Structure of (rGGCGAGCC)$_2$ in Solution from NMR and Restrained Molecular Dynamics†

John SantaLucia, Jr.,*,‡,§ and Douglas H. Turner§

Department of Chemistry, University of California, Berkeley, California 94720, and Department of Chemistry, University of Rochester, Rochester, New York 14627

Received June 28, 1993; Revised Manuscript Received September 9, 1993®

ABSTRACT: The duplex (rGGCGAGCC)$_2$ contains tandem G·A mismatches—a common motif in the secondary structures of biological RNAs. The three-dimensional structure of (rGGCGAGCC)$_2$ was derived using molecular dynamics and energy minimization with NMR-derived restraints for 78 interproton distances (per strand), 18 hydrogen bonds for the six Watson–Crick G·C pairs, and 26 dihedral angles (per strand). The G·A mismatch structures are similar to those observed in a DNA duplex [Li, Y., Zon, G., & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 80,* 26–30] and an RNA hairpin [Heus, H. A., & Pardi, A. (1991) *Science 253,* 191–193], with hydrogen bonds from guanine 2-amino and N3 to adenine N7 and 6-amino, respectively. The other G 2-amino and A 6-amino protons are within hydrogen-bonding distance of a phosphate oxygen and 2′-oxygen, respectively. Strong interstrand A–A and G–G stacking is observed between the G·A mismatches. This contrasts with the poor stacking observed between the G·A mismatches and closing G·C base pairs. The stems are basically A-form with all bases in the anti conformation and all nonterminal sugars in the C3′-endo conformation. The structure rationalizes previous thermodynamic, circular dichroism, and imino proton NMR results and suggests tandem G·A mismatches in RNA may provide a contact site for tertiary interactions.

G·A mismatches are commonly observed in the secondary structures of biological RNA molecules (Traub & Sussman, 1982; Noller, 1984). Thermodynamic studies indicate tandem G·A mismatches contribute more than other non-G·U mismatches to the free energies of RNA folding (SantaLucia et al., 1991b). Tandem G·A mismatches in the order 5′G–A3′ appear to be more prevalent than tandem G·A mismatches in the order 5′A–G3′. For example, a search of phylogenetically determined secondary structures of 21 16S rRNAs (Gutell et al., 1985) and 38 23S rRNAs (Gutell & Fox, 1988) revealed that the 5′G–A3′ sequence occurred 45 times while the 5′A–G3′ sequence never occurred (SantaLucia et al., 1990). Furthermore, the conserved sequences of the hammerhead ribozyme include a 5′G–A3′ sequence double G·A mismatch capping a helix (Forster & Symons, 1987a,b). Thus, we decided to determine structural features for the duplex (rGGCGAGCC)$_2$ (underlined residues are mismatched). In this paper, a model is derived for the structure of (rGGCGAGCC)$_2$ based on interproton distances and dihedral angles determined by NMR.

MATERIALS AND METHODS

RNA Synthesis and Purification. rGGCGAGCC was synthesized on solid support using nucleoside methoxyphosphoramidites with the 2′-hydroxyl protected as the *tert*-butyl dimethyl silyl ether (Usman et al., 1987). Base amino groups were protected with phenoxyacetyl (for A and G) and acetyl (for C) blocking groups. Upon completion of the coupling reactions, the phosphate methyl groups were removed by treatment with thiophenol–dioxane–triethylamine (1:3:1 by volume). The oligomer was removed from solid support, and base blocking groups were also removed by treatment with concentrated ammonia in ethanol (3:1 v/v) at 50 °C overnight. The silyl protection was removed by treatment with freshly made 1.0 M triethylammonium hydrogen fluoride (50 equiv) in pyridine at 50 °C for 48 h. The crude reaction mixture was then dried and partitioned between water and diethyl ether. The water phase was evaporated to dryness. Dimethoxytrityl was removed by treating with 80% acetic acid for 45 min. The sample was then evaporated to dryness, dissolved in 10 mL of 10 mM ammonium bicarbonate (pH 5.5), and passed over a Sep-pak C-18 cartridge (Waters) to desalt it. The RNA was eluted with 30% acetonitrile buffered with 10 mM ammonium bicarbonate at pH 7.0. Product was purified on a Si500F thin-layer chromatography plate (Baker) eluted for 5 h with 1-propanol–ammonia–water (55:35:10 by volume) (Chou et al., 1989). Bands were visualized with an ultraviolet lamp, and the least mobile band was cut out and eluted three times with 1 mL of doubly distilled water. We find that TLC gives superior separation compared to HPLC for small G-rich oligonucleotides. Presumably, this is due to the strongly denaturing conditions used for TLC. Oligonucleotides were desalted and further purified with a Sep-pak C-18 cartridge. The purity of rGGCGAGCC was checked by analytical C-8 HPLC (Beckman) and was greater than 95%.

Sample Preparation for NMR. The RNA was dissolved in doubly distilled H_2O and extensively dialyzed by continuous flow (BRL) against (A) 10 mM phosphate and 0.5 mM EDTA, pH 7, (B) doubly distilled H_2O, and (C) 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM EDTA, pH 6.4. The sample was then lyophilized to dryness and lyophilized twice from 99.8% D_2O and once from 99.96% D_2O. The sample was dissolved, under dry nitrogen atmosphere, in 0.65 mL of 99.996% D_2O (Cambridge Isotope Laboratories). The strand concentration of rGGCGAGCC for this sample was 3 mM.

† This work supported by National Institutes of Health Grant GM22939 to D.H.T. and NIH Postdoctoral Fellowship GM14682 to J.S.L., by NIH Grant RR 03317 and NSF Grant DBM 8611927 for the purchase of the Varian NMR spectrometer at Rochester, and by DOE Grant DE FG05-86ER 75281 and NSF Grants DMB 86-09305 and BBS 87-20134 for the purchase of the Bruker NMR spectrometer at Berkeley.

* Address correspondence to this author at Berkeley.
‡ University of California.
§ University of Rochester.

® Abstract published in *Advance ACS Abstracts,* November 1, 1993.

0006-2960/93/0432-12612$04.00/0

NMR spectra of the nonexchangeable protons of 3 mM rGGCGAGCC were recorded at 35 or 40 °C.

NMR spectra of the exchangeable protons of 1 mM rGGCGAGCC were recorded at 10–25 °C in 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM EDTA, pH 6.4, in 90% H_2O/10% D_2O, unless otherwise noted. A higher RNA concentration was not used because aggregation occurs at the low temperatures required for exchangeable proton studies.

NMR Spectroscopy. Nuclear Overhauser enhancement spectroscopy (NOESY) and 1-D NOE difference spectra were acquired on a Varian VXR-500S spectrometer and processed by a Sun 3/160 computer running Varian VNMR software. ^{31}P-Decoupled double quantum filtered correlated spectroscopy (DQF-COSY) and ^{1}H{^{13}C} heteronuclear multiple quantum correlation spectra (HMQC) were acquired on a Bruker AMX-600 spectrometer and processed with a microVax computer by running the programs FTNMR and FELIX (Hare Research, Inc.). The ^{1}H{^{31}P} heteronuclear correlation spectrum (HETCOR) was acquired on a Bruker AMX-400 spectrometer. TSP [3-(trimethylsilyl)propionate] was added to all samples as an internal chemical shift reference for proton and carbon resonances. Phosphorus resonances were referenced to internal phosphate buffer, pH = 6.4, at 0.0 ppm (see Table I footnote).

Exchangeable proton spectra were recorded using the solvent suppression $1\bar{3}1$ pulse sequence (Hore, 1983). Spectra were collected with 16K complex points over a spectral width of 10 kHz and multiplied by a 4.0-Hz line-broadening exponential function. The delays were set to optimize the signal to noise ratio at 12.5 ppm. One-dimensional NOE difference spectra of the imino and amino protons were acquired by using the $1\bar{3}1$ pulse sequence as the read pulse. Experiments with irradiation on- and off-resonance were collected in blocks of 16 scans and interleaved to correct for long-term instrumental drift. A presaturation period of 0.5–1.0 s was used to achieve a steady state. In cases where resonances partially overlapped, the selective saturation was achieved by applying a low-power radio-frequency (rf) field 5 Hz off-resonance (away from the overlapped resonance). To minimize spillover artifacts, the control off-resonance saturation frequency was offset by the difference of the on-resonance saturation frequency and the resonance frequency for which spillover was a concern (Varani et al., 1989).

All 2D NMR spectra were recorded in the phase-sensitive mode using the States–Haberkorn method (States et al., 1982) for NOESY spectra and TPPI for all other experiments (Marion & Wuthrich, 1983). The residual HDO peak was reduced by preirradiation at low decoupler power.

