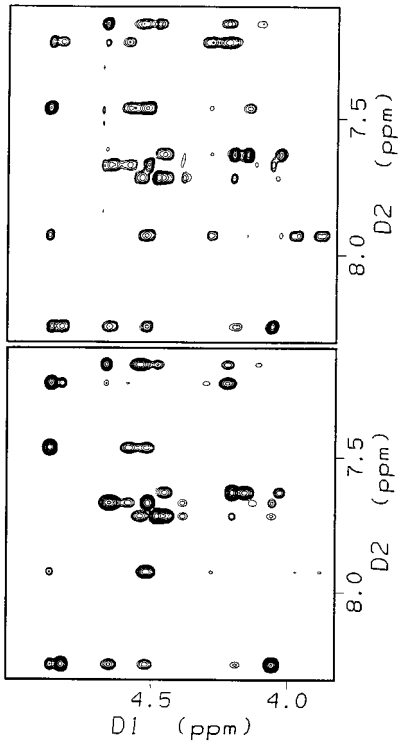

FIGURE 14: Experimental (upper) and back-calculated (lower) 400 ms NOESY spectra of the base H8/H6/H2 to sugar region.

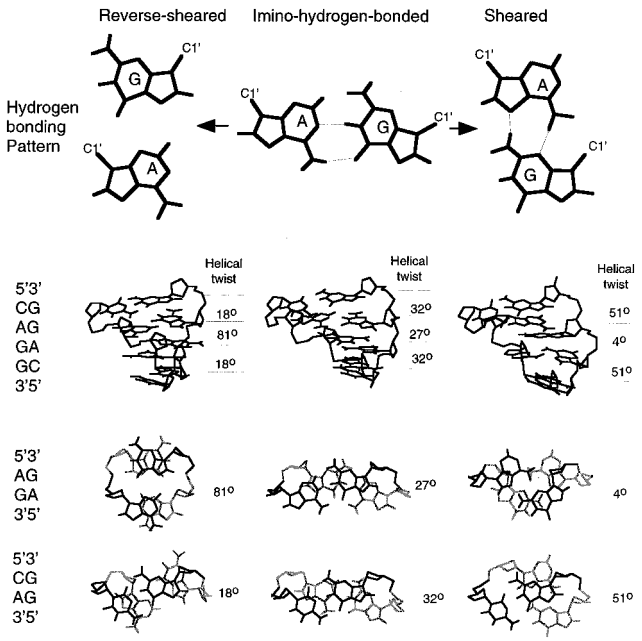

FIGURE 15: Structural comparison of the imino hydrogen-bonded (experimental), reverse-sheared [by flipping A and G of GA mismatches of (rGGCGAGCC)$_2$, keeping the coordinates], and sheared [by forcing the G and A into the sheared conformation with artifical restraints while maintaining the distance and dihedral angle restraints for Watson−Crick GC pairs of (rGGCAGGCC)$_2$] conformations for (rGGCAGGCC)$_2$. GA mismatch geometry is shown at the top with dotted lines indicating the hydrogen bonds. Below these are the 3-D structures of tandem AG mismatches and two adjacent CG pairs with helix twist angles listed for each step. Stacking between two AG mismatches and between the AG mismatch and CG closing base pairs is shown at the bottom with helical twist angles.

facilitates interstrand stacking between them. The GA mismatch conformation in (rGGCAGGCC)$_2$ is totally different from the sheared GA conformation in (rGGC-

Table 5: Dihedral Angles of the Final Structure of (rGGCAGGCC)$_2$ (deg)