NOESY spectra were recorded at 60-, 100-, and 400-ms mixing times as described by Macura and Ernst (1979). A total of 450–550 FID's were accumulated with 4K complex points and 5000-Hz spectral width. A total of 32–128 scans were averaged for each FID. Data were zero-filled to 1K real points in the t_1 dimension and apodized with Gaussian or phase-shifted sine-bell functions.

For ^{31}P-decoupled DQF-COSY, 650 FID's were accumulated with 4K complex points and with the carrier at 4.6 ppm, 1860-Hz (3.1 ppm) spectral width (digital resolution <1 Hz/point in t_2) in both dimensions, and 64 scans per transient. Data were zero-filled to 2K real points in the t_1 dimension and apodized with 0–30° phase-shifted sine-bell functions. In some instances a line-broadening exponential function was also used to skew the sine bell and improve the signal to noise ratio. The diagonal and cross-peaks of DQF-COSY spectra were phased with antiphase-absorption line shape in both dimensions.

The ^{1}H–^{13}C HMQC spectrum was acquired with the pulse sequence previously described (Varani & Tinoco, 1991a). The excitation delay, τ ($=^1/_2$ J), was set to 3 ms to give optimum sensitivity for the sugar 1-bond ^{1}H–^{13}C correlations. The same delay period was used prior to acquisition and GARP1 decoupling so that doublet components would be in phase. The spectral widths were 5000 Hz (8.3 ppm) for proton and 6250 Hz (41.4 ppm) for carbon dimensions. A total of 155 FID's of 2K complex points were collected. A repetition delay of 3 s was used with 288 scans averaged for each FID for a total acquisition time of 43 h. Data were zero-filled to 1K real points in t_1 and apodized with 60° phase-shifted sine bells in both dimensions.

The ^{1}H–^{31}P HETCOR was acquired with the pulse sequence described by Sklenar et al. (1986). A total of 132 FID's were acquired with 2K complex points with 128 scans averaged for each transient. Spectral widths were 800 Hz (2 ppm) for proton and 1000 Hz (6.2 ppm) for phosphorus dimensions. Spectra were processed with Gaussian or phase-shifted sine-bell functions.

The T_1 relaxation times were measured by inversion recovery and are 2.9–4 s for base protons and 5.6 s for the AH2 proton (SantaLucia, 1991). These results agree with previous observations of long relaxation times in RNA (Petersheim & Turner, 1983; Wang et al., 1992). Thus, the repetition delay was set to 5 s in correlated and long mixing time NOESY experiments to optimize the signal to noise ratio (Ernst et al., 1987). For short mixing time NOESY spectra, spin-state equilibrium is required for accurate calculation of distances from NOE measurements (Macura & Ernst, 1979). Thus, in the 100-ms mixing time NOESY spectrum, a repetition delay of 25 s was used. Subsequent experiments collected with 6-s recycle delay gave essentially the same results as NOESY data collected with 25-s recycle delay. Presumably, this is due to relatively uniform relaxation of all spins, since for RNA the base and adenine H2 resonances relax at similar rates. For molecules such as DNA, where adenine H2 relaxation rates are factors of 2–5 longer than other protons, more caution is required.

The NOESY-JRE (NOESY with jump–return–echo solvent suppression sequence for the read pulse) (Sklenar & Bax, 1987; Sklenar & Feigon, 1990) spectrum of rGGCGAGCC dissolved in 90% H_2O was recorded with 130-ms mixing time. Spectra were acquired with 8K complex points and spectral widths of 10 kHz, and the delays were set to optimize the signal to noise ratio at 8 ppm for optimum detection of cross-peaks involving amino resonances.

NOESY cross-peak volumes were measured by encircling the peak of interest and using the Varian VNMR software to integrate. All internuclear vectors were assumed to vary with the same rotational correlation time. Proton–proton distances were determined by measuring cross-peak volumes of 60- and 100-ms mixing time NOESY spectra and comparing to the volume for cytidine H5–H6 (2.45 Å) cross-peaks and scaling by $1/r^6$. Error bars were assigned to each distance as follows: (1.7–2.6) ± 0.5 Å, (2.7–3.4) ± 0.6 Å, (3.5–4.0) ± 0.8 Å, and >4.0 ± 1.0 Å. Distances between protons with clearly nonexistent NOEs in the 400-ms mixing time NOESY spectrum were restrained to >4.5 Å. Occasionally error ranges were adjusted to larger values if overlap, noise ridges, or spin diffusion was thought to be large. These conservative distance estimates avoid overinterpretation of the data and reduce the potential for inconsistencies among constraints (Metzler et al., 1990).

The A5H2 to A5′H1′ (cross-strand) distance is particularly accurate for the following reasons: (1) a recycle delay time of 25 s was used in the 100-ms mixing time NOESY to ensure spin equilibrium, (2) the two-spin approximation for this distance is particularly good since no other cross-peaks involving the A5H2 are observed, (3) the measured cross-peak volumes for NOESY experiments acquired with both 60- and 100-ms mixing times are consistent with a distance of 3.2 ± 0.5 Å, and (4) the cross-peak is well resolved at 40 °C.

Scalar Coupling Measurements. ^{1}H–^{1}H *J*-couplings were measured in the high-resolution ^{31}P-decoupled DQF-COSY as described by Varani and Tinoco (1991a). Line widths of 3–4 Hz for the sugar protons allowed *J*-couplings to be measured with an uncertainty of 1–2 Hz. Measurement of H2′–H3′ couplings, which are 5 ± 1 Hz for both C2′-endo and C3′-endo sugar puckers (Hosur et al., 1988), served as useful controls that distortions due to line-width effects and superposition of multiplet components are small compared to the reported uncertainties. When the *J*-coupling is smaller than the line width, cancellation of antiphase components causes considerable attenuation of cross-peak intensity. Thus, for attenuated cross-peaks the *J*-coupling was estimated from the cross-peak intensity. When cross-peaks were very weak or missing, the *J*-coupling was estimated to be <3 Hz. Heteronuclear ^{1}H–^{31}P *J*-couplings were determined from the high-resolution ^{1}H–^{31}P HETCOR spectrum. The heteronuclear couplings were obtained by subtracting the passive ^{1}H–^{1}H couplings, measured as described above, from the total multiplet width. With this method the ^{1}H–^{31}P *J*-coupling uncertainties are about ±2 Hz.

Structure Modeling. Models of (rGGCGAGCC)$_2$ consistent with NMR data were derived from restrained molecular dynamics (rMD) and restrained energy minimization. Calculations were done and structures visualized on a Silicon Graphics SGI 4D/35TG computer running X-PLOR (Brunger, 1990) and Biosym InsightII software. The all-atom X-PLOR force field and charges were used for all molecular dynamics simulations and energy minimizations. Parameters for the sugar and base bond angles, however, were modified to be similar to those given for RNA by Saenger (1984). This resulted in sugar pucker pseudorotations and amplitudes closer to those observed in canonical A-form duplexes. Calculations were performed *in vacuo* (i.e., excluding solvent and counterions) including potential terms for bond stretching, bond angle bending, maintaining improper orientations, torsional rotation, Coulombic electrostatic interactions, and van der Waals interactions (either repulsive or Lennard-Jones). Bond length and bond angle force constants were set to 1000 kcal/(mol Å^2) and 500 kcal/(mol rad^2), respectively. Nonbonded interactions were cut off at 8.5 Å. NOE and dihedral angle restraints were added to the force field as pseudopotentials that are flat (zero penalty) within the error limits of the restraints and quadratic outside the error bounds. NOE and dihedral angle force constants were set to 50 kcal/mol Å^2 and 50 kcal/(mol rad^2), respectively. A total of 78 interproton distances (22 intranucleotide and 56 internucleotide distances per strand), 18 hydrogen bonds (included as NOE restraints of 1.8 ± 0.2 Å) for six Watson–Crick base pairs, and 26 dihedral angles (per strand) were included as constraints. Note that 52 dihedral angles specify the structure of an octamer strand. No hydrogen-bonding restraints between G·A mismatches were used since no direct NMR evidence for a particular geometry was observed. This also prevents bias in the conformational searches. Dihedral angle restraints for α and ζ were not included, since phosphorus chemical shifts might be misleading. The NOE and dihedral angle restraints bias molecular dynamics and energy minimization calculations toward the correct structure. Since rGGCGAGCC forms a self-complementary duplex, the same restraints were applied to both strands, and 2-fold symmetry was applied as a restraint during the final stage of refinement by including NCS (noncrystallographic symmetry) as a term in the X-PLOR force field.

Eleven starting structures with random α, β, γ, ϵ, and ζ dihedral angles were generated with the internal coordinate facility of version 2.1 of X-PLOR. A 12th starting structure was generated by constructing an A-form duplex in the Biopolymer module of InsightII. Structure calculations were done in two stages: (1) global fold and (2) refinement with a protocol similar to that described by Wimberly et al. (1992). The global-fold protocol was omitted for the A-form starting structure.

During the global fold calculation, electrostatic interactions are turned off, and van der Waals repulsive interactions are turned on with the following protocol: (1) 10-ps rMD at a simulated temperature of 1000 K with no van der Waals interactions or dihedral angle restraints to allow sufficient sampling of conformation space so that NOE and hydrogen bond restraints can be satisfied; (2) 10-ps rMD at 1000 K while the force constant for van der Waals repulsion is gradually increased; (3) 14-ps rMD while the temperature of the system is gradually reduced to 300 K; (4) 1000 steps of restrained energy minimization. The criterion for convergence of the global-fold algorithm is the proper formation of Watson–Crick base pairs in the stem regions. We find that approximately 70% of the structures converge. Those structures which did not converge were resubmitted to the global-fold algorithm.

During the refinement calculations, dihedral angle restraints, Coulombic electrostatic forces, and Lennard-Jones van der Waals forces are turned on and the following protocol is performed: (1) 500 steps of restrained energy minimization; (2) 6-ps rMD at 1000 K while the dihedral angle force constant is gradually increased to 50 kcal/(mol rad^2) for angles derived from NMR data; (3) 7-ps rMD while being cooled from 1000 to 300 K; (4) 3-ps rMD at 300 K (this allowed for small energy barriers to be overcome so that the molecule folds into deep energy wells); (5) 1000 steps of restrained energy minimization; (6) turning on the NCS pseudopotential [force constant = 300 kcal/(mol A^2)] and performing 2-ps rMD at 300 K and 3000 steps of restrained energy minimization.

RESULTS

Assignment of Nonexchangeable Proton, Carbon, and Phosphorus Resonances. NMR resonances of (rGGCGAGCC)$_2$ were assigned essentially as described by Varani and Tinoco (1991a). Spectral overlap required assignments based on spectra recorded at 35 and 40 °C. Table I summarizes all chemical shift assignments at 35 °C (small variations in chemical shifts are observed upon changing the temperature to 40 °C). The base (H8/H6/H2) to H1′/H5 region of the NOESY spectrum recorded at 400-ms mixing time at 35 °C is shown in Figure 1. The assignments follow the standard connectivity pathway from residues G1 to G6, where the connections stop. Weak NOE's from G6H8 to C7H5 and C7H6 (data not shown) allow the connectivity pathway to be completed. Note that the G6H1′ resonance has not been located. One possible explanation for this is that the G6H8–H1′ and the C7H6–G6H1′ cross-peaks are very weak. Examination of the 5.0–6.0 ppm region of the 1-D NMR

Table I: Chemical Shifts for rGGCGACGG at 35 °C in ppm[a]

base	proton assignments								
	H8/H6	H5/H2	H1′	H2′	H3′	H4′	H5′[b]	H5″[b]	imino/amino[c]
G1	7.98	na[d]	5.77	4.88	4.54	4.33	3.99	3.88	12.5
G2	7.54	na	5.91	4.60	4.49	4.18	4.43	4.32	13.3
C3	7.33	5.13	5.48	4.57	4.31	4.18	4.48	4.42	8.25/6.65
G4	7.98	na	5.78	4.39	4.87	4.54	4.47	4.21	10.1
A5	7.88	7.92	5.92	4.92	4.42	4.60	4.55	4.29	(7.60/6.85)
G6	7.44	na	(4.32)	4.23	4.31	(4.35)	4.32	4.11	13.2
C7	7.59	5.23	5.52	4.26	4.40	4.35	4.47	3.99	8.51/6.88
C8	7.72	5.53	5.79	4.05	4.15	4.55	4.47	4.03	8.35/6.95

base	carbon assignments							phosphorus assignments
	C8/C6	C5/C2	C1′	C2′	C3′	C4′	C5′	
G1	138.2	na	90.3	74.1	(73.2)	81.0	61.8	na
G2	135.2	na	92.3	74.5	(71.0)	82.5	64.2	−1.67
C3	139.7	96.5	93.4	(71.8)	71.8	82.5	64.5	−1.62
G4	137.2	na	91.9	76.2	74.1	81.2	64.2	−1.92
A5	139.7	153.2	91.9	74.7	72.6	81.5	65.9	−1.46
G6	134.2	na	(81.0)	73.9	71.8	(81.3)	64.8	−1.57
C7	140.7	95.9	93.2	74.6	71.0	83.7	63.3	−2.36
C8	141.2	97.2	91.9	76.6	68.7	82.4	64.1	−1.98

[a] Chemical shifts are relative to TSP for proton and carbon resonances and to phosphate buffer (pH = 6.4) for phosphorus resonances. Phosphorus chemical shifts can be converted to the TMP scale by subtracting 2.06 ppm from the phosphorus chemical shifts given here. Assignments in parentheses are tentative. [b] The H5′ and H5″ assignments are not stereospecific. [c] Exchangeable proton chemical shifts at 22 °C. [d] Not applicable.

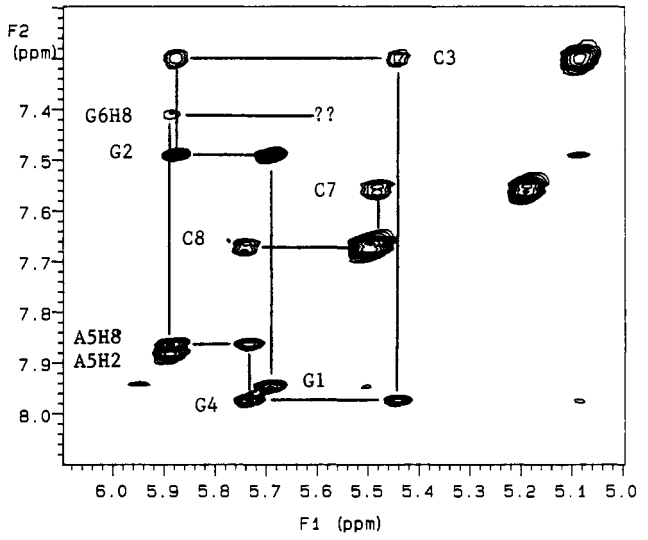

FIGURE 1: Expansion of the 400-ms mixing-time NOESY spectrum showing the base H8/H6/H2 to sugar H1′ region of rGGCGAGCC at 35 °C in 80 mM NaCl, 10 mM phosphate, and 0.5 mM EDTA, pH 6.4, dissolved in 99.996% D_2O. The sequential connectivity pathway is shown. Intraresidue base H8/H6 to sugar H1′ cross-peaks are labeled. Note that there is a break in the connectivity pathway at G6.

spectrum (supplementary material; see paragraph at end of paper regarding supplementary material), however, shows only 10 resonances, where 11 are expected (one H1′ for each sugar and three H5 resonances from C residues). Thus the G6H1′ resonance must be shifted to the crowded region between 4 and 4.6 ppm, making it difficult to locate. We attempted to assign the G6H1′ resonance from the ^{13}C decoupled ^1H–^{13}C HMQC spectrum (supplementary material), and an extra H–C correlation is observed in the C4′ region (80–85 ppm). Except for G6, all of the H4′–C4′ cross-peaks have been assigned. Thus, the two remaining peaks are tentatively assigned to G6H1′ and G6H4′. Assuming the G6C1′ assignment is correct, the observed chemical shift of 81.0 ppm is remarkable. In contrast, Varani and Tinoco (1991a) observed that an H1′ proton in a similar environment was shifted to 4.38 ppm, but the carbon resonated at 92.2 ppm.

The remaining base resonance at 7.88 ppm is assigned to A5H2. This assignment was confirmed by the ^1H–^{13}C HMQC spectrum (data not shown). This resonance is well resolved from A5H8 at 40 °C, and a medium-intensity NOE to the A5H1′ is observed in both the 60- (data not shown) and 100-ms mixing time NOESY (supplementary material). Since the AH2 proton cannot be closer than 4.5 Å to its own H1′ proton (Wuthrich, 1986), this must result from an interstrand transfer (recall the *C2* symmetry of a self-complementary duplex). This is important because it suggests that the minor groove is narrower than typically observed for A-form duplexes. This contrasts with the standard G·A mismatch conformation observed in $(rCGCAGGCG)_2$ where the minor groove is wider than typical A-form, in order to relieve steric clashes between purine residues (SantaLucia, 1991). NOE's at short and long mixing times from the A5H2 to H2′, H3′, H4′, H5′, and H5″ sugar resonances are very weak or nonexistent (Figure 2 and supplementary material), suggesting that the A5H2 is relatively isolated and that the two-spin approximation is particularly applicable. The spatial isolation of the A5H2 is confirmed by its long T_1 relaxation time of 5.6 s (SantaLucia, 1991).