residue	α	β	γ	ν_1	ν_2	δ	ε	ζ	χ	P[a]
G1	na[b]	na	66.1	−28.8	45.1	62.7	−173.2	−65.1	−161.2	20.8
G2	−71.4	−180.3	57.0	−33.3	45.5	71.1	−171.0	−60.5	−166.4	12.3
C3	−74.6	−176.6	52.5	−36.9	47.7	62.5	−169.9	−61.3	−156.7	8.7
A4	−72.7	−185.1	49.8	−31.0	39.8	76.8	−173.5	−59.2	−152.2	5.8
G5	−73.3	−182.1	61.3	−35.8	48.3	64.4	−168.2	−58.4	−175.4	10.1
G6	−71.1	−172.7	53.4	−31.7	44.3	72.8	−169.9	−58.3	−160.0	14.1
C7	−73.6	−179.5	54.1	−39.2	46.6	71.8	−172.0	−57.2	−155.6	1.4
C8	−74.8	−181.1	51.6	−15.8	34.2	76.1	na	na	−164.3	32.3
A-form[c]	−68	178	54	−25	37	82	−153	−71	−158	18

[a] Pseudorotation angles. [b] Not applicable. [c] A-form values are from Saenger (1984) and Roberts (1993).

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

base steps	helical twist (deg)	rise (Å)	slide (deg)	roll (deg)	tilt (deg)	interstrand P−P distance[c] (Å)
1	32.1	2.4	−1.8	3.0	0.5	17.3
2	35.9	2.9	−1.2	−2.1	−0.6	17.5
3	31.6	2.9	−1.3	1.0	−1.9	18.0
4	27.1	2.1	−2.3	6.0	0.1	18.9
5	31.7	2.9	−1.3	1.1	2.1	18.0
6	35.8	2.9	−1.2	−2.4	0.5	17.5
7	31.9	2.4	−1.8	3.0	−0.7	17.3
average	32.3	2.7	−1.6	1.4	0.0	17.8
A-form	32.7	2.8	−1.5	−0.4	13.0	17.5

[a] The parameters were calculated using NEWHEL92 from the Brookhaven Protein Data Bank. [b] Base steps are defined as $^{G}_{C}$−1−$^{C}_{G}$−2−$^{C}_{G}$−3−$^{A}_{G}$−4−$^{G}_{A}$−5−$^{G}_{C}$−6−$^{C}_{G}$−7−$^{C}_{G}$. [c] The interstrand phosphorus−phosphorus distance is defined as the distance between *nP* and (10 − *n*)′*P*, *n* = 2, 3, 4, ..., 8, e.g., G2P−C8′P, where the prime indicates the opposite strand.

GAGCC)$_2$ (SantaLucia & Turner 1993) but similar to that in (rGCGGACGC)$_2$ (Wu & Turner, 1996).

Torsion angles and helical parameters are shown in Tables 5−7. (rGGCAGGCC)$_2$ has an overall A-form geometry with minor distortion at the mismatch region. Purine−purine GA mismatches are incorporated into the A-form geometry evolved for Watson−Crick purine−pyrimidine pairs. The bulkiness of the two purines in the GA mismatches expands the backbone at the mismatch region. This explains the long interstrand phosphorus−phosphorus distance of 18.9 Å (compared to 17.5 Å for A-form), the long interstrand C1′−C1′ distance of 13.0 Å (compared to 10.9 Å for A-form), and the short helical rise of 2.1 Å (compared to 2.8 Å for A-form) at the mismatch step. The expansion of the backbone and decrease in helical rise propagate only to the adjacent base pairs along the helix. This is consistent with the nearest-neighbor model often used for prediction of RNA secondary structure based on thermodynamics (Turner et al., 1988).

DISCUSSION

Many important RNA sequences are being discovered, but little is known about structure−function relationships in RNA because only a few three-dimensional structures of natural RNAs or fragments thereof have been determined (Kim et al., 1974; Robertus et al., 1974; Moras et al., 1980; Woo et al., 1980; Varani & Tinoco, 1991; Heus & Pardi, 1991; SantaLucia & Turner, 1993; Pley et al., 1994a,b; Scott et al., 1995; Szewczak & Moore, 1995; Puglisi et al., 1995; Greenbaum et al., 1995; Ye et al., 1995; Allain & Varani, 1995; Gubser & Varani, 1996). Despite this lack of structural information, many RNAs are being considered as targets for therapeutics (Mol & van der Krol, 1991; Agrawal, 1995). A rational approach to modeling RNA structures and designing such therapeutics requires knowledge of the interactions that are important for determining RNA structure and stability. Tandem GA mismatches are an attractive motif for providing information on these interactions because their structures and stabilities are very sensitive to sequence, both in RNA (SantaLucia et al., 1990; Walter et al., 1994; Wu & Turner, 1996) and in DNA (Li et al., 1991a,b; Cheng et al., 1992; Li & Agrawal, 1995).