Assignments of the H2′ resonances follow from strong cross-peaks to H1′ resonances observed in the short mixing time NOESY spectrum and from weak cross-peaks observed in DQF-COSY and 2-quantum spectra (SantaLucia, 1991). Figure 3 shows the H1′ to sugar region of the 400-ms mixing time NOESY spectrum. These assignments are confirmed in the base to sugar region of the 100-ms mixing time NOESY (Figure 2) by strong H8/H6(n) to H2′($n-1$) cross-peaks, as typically observed for the A-form conformation. The only remaining H2′ resonance not assigned is G6H2′, which is discussed further below.

Assignments of the H3′ and H4′ resonances were determined primarily from *J*-couplings (Table II) observed between H2′–H3′ and H3′–H4′ in DQF-COSY (data not shown) and from H3′–P5′ correlations in the ^1H–^{31}P HETCOR (supplementary material). The H1′ to sugar and base to sugar regions of 100- and 400-ms mixing time NOESY spectra were used to confirm assignments. In the H1′ to sugar region of the 400-ms mixing time NOESY (Figure 3), cross-peaks from the H1′ to the

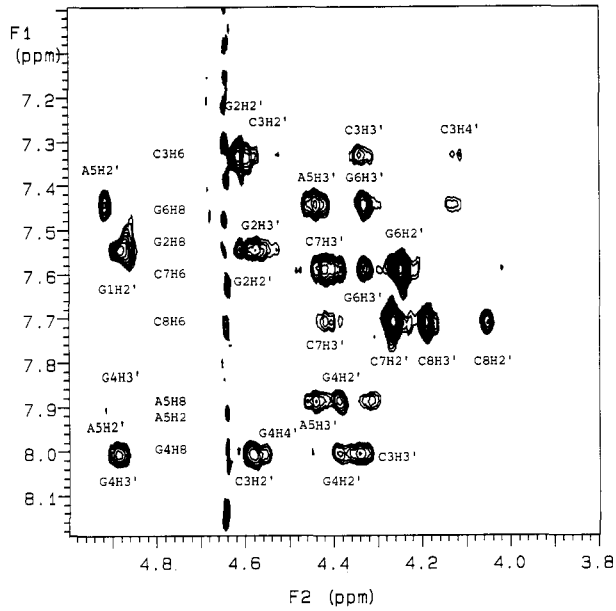

FIGURE 2: Expansion of the 100-ms mixing-time NOESY spectrum of rGGCGAGCC at 40 °C, showing the base H8/H6/H2 to sugar H2′/H3′/H4′/H5′/H5″ region. Important cross-peaks are labeled. This region contains most of the structural information used for restraints in molecular dynamics calculations. Note that no cross-peaks involving A5H2 are observed.

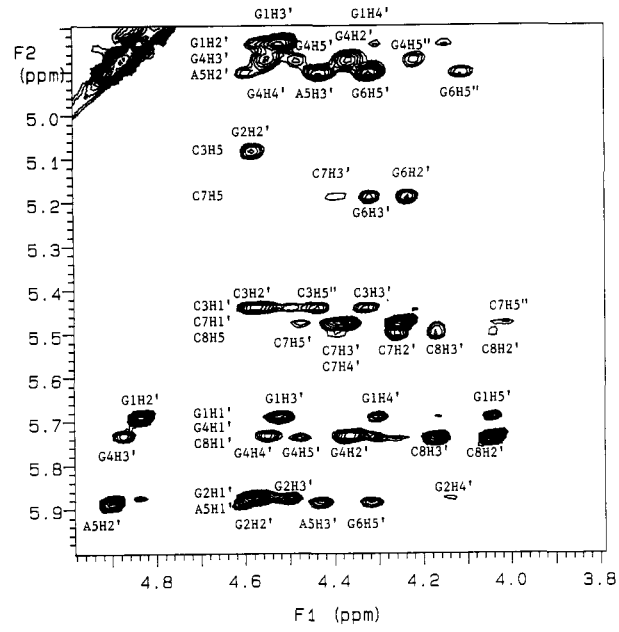

FIGURE 3: Expansion of the 400-ms mixing time NOESY spectrum of rGGCGAGCC at 35 °C, showing the H1′ to sugar H2′, H3′, H4′, H5′, H5″ region. Important resonances are labeled. Intraresidue cross-peaks from the H1′ to H2′, H3′, H4′, and sometimes H5′ and H5″ are observed via spin diffusion. This region is useful for deducing which cross-peaks in the base H8/H6 to sugar H2′/H3′/H4′/H5′/H5″ region are inter- vs intraresidue cross-peaks.

H3′, H4′ and sometimes also H5′ and H5″ resonances are observed from spin diffusion via H2′. This region is important for deducing which base to sugar cross-peaks arise from intra- or interresidue contacts. Thus the H1′ to sugar region of the 400-ms mixing time NOESY gives information on RNA assignments analogous to that derived from the TOCSY experiment for DNA. For RNA the NOESY experiment is more sensitive than the TOCSY experiment because coherence transfer from H1′ to H2′ is inefficient for C3′-endo sugar pucker.

Table II: *J*-Couplings for rGGCGAGCC[a]

	^{1}H–^{1}H				
base	1′–2′	2′–3′	3′–4′	4′–5′	4′–5′
G1	2	overlap	8 ± 2	6	<3
G2	<2	5	9	4	<3
C3	<2	overlap	overlap	<3	<3
G4	2	5	10 ± 2	<3	<3
A5	<2	5	10	4	<3
G6	–	5	–	–	–
C7	<2	5	9 ± 2	<3	<3
C8	3	5	9 ± 2	<3	<3

	^{1}H–^{31}P			
base	H3′–P3′[b]	P5′–H5′	P5′–H5″	P5′–H4′
G2	9	overlap	7	<3
C3	7 ± 3	overlap	overlap	<3
G4	11	4	<3	overlap
A5	7	5	5 ± 3	<3
G6	10	overlap	5	overlap
C7	10 ± 3	5	<3	5
C8	10	<3	4	5

[a] The ^{1}H–^{1}H *J*-coupling uncertainty is ±1 Hz unless otherwise noted. Overlap indicates measurement not possible due to overlap. A dash indicates cross-peak not observed. The ^{1}H–^{31}P *J*-coupling uncertainty is ±2 Hz unless otherwise noted. [b] The coupling given is from the H3′ of the $n-1$ residue to the P of the n residue.

The sequence-specific assignments of H5′ and H5″ resonances were determined from weak H4′–H5′ and H4′–H5″ cross-peaks in the DQF-COSY spectrum, H5′/H5″ to P correlations in the ^{1}H–^{31}P HETCOR (Varani & Tinoco, 1991b; Sklenar et al., 1986), and by elimination from the base to sugar region of the 400-ms mixing time NOESY spectrum (supplementary material). The H5′/H5″ proton resonance which is more downfield shifted is assumed to be from the H5′ proton (Varani & Tinoco, 1991b), but no stereospecific restraints are derived from these assignments.

Assignment of the G6H2′ and G6H3′ resonances required special care. A cross-peak in the ^{31}P-decoupled DQF-COSY is observed between G6H2′ and G6H3′ (data not shown). Interestingly, the G6H2′–H3′ cross-peak in the DQF-COSY spectrum shows no passive couplings in either dimension; thus there is no way to distinguish which resonance is from H2′ and which is from H3′. A ^{31}P-*coupled* DQF-COSY spectrum (SantaLucia, 1991), however, shows passive coupling from G6H3′ to C7P, thus resolving the ambiguity without conformational assumption. This assignment is confirmed by the strong G6H3′–C7P correlation in the ^{1}H–^{31}P HETCOR (supplementary material). Strong cross-peaks in the short mixing time NOESY spectra from G6H8 to G6H3′ and from G6H2′ to C7H6 and a weak cross-peak from G6H8 to G6H2′ suggest the G6 sugar is predominantly C3′-endo. No DQF-COSY G6H3′–H4′ or G6H2′–H1′ cross-peaks are observed, however, to confirm this. The G6H4′ resonance was tentatively assigned by elimination from the ^{1}H–^{13}C HMQC spectrum as described above and shows that G6H4′ and G6H3′ have nearly the same chemical shift, which explains why no G6H3′–H4′ cross-peak is observed and no H3′–H4′ passive couplings are observed in H2′–H3′ cross-peaks. Also no G6H4′–H5′/H5″ cross-peaks are observed. Normally, this would suggest γ is gauche (g^+), not trans (t). For G6, however, H4′ has nearly the same chemical shift as H5′, so that no *J*-coupling would be observed whether γ is t or g^+. Thus, no restraints were placed on either G6 γ or δ. All cross-peaks in correlated experiments have been assigned. The completeness of assignments adds confidence in their reliability.

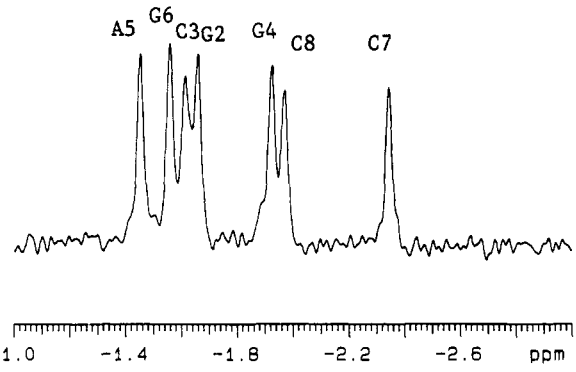

FIGURE 4: One-dimensional 202.3-MHz ^{31}P NMR spectrum (−1 to −3 ppm) with proton WALTZ decoupling of rGGCGAGCC at 35 °C. Note that seven resonances are observed as expected for an octamer without terminal phosphates. These results suggest the G·A mismatches in rGGCGAGCC are incorporated into the duplex without major distorsion of the backbone.

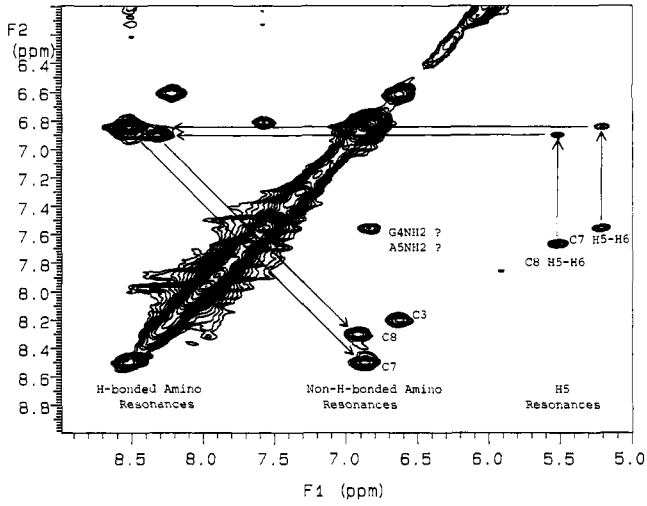

FIGURE 5: Expansion of the 130-ms mixing time NOESY-JRE spectrum of rGGCGAGCC dissolved in 90% H_2O at 22.8 °C, showing the H5/H6 to non-hydrogen-bonded and hydrogen-bonded amino region. The excitation maximum was set to 8 ppm. Assignments for C7 and C8 aminos follow directly from the H5–H6 assignments made in nonexchangeable spectra. The C3H5 resonance is not detected because it is too close to the water resonance (~0.3 ppm separation). Thus the assignment for C3 is made by elimination. An extra cross-peak between resonances at 7.60 and 6.85 ppm is tentatively assigned to $G4NH_2$ or $A5NH_2$.

Phosphorus assignments followed directly from the strong H3′–P5′ three bond correlations observed in the ^1H–^{31}P HETCOR (supplementary material). Seven ^{31}P resonances are observed (Figure 4) within a 1 ppm range (−1.4 to −2.4 ppm), suggesting that none of the phosphate α or ζ dihedral angles are trans (Gorenstein, 1984). This information was not included as restraints for structural calculations, however, since chemical shifts might be misleading for unusual structures.

Assignments of the Exchangeable Protons. The imino (10–15 ppm) and amino (6–9 ppm) regions of the ^1H NMR spectra recorded in H_2O provide information on hydrogen bonding between base pairs and mismatches. Imino proton spectra were recorded at different temperatures, pHs, and Mg^{2+} concentrations in order to find optimal conditions for making exchangeable proton assignments (SantaLucia, 1991).

Exchangeable amino proton resonances were assigned from the NOESY-JRE (Figure 5). Assignments of the C7 and C8 hydrogen-bonded and non-hydrogen-bonded amino resonances follow unambiguously from the H5–H6 resonances assigned

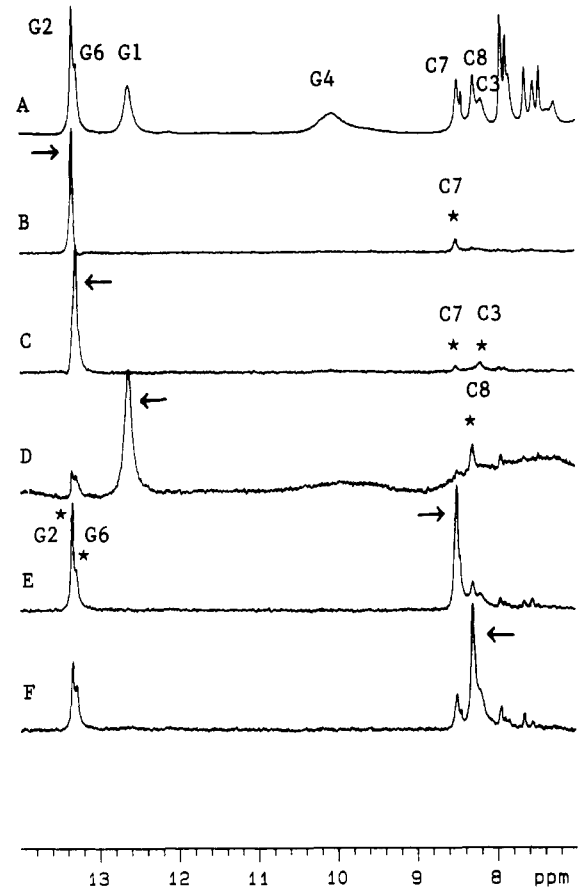

FIGURE 6: (A) 500-MHz ^1H NMR spectrum of 1 mM rGGCGAGCC at 20 °C in 80 mM NaCl, 10 mM phosphate, and 0.5 mM EDTA in 90% H_2O, pH 6.4. Difference spectra following 1-s saturation of (B) the resonance at 13.3 ppm, (C) the resonance at 13.25 ppm, (D) the resonance at 12.5 ppm, (E) the C7 hydrogen-bonded amino proton (8.51 ppm), and (F) the C8 hydrogen-bonded amino proton (8.35 ppm). For spectra B and C the saturation was applied 5-Hz off-resonance, and a power level which resulted in ~50% saturation was used to achieve selectivity. For spectrum C, high power was necessary to saturate the broad resonance. The saturated resonance is designated by an arrow while the observed NOE's are designated by asterisks. Assignments are shown above spectrum A.

in the nonexchangeable 2-D NMR spectra. The C3 amino resonances are at 8.20 and 6.63 ppm, as described below. The C3 H5–H6 cross-peak is not detected because it is too close to the H_2O resonance. An extra cross-peak between resonances at 7.54 and 6.82 ppm is observed and is tentatively assigned to $G4NH_2$, $A5NH_2$, or both. Specific assignment of the C3, C7, and C8 amino resonances is important for the assignment of imino resonances.

Figure 6 shows the imino and amino region of the 1-D NMR spectrum (7–14 ppm) and the NOE difference experiments used to make imino proton assignments. Four resonances are observed above 10 ppm, as expected for a self-complementary octamer with four G residues. Three resonances are observed between 12.5 and 13.5 ppm, typical of Watson–Crick G·C pairs (Carbonnaux et al., 1991). Upon irradiation of the resonance at 12.6 ppm (Figure 6D), a medium-intensity NOE is observed to the C8 hydrogen-bonded amino proton, allowing the assignment of the resonance at 12.6 ppm to G1. This assignment is consistent with the chemical shifts of G1 observed for other duplexes starting with GGC (SantaLucia et al., 1991b). A weak NOE is also observed from G1 to the hydrogen-bonded amino proton of C7. This NOE is likely due to spin diffusion via the C8 hydrogen-bonded amino. The "NOE's" observed at 13.25

and 13.3 ppm are spillover artifacts from the high decoupler power used to saturate the broad G1 resonance. Irradiation of the C7 hydrogen-bonded amino proton (Figure 6E) gives rise to a very strong NOE at 13.3 ppm, which is assigned to the G2 imino proton, and a weak NOE at 13.25 ppm, which is assigned to the G6 imino proton. Selective irradiation of the G2 resonance (Figure 6B) gives rise to an NOE at the C7 hydrogen-bonded amino proton, confirming the G2 assignment. Importantly, no NOEs are observed to C8 and C3, making the G2 assignment unambiguous and proving that selective irradiation was achieved. Selective irradiation of the resonance at 13.25 ppm (Figure 6C) gives rise to NOE's to C3 and C7. Importantly, no NOE is observed to C8. Thus the resonance at 13.25 ppm is assigned to G6. Irradiation of the C8 hydrogen-bonded amino proton (Figure 6F) does not give rise to an NOE at G1. This is not surprising, since chemical exchange with solvent is known to attenuate the NOE (Wemmer & Reid, 1985). The NOE at 13.25 ppm is likely from spillover since the C3 hydrogen-bonded proton is broad and resonates close to the C8 hydrogen-bonded proton. The NOE at 13.3 ppm (G2 imino) is likely from spin diffusion via the $C7NH_2$. By elimination, the broad (~80 Hz at 10 °C) resonance at 10.1 ppm is assigned to G4. The imino and amino assignments given above for $(rGGCGAGCC)_2$ were confirmed by analogy with $(rGAGCGAGCUC)_2$ whose imino resonances are well resolved (unpublished results).