The $^{5'CAGG3'}_{3'GGAC5'}$ motif studied here is rare in natural secondary structures of RNA (SantaLucia et al., 1990), while the related $^{5'CGAG3'}_{3'GAGC5'}$ motif is common in nature. This natural

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

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′	−17.3	12.6	1.3	−3.6	10.9	57.6	49.8
G2−C7′	−5.9	13.2	6.9	−3.6	10.9	55.9	52.6
C3−G6′	−11.8	12.6	2.7	−3.7	11.1	49.3	51.2
A4−G5′	−6.6	10.7	−6.7	−3.3	13.0	48.0	46.8
G5−A4′	−6.8	10.8	7.4	−3.3	13.0	47.4	47.1
G6−C3′	−12.2	12.9	−2.4	−3.7	11.1	51.2	48.9
C7−G2′	−5.7	13.4	−6.6	−3.6	10.9	52.5	55.8
C8−G1′	−17.7	12.6	−1.8	−3.6	10.9	49.5	57.5
average	−10.5	12.3	0.1	−3.6	11.5	51.4	51.2
A-form	−13.8	16.9	0.2	−4.4	10.9	52.2	52.2

[a] The parameters were calculated using NEWHEL92 from the 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 the left and right base, respectively, in column 1.

selection is probably due to the hydrogen-bonding pattern within the GA mismatches. The sheared GA mismatches in the $^{5'CGAG3'}_{3'GAGC5'}$ motif (Figure 1) have hydrogen-bonding positions available for formation of tertiary contacts. For example, a base triple has been observed to a sheared GA pair in the crystal structure of *Thermus thermophilus* serine tRNA (Biou et al., 1994). In the crystal structure of the hammerhead ribozyme, a tandem sheared GA mismatch is a binding site for a metal ion (Pley et al., 1994a). Such tertiary contacts are not possible with the imino hydrogen-bonding pattern in the three-dimensional structure found for the $^{5'CAGG3'}_{3'GGAC5'}$ motif studied here. Another tandem GA motif with imino hydrogen bonding (Wu & Turner, 1996), $^{5'GGAC3'}_{3'CAGG5'}$, also is rare in known secondary structures. An understanding of why some sequence contexts promote a biologically active tandem GA structure whereas others do not would be a step toward modeling non-Watson−Crick regions in RNA.

On the basis of NMR studies of the $^{5'CGAG3'}_{3'GAGC5'}$ and $^{5'GGAC3'}_{3'CAGG5'}$ motifs, it has been suggested that the opposite orientations of the CG base pairs induce different hydrogen-bonding patterns in the mismatches because this minimizes unfavorable interactions between the partial charges on the GA mismatches and CG base pairs (Wu & Turner, 1996). For the $^{5'CAGG3'}_{3'GGAC5'}$ motif determined here, no strong unfavorable electrostatic interactions are present (see Supporting Information). Thus there is no major electrostatic interaction disfavoring the imino structure.