Studies in DNA suggest that imino resonances below 11 ppm correspond to non-hydrogen-bonded imino protons as observed in loop structures (Haasnoot et al., 1980, 1983) or exposed imino protons from guanosine residues in the *syn* conformation (Gao & Patel, 1988; Kouchakdjian et al., 1989; Carbonnaux et al., 1991). Interestingly, however, the imino protons of hydrogen-bonded G·U (He et al., 1992) and U·U (SantaLucia et al., 1991b) mismatches often resonate below 11 ppm but are generally as sharp as Watson–Crick base pairs. The chemical shift and broadness of the G4 imino proton suggest it is not hydrogen bonded. With increasing temperature, G1, G2, and G6 imino resonances become narrower, presumably due to less aggregation, while the G4 imino resonance (10.1 ppm) becomes broader, presumably due to increased solvent exchange. Thus any structure deduced for $(rGGCGAGCC)_2$ must be able to rationalize why the G4 imino proton is not hydrogen bonded, even though a very stable structure is formed (SantaLucia et al., 1990).

Several studies of DNA oligonucleotides have found that G·A mismatch structure is dependent on pH (Gao & Patel, 1988; Kouchakdjian et al., 1989; Carbonnaux et al., 1991). The pH dependence of the imino region of the ^1H NMR spectrum indicates three of the resonances broaden slightly at pH 4.6 and one broadens substantially (G6), suggesting that lowering the pH induces chemical exchange (SantaLucia, 1991). This is consistent with previous results (SantaLucia et al., 1991) where, upon lowering the pH to 5.5, thermodynamic stability was decreased by ~0.4 kcal/mol. Studies of a hammerhead self-cleaving domain have also revealed that addition of Mg^{2+} can affect structure (Heus & Pardi, 1991a). Addition of 10 mM Mg^{2+} does not significantly affect the imino region of the NMR spectrum of $(rGGCGAGCC)_2$, suggesting that addition of Mg^{2+} does not change the structure. This is consistent with previous CD results for a similar RNA sequence with consecutive G·A mismatches (SantaLucia et al., 1990).

Structure Determination. Distance and dihedral angle restraints used in the structure determination are listed in Tables III and IV. Eleven of the structures generated by the rMD protocol described in the Materials and Methods section converged to satisfy *all* distance and dihedral angle restraints within 0.1 Å and 1°, respectively, with total violation energies less than 0.5 kcal/mol. These structures form G·A hydrogen bonds similar to those described by Li et al. (1991) and Heus and Pardi (1991b) (Figure 7). These have mutually similar simulated total energies (~−330 kcal/mol). Superposition of these structures revealed that the overall structure as well as local features such as dihedral angles (Table V), base pairing, and base stacking (Figure 8 and supplementary material) is well determined. The average RMSD for the all-atom pairwise superposition of these 11 structures is 0.70 Å. We note that the average RMSD observed depends on the number of energy minimization steps performed at the end of the refinement protocol. Fewer steps result in poorer superposition and total energy and vice versa. We chose 3000 steps as a compromise between obtaining structures with low

Table III: Distance Restraints (Å) for rGGCGAGCC[a]

intraresidue	H8/H6–H1′	H8/H6–H2′	H8/H6–H3′
G1	3.9 ± 0.8	4.0 ± 0.8	3.5 ± 0.6
G2	4.4 ± 0.8	3.6 ± 0.8	>2.4
C3	3.5 ± 0.8	3.3 ± 0.6	3.1 ± 0.6
G4	3.7 ± 0.8	3.5 ± 0.7	2.8 ± 0.6
A5	3.7 ± 0.8	4.0 ± 0.8	3.0 ± 0.6
G6	NR[b]	>3.5	2.9 ± 0.6
C7	3.6 ± 0.8	NR	2.7 ± 0.5
C8	3.7 ± 0.8	3.2 ± 0.6	2.5 ± 0.5

interresidue	H1′–H8/H6	H2′–H8/H6	H3′–H8/H6
G1–G2	4.7 ± 0.8	2.6 ± 0.5	>2.4
G2–C3	4.2 ± 0.8	2.7 ± 0.5	3.8 ± 0.8
C3–G4	4.2 ± 0.8	2.7 ± 0.5 (2.2)	3.0 ± 0.6 (2.6)
G4–A5	4.4 ± 0.8	3.2 ± 0.6 (2.7)	>4.0 (4.5)
A5–G6	4.5 ± 0.8	3.1 ± 0.6 (3.3)	2.7 ± 0.5 (2.2)
G6–C7	NR	2.3 ± 0.5	3.1 ± 0.6
C7–C8	3.8 ± 1.0	2.3 ± 0.5	3.5 ± 0.6

Adenine H2 Restraints			
A5(H2)–A5′(H1′)	3.2 ± 0.5 (2.8)	A5(H2)–A5′(H2′)	>4.5
A5(H2)–G4(H1′)	>4.5	A5(H2)–G4′(H2′)	>4.5
A5(H2)–G4′(H1′)	>4.5		
Base–Base Restraints			
G1(H8)–G2(H8)	4.5 ± 1.0	A5(H8)–G6(H8)	>4.5
G2(H8)–C3(H6)	>4.0	G6(H8)–C7(H6)	>4.5
C3(H6)–G4(H8)	>4.0	C7(H6)–C8(H6)	>4.5
G4(H8)–A5(H8)	>4.5		
Imino Restraints			
G1(H1)–G2(H1)	4.0 ± 1.0	G2(H1)–G6′(H1)	4.0 ± 1.0
H2′–H1′ Restraints			
G1(H2′)–G2(H1′)	4.2 ± 0.8	G6(H2′)–C7(H1′)	4.0 ± 0.8
C3(H2′)–G4(H1′)	>4.0	C7(H2′)–C8(H1′)	4.2 ± 0.8
G4(H2′)–A5(H1′)	>4.5		
Pyrimidine H5 Restraints			
C3(H5)–G2(H1′)	>4.5	C7(H5)–G6(H3′)	4.0 ± 1.0
C3(H5)–G2(H2′)	4.5 ± 1.0	C7(H5)–G6(H8)	>4.0
C3(H5)–G2(H8)	>4.0	C7(H5)–C8(H5)	3.8 ± 1.0
C3(H5)–G4(H1′)	>4.5	C7(H5)–C8(H6)	>4.0
C3(H5)–G4(H8)	>4.0	C8(H5)–C7(H2′)	4.5 ± 1.0
C7(H5)–G6(H2′)	4.5 ± 1.0		
Other Restraints			
G4(H3′)–C3(H2′)	>4.0	A5(H8)–G4(H4′)	>4.0
G4(H3′)–C3(H3′)	>4.0	G4(H1′)–A5(H1′)	>4.5
G4(H3′)–A5(H3′)	>4.5	G4(H1′)–A5(H4′)	>4.5

[a] Distances are given only for strand 1. Primed residues are from the complementary strand. For strand 2, equivalent distances were applied. Hydrogen bonds between Watson–Crick base pairs were also included as distance restraints (1.8 ± 0.2 Å). Hydrogen bonds between G·A mismatches were not used as restraints. Values in parentheses are averages of the observed distances in the 11 final structures which were critical for the structure determination. [b] Not restrained.

Table IV: Torsion Angle Restraints (deg) for (rGGCGAGCC)$_2$[a]

residue	α	β	γ	δ	ε	ζ	χ
G1	na	na	NR	80 ± 20	−155 ± 15	NR	NR
G2	NR	180 ± 20	60 ± 15	80 ± 20	−155 ± 30	NR	NR
C3	NR	180 ± 40	60 ± 15	80 ± 20	−120 ± 40	NR	NR
G4	NR	180 ± 15	60 ± 15	80 ± 20	−120 ± 60	NR	NR
A5	NR	180 ± 60	60 ± 40	80 ± 20	−120 ± 40	NR	NR
G6	NR	180 ± 40	NR	NR	−155 ± 15	NR	NR
C7	NR	180 ± 30	60 ± 15	80 ± 20	−155 ± 15	NR	NR
C8	NR	180 ± 20	60 ± 15	NR	na	na	NR
A-form[b]	−68	178	54	82	−153	−71	−158

[a] NR indicates dihedral angle was not restrained. na indicates not applicable. [b] Saenger, 1984.

Table V: Average Torsion Angles (deg) for the 11 Final Structures of (rGGCGAGCC)$_2$[a]

residue	α	β	γ	δ	ε	ζ	χ	P
G1	na	na	na	93.3 ± 1.1	**−142.6 ± 2.5**	−78.4 ± 1.6	**177.1 ± 1.8**	−2.6 ± 2.2
G2	−76.6 ± 2.0	175.3 ± 4.8	57.6 ± 1.2	90.4 ± 0.3	−153.7 ± 3.1	−63.7 ± 4.7	**−169.8 ± 1.1**	5.0 ± 0.8
C3	−77.9 ± 1.2	174.4 ± 3.5	59.3 ± 1.0	88.7 ± 1.1	−158.9 ± 1.6	**−53.6 ± 1.6**	−158.4 ● 0.7	9.1 ± 2.6
G4	**−81.1 ± 1.3**	−177.8 ± 1.2	54.4 ± 0.2	90.6 ± 1.6	**−179.0 ± 0.9**	**−106.1 ± 2.3**	−158.3 ± 2.3	6.3 ± 3.2
A5	−73.5 ± 3.3	**−165.4 ± 2.7**	55.7 ± 0.5	94.4 ± 1.5	**−138.5 ± 1.4**	**−56.4 ± 6.3**	−168.2 ± 2.6	−2.9 ± 0.8
					149.5 ± 1.2[b]			
G6	−67.6 ± 1.0	**165.0 ± 1.0**	61.6 ± 0.5	91.4 ± 2.6	−156.7 ± 0.7	−67.6 ± 1.5	−167.3 ± 0.8	2.4 ± 4.5
	134.2 ± 1.0[b]	**−146.0 ± 2.8[b]**	**178.5 ± 0.6[b]**		−146.0 ± 0.8[b]		**175.0 ± 0.6[b]**	
C7	−78.0 ± 0.8	178.4 ± 2.7	55.8 ± 0.8	89.6 ± 0.4	−153.2 ± 1.9	−68.3 ± 2.2	−161.8 ± 3.2	7.5 ± 0.8
C8	−72.4 ± 1.7	170.5 ± 2.6	56.7 ± 0.7	81.7 ± 0.6	na	na	−157.2 ± 0.8	14.1 ± 1.7
A-form[c]	−68	178	54	82	−153	−71	−158	

[a] Errors given are standard deviations. P is the pseudorotation angle. Entries in boldface differ from A-form by more than 10°. [b] Average angles for the five structures with G6 γ *t* (see Discussion). [c] Saenger, 1984.

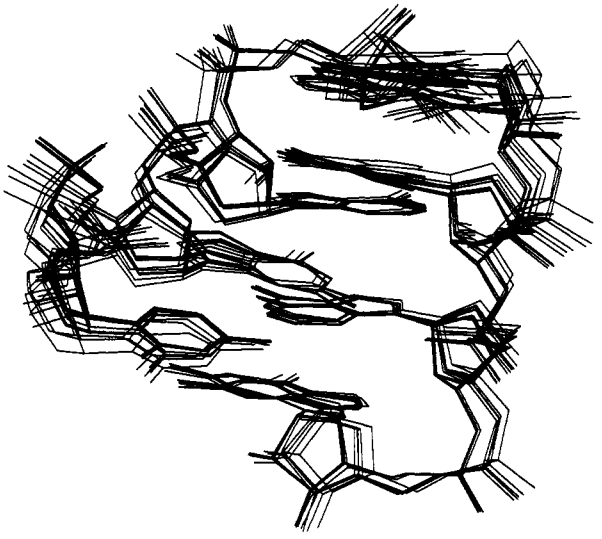

FIGURE 7: Minor groove view of the superposition of the 11 final structures. Shown are the G·A mismatches and the closing base pairs.

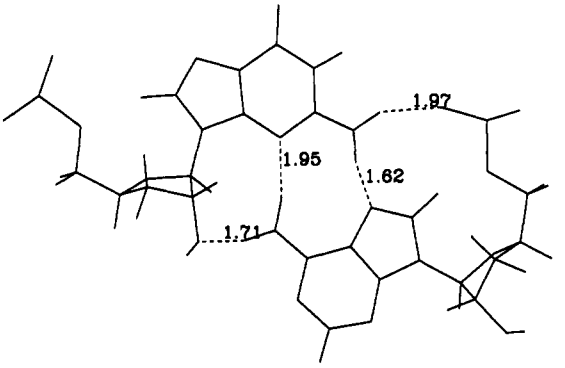

FIGURE 8: G·A mismatch from one of the structures. Dashed lines indicate functional groups which are close enough to form hydrogen bonds. The numbers indicate the distance in angstroms between the hydrogen and hydrogen-bond acceptor of the closest dashed line. The G4 imino proton is 3.6 Å from the A5′ phosphate oxygen. Model building suggests a water molecule could form bridging hydrogen bonds between these functional groups.

energy and fairly representing the precision of the structure determination. The terminal residues are not as well determined as the internal residues (supplementary material), presumably because of fewer restraints on these residues and propagation of small errors from the internal base pairs.

The remaining structure was approximately 40 kcal/mol less stable than the other 11, due to less favorable van der Waals and electrostatic interactions. This structure has essentially the same dihedral angles as the other 11 except the G4 and G4′ χ dihedral angles are +170° instead of −155° (35° difference). This causes the guanines from the two mismatches to stack on top of one another with the opposite faces used in the other 11 structures. This structure has a marginally higher total violation energy of 3.5 kcal/mol and does not rationalize the thermodynamic consequences of functional group substitutions (SantaLucia et al., 1991a). The energy of this structure suggests it is stuck in a suboptimal energy well; thus, this structure is not discussed further.

DISCUSSION

Qualitative Structural Analysis. Data from exchangeable ^{1}H NMR indicate that Watson–Crick base pairs are formed between G1–C8, G2–C7, and C3–G6. The chemical shift and broad line width of the G4 imino proton suggest the G4 imino proton is not hydrogen bonded. Except for G6, the base to H1′ region indicates all bases are in the *anti* conformation. G6 is also thought to be in the *anti* conformation since the imino proton is hydrogen bonded and a strong G6H8 to G6H3′ NOE is observed. Weak or nonexistent cross-peaks in DQF-COSY and 2-quantum spectra (SantaLucia, 1991) indicate H1′–H2′ couplings of <2 Hz, which suggest predominant C3′-endo sugar pucker for all nonterminal residues. These sugar pucker assignments are confirmed by large (8–10 Hz) H3′–H4′ couplings (Table II). The G6 sugar is also thought

Table VI: Helical Parameters for (rGGCGAGCC)$_2$[a]

base pair	helical twist (deg)	slide (Å)	roll (deg)	propeller twist (deg)	buckle (deg)	inclination (deg)	X-DSP (Å)
G1–C8′	35.1	−1.7	−0.4	−7.3	11.6	2.4	−2.6
G2–C7′	32.9	−1.1	−0.3	−8.4	8.4	3.3	−2.2
C3–G6′	17.9	−1.2	3.6	−12.0	−5.9	4.0	−1.3
G4–A5′	80.8	−1.5	0.3	2.0	−6.4	3.0	−1.3
av[b]	36.1	−1.4	0.9	−6.4	0.0	3.2	−1.9
A-form	32.7	−1.5	−0.4	13.8	0.2	16.9	−4.4

[a] Parameters given are for 1 of the 11 final structures (G6 γ is g^+). Helical parameters are listed according to the Cambridge convention (Dickerson, 1989) and calculated with the program NEWHELIX93. Slide, roll, propeller twist, buckle, inclination, and X-DSP are calculated from the best plane through both bases. Helical twist is calculated from C1′–C1′ vectors of consecutive base pairs. X-DSP is the displacement of base pairs from the helix center. [b] The averages are for the entire duplex.