The two potential sheared conformations for $(rGGCAGGCC)_2$ are compared with the experimentally observed imino hydrogen-bonded conformation in Figure 15. Maintaining the anti glycosyl angle for both G and A, there are two ways to convert imino hydrogen-bonded GA mismatches into side-by-side conformations (see left and right panels in Figure 15). The reverse-sheared structure shown on the left in Figure 15 is generated by interchanging the A and G of the mismatches in $(rGGCGAGCC)_2$ (SantaLucia & Turner 1993), keeping the rest of the coordinates. In the reverse-sheared structure, all the helical properties are basically the same as those of $(rGGCGAGCC)_2$. Helical underwinding (18°) at the $^{5'CA3'}_{3'GG5'}$ step facilitates intrastrand stacking. Helical overwinding (81°) at the $^{5'AG3'}_{3'GA5'}$ step facilitates interstrand stacking. There are no hydrogen bonds, however, between the G and A of the mismatch. Presumably this makes this potential structure unstable. The sheared structure of $(rGGCAGGCC)_2$ on the right in Figure 15 was created by forcing the mismatched G and A into the sheared structure observed in $(rGGCGAGCC)_2$, while keeping the six stem base pairs in standard A-form geometry by using distance and dihedral angle restraints observed for $(rGGCAGGCC)_2$, and then energy minimizing the structure. This geometry of sheared GA mismatches results in overwinding by 18° (=51° − 33°) at the $^{5'CA3'}_{3'GG5'}$ step and underwinding by 29° (=33° − 4°) at the $^{5'AG3'}_{3'GA5'}$ step in the sheared structure of $(rGGCAGGCC)_2$ (Figure 15). This sheared structure has two hydrogen bonds but requires a major distorsion of the backbone and has relatively little stacking overlap between bases. Formation of sheared GA mismatches in the $^{5'CGAG3'}_{3'GAGC5'}$ motif also requires significant backbone distortion (SantaLucia & Turner, 1993). This distortion, however, involves overwinding of the $^{5'GA3'}_{3'AG5'}$ step by 48° (=2 × 24°), allowing formation of a hydrogen bond between the exocyclic NH_2 group of G and a nonbridging oxygen of the phosphate 5′ of the A in the opposite strand. In addition, the helical geometry of CGAG allows the formation of a hydrogen bond from the adenine amino to the G2′ oxygen. These hydrogen bonds presumably provide stabilization energy that compensates for the helical distortion. Equivalent hydrogen bonds cannot be made with the sheared $^{5'CAGG3'}_{3'GGAC5'}$ because, in the context of a right-handed helix, the overwinding by only 18° (=51° − 33°) at the $^{5'CA3'}_{3'GG5'}$ step is not enough to bring the phosphate 5′ of the A close to the exocyclic NH_2 group of the G in the AG mismatch. The overwinding by 18° at the $^{5'CA3'}_{3'GG5'}$ step and the underwinding by 29° at the $^{5'AG3'}_{3'GA5'}$ step also result in poor stacking at these two steps. Thus this potential sheared conformation of $^{5'CAGG3'}_{3'GGAC5'}$ is also very unfavorable. A modeling study of GA mismatches concluded that a sheared structure cannot accommodate a Watson−Crick pair 5′ of A and still maintain A-form geometry (Gautheret et al., 1994). This is consistent with the structures and models presented here. Preliminary NMR experiments on $(CGCAGGCG)_2$ are also consistent with this idea, since the GA mismatches are imino hydrogen bonded (SantaLucia, 1991).

A reasonable working hypothesis is that the structure of a GA mismatch depends on at least three factors: (1) electrostatic interactions between partial charges, (2) the number of hydrogen bonds, including hydrogen bonds to the backbone, and (3) backbone distortion. There have recently been significant advances in the computational modeling of nucleic acids (Ponnuswamy & Gromiha, 1994; Jiang et al., 1994; York et al., 1995; Zichi, 1995; Cheatham et al., 1995; Searle & Williams, 1993). The structures and thermodynamics of $(GGCAGGCC)_2$, $(GGCGAGCC)_2$, and $(GCGGACGC)_2$ provide good tests of these modeling parameters and protocols.

SUPPORTING INFORMATION AVAILABLE

One table of T_1 relaxation times, one figure showing the 1-D 500 MHz proton spectrum and two figures showing 400 ms NOESY spectra in D_2O at 35 °C, three figures comparing experimental and back-calculated 400 ms NOESY spectra, one figure showing a stereoview of 15 final overlapped structures of GA mismatches, and one figure showing overlap of electrostatic potentials of AG mismatch with the adjacent CG pair (9 pages). Ordering information is given on any current masthead page.

REFERENCES

Agrawal, S. (1995) *Methods in Molecular Medicine: Antisense Therapeutics*, Humana Press, Totowa, NJ.

Allain, F. H.-T., & Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.

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.

Cheatham, T. E., III, Miller, J. L., Fox, T., Darden, T. A., & Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 4193−4194.

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.

Gautheret, D., Konings, D., & Gutell, R. R. (1994) *J. Mol. Biol. 242*, 1−8.

Gorenstein, D. (1984) in *^{31}P NMR, Principles and Applications*, Academic Press, New York.
Gorenstein, D. G., Meadows, R. P., Metz, J. T., Nikonowicz, E., & Post, C. B. (1990) in *Advances in Biophysical Chemistry* (Bush, C. A., Ed.) JAI Press, Greenwich, CT.
Greenbaum, N. L., Radhakrishnan, I., Hirsh, D., & Patel, D. J. (1995) *J. Mol. Biol. 252*, 314−327.
Gubser, C. C., & Varani, G. (1996) *Biochemistry 35*, 2253−2267.
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. T. (1988) *Magn. Reson. Chem. 26*, 927−944.
Hunter, C. A. (1993) *J. Mol. Biol. 230*, 1025−1054.
Jiang, S. P., Raghunathan, G., Ting, K. L., Xuan, J. C., & Jernigan, R. L. (1994) *J. Biomol. Struct. Dyn. 12*, 367−382.
Kellogg, G. W., Szewczak, A. A., & Moore, P. B. (1992) *J. Am. Chem. Soc. 114*, 2727−2728.
Kim, S. H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C., & Rich, A. (1974) *Science 185*, 435−440.
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., & Agrawal, S. (1995) *Biochemistry 34*, 10056−10062.
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.
Mol, J. N. M., & van der Krol, A. R. (1991) *Antisense Nucleic Acids and Proteins, Fundamentals and Applications*, Marcel Dekker, Inc., New York.
Moras, D., Comarmond, M. B., Fischer, J., Weiss, R., Thierry, J. C., Ebel, J. P., & Giege, R. (1980) *Nature 288*, 669−674.
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.
Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 264−268.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994a) *Nature 372*, 68−74.
Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994b) *Nature 372*, 111−113.
Ponnuswamy, P. K., & Gromiha, M. M. (1994) *J. Theor. Biol. 169*, 419−432.
Puglisi, J. D., Chen, L., Blanchard, S., & Frankel, A. D. (1995) *Science 270*, 1200−1203.
Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. D., Clark, B. F. C., & Klug, A. (1974) *Nature 250*, 546−551.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
SantaLucia, J., Jr. (1991) Ph.D. Thesis, University of Rochester, Rochester, NY.
SantaLucia, J., Jr., & Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813−8819.
Scott, W. G., Finch, J. T., & Klug, A. (1995) *Cell 81*, 991−1002.
Searle, M. S., & Williams, D. H. (1993) *Nucleic Acids Res. 21*, 2051−2056.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94−98.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Szewczak, A. A., & Moore, P. B. (1995) *J. Mol. Biol. 247*, 81−98.
Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
Varani, G., & Tinoco, I., Jr. (1991) *Q. Rev. Biophys. 24*, 479−532.
Walter, A. E., Wu, M., & Turner, D. H. (1994) *Biochemistry 33*, 11349−11354.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230−252.
Wijmenga, S. S., Mooren, M. M. W., & Hilbers, C. W. (1993) NMR of Nucleic Acids; from spectrum to structure, in *NMR of Macromolecules, A Practical Approach*, Chapter 8, Oxford University Press, Oxford OX2 6DP.
Woo, N. H., Roe, B. A., & Rich, A. (1980) *Nature 286*, 346−351.
Wu, M., & Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.
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.
Ye, X. M., Kumar, R. A., & Patel, D. J. (1995) *Chem. Biol. 2*, 827−840.
York, D. M., Yang, W., Lee, H., Darden, D., & Pedersen, L. G. (1995) *J. Am. Chem. Soc. 117*, 5001−5002.
Zichi, D. A. (1995) *J. Am. Chem. Soc. 117*, 2957−2969.

BI9625915