to be C3′-endo on the basis of H8 to H2′ and H3′ NOE's as described in the Results section. Strong cross-peaks are observed in the 100-ms mixing time NOESY spectrum from base H8/H6 to 5′-neighboring H2′ protons for base protons of residues G2, C3, C7, and C8, consistent with an A-form-like structure. The base protons of residues G4, A5, and G6 show medium-strength NOE's to 5′-neighboring H2′ protons, suggesting deviations from the A-form structure near the G·A mismatches. All resonances in the 1-D ^{31}P spectrum are within a 1 ppm range, suggesting that the G·A mismatches are incorporated into the helix without trans α or ζ dihedral angles (Gorenstein, 1984). The chemical shift of G6H1′ is probably between 4 and 4.6 ppm, suggesting an unusual structure on the 3′ side of the internal loop. A medium-strength NOE (corresponding to a distance of 3.2 ± 0.5 Å) is observed from the A5H2 to the cross-strand A5′H1′. This suggests that the minor groove near the mismatches is significantly smaller than would be predicted for a helix with G·A mismatches forming hydrogen bonds involving Watson–Crick positions. The G·A mismatches are also known to involve hydrogen bonds from the A5NH$_2$ to some functional group, since the stability of (rGGCGAGCC)$_2$ decreases by 3.0 kcal/mol when A5 residues are replaced with purines (SantaLucia et al., 1991a). The H8, H6, and H5 resonances of G1, G2, G4, A5, C7, and C8 are sharp (3–5 Hz line width), while those for C3 and G6 are slightly broadened (6 and 7 Hz, respectively). This suggests that the structure is somewhat dynamic at the interfaces of the helical stems and the internal loop.

Quantitative Structural Analysis. The results given in Table V indicate the structure of (rGGCGAGCC)$_2$ is well determined. Close examination of the structures revealed two families of structures equally consistent with the NMR data (Table V). These two different structures are the result of the G6 γ dihedral angle in either g^+ or t conformations. Changes in G6 α and β dihedral angles along with minor perturbations in G6 ϵ and χ compensate for the change in γ so that the two families of structures have nearly equivalent stacking (Figure 7). The phosphorus chemical shift data suggest none of the phosphates have t α or ζ angles, which would argue against the G6 γ t conformation. The G6 α is at the edge of the t range, and the structure is unusual at this base-pair step; thus ^{31}P chemical shift arguments may be misleading. Sugar pucker amplitudes are not reported in Table V because J-couplings were not measured precisely enough to define this parameter. The observed pucker amplitudes ranged from 41° to 44°, somewhat higher than the usual range of 35 ± 5° observed for C3′-endo pucker. This is probably an artifact of the force field.

The functional groups of the G·A mismatches are close enough in the 11 structures to form hydrogen bonds similar to those observed in a DNA duplex (Li et al., 1991) and an RNA hairpin (Heus & Pardi, 1991). Analysis of the structures shows that average hydrogen bond distances in the G·A mismatches are 1.7 ± 0.1, 1.9 ± 0.1, 1.7 ± 0.1, and 1.9 ± 0.1 for adenine amino proton to guanosine 2′-oxygen, adenine amino proton to guanine N3, guanine amino proton to adenine N7, and guanine amino proton to adenine phosphate oxygen, respectively (Figure 8). Examination of structures just before the last energy minimization showed that the guanine amino proton to adenine phosphate oxygen distance is less than 4 Å for all structures before energy minimization and that the distance decreases during energy minimization, without substantially improving restraint violations. Since the force field used for molecular dynamics and energy minimization does not include solvent, we cannot rule out the possibility that the guanine amino to adenine phosphate interaction is mediated by a water molecule (SantaLucia et al., 1992).

Several NOE's were critical for determining the geometry of the G·A mismatches (Table III). The cross-strand A5H2 to A5′H1′ distance determines the relative positions of the cross-strand adenines and creates a narrow minor groove. For A-form geometry, the base H8/H6 to n − 1 H2′ and H3′ distances are 2.1 and 3.0 Å, respectively (Wuthrich, 1986). For the G4–A5 step, however, the G4H2′–A5H8 NOE is weaker than normal and the G4H3′–A5H8 NOE is not observed at short mixing time (Figure 2). The resulting restraints push the G4 and A5 bases away from one another so that the helix is overwound at the G4–A5 step (helical twist = 80.8°; cf. Table VI). On the other hand, at the A5–G6 step the A5H2′–G6H8 NOE is weaker than observed for A-form geometry, but the A5H3′–G6H8 NOE is stronger (Figure 2). The resulting restraints push the A5 and G6 bases closer to one another so that the helix is underwound at the A5–G6 step (helix twist = 17.9°; cf. Table VI). The other inter- and intraresidue NOEs and dihedral restraints for these residues are fairly normal but do restrict the conformation space consistent with the unusual NOE's. The fact that no hydrogen bond orientation between G·A mismatches was assumed and 11 of the structures form the same hydrogen-bonding orientation adds confidence to the proposed structure. The structure also rationalizes qualitative aspects of the NMR data not included as restraints (e.g., the unpaired guanine imino proton, phosphorus chemical shifts, and G6H1′ chemical shift).

Figure 7 indicates the stacking between G·A mismatches and between G·A mismatches and closing G·C base pairs is well determined. Table VI gives helical parameters for (rGGCGAGCC)$_2$ according to the Cambridge convention (Dickerson, 1989) and calculated with the program NEWHELIX (provided by R. E. Dickerson). The twist angle between G·A mismatches is 80.8°, considerably overwound compared to A-form. On the other hand, the twist angle between closing G·C base pairs and G·A mismatches is 17.9°, considerably underwound compared to A-form. The net effect is that the overall helix geometry is similar to an A-form duplex. The unusual twist angles are required to accommodate the short C1′–C1′ distances (~9.1 Å) for the G·A mismatches

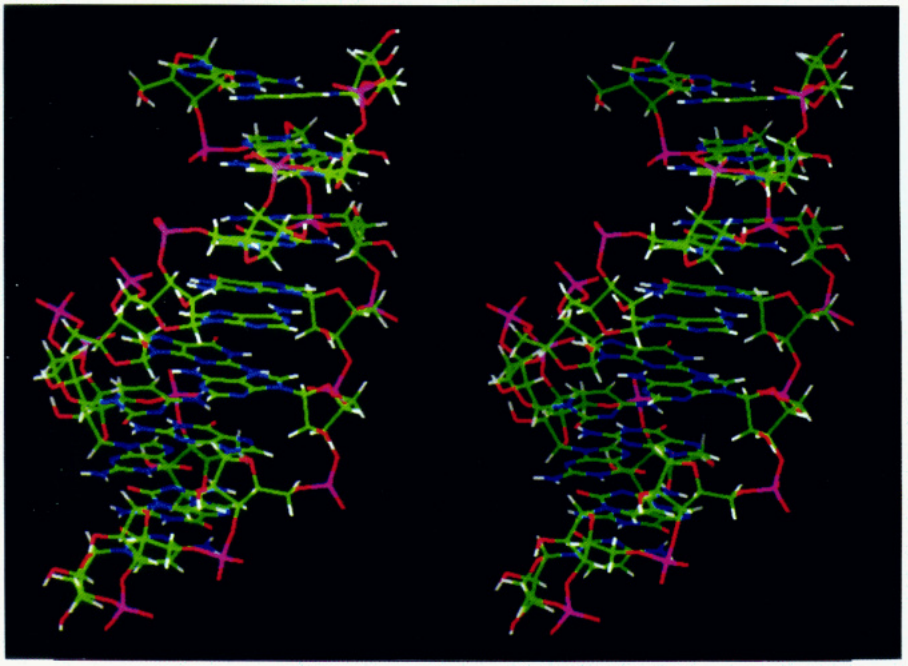

FIGURE 9: Stereoview into the minor groove of 1 of the 11 final structures (G6 γ is g^+). White, green, blue, red, and purple atoms are hydrogen, carbon, nitrogen, oxygen, and phosphorus, respectively.

compared with Watson–Crick G·C base pairs (10.4 Å). Figure 9 shows the minor grove stereoview of 1 of the 11 (rGGCGAGCC)$_2$ structures (G6 γ g^+).

Comparison of RNA and DNA Structures. Recently, several groups have performed detailed NMR investigations of DNA duplexes with tandem G·A mismatches similar to those reported here (Chou et al., 1992a,b; Ebel et al., 1992; Lane et al., 1992). The DNA and RNA structures share many common features but also have some important differences. The hydrogen bonding between bases and stacking between adjacent G·A mismatches are nearly identical in the two structures [compare Figure 10 with Figure 10c from Chou et al. (1992b)]. However, DNA and RNA accomplish this stacking orientation differently. In the DNA structure, the phosphate between the G-A dinucleoside steps has α and ζ dihedral angles of –142° and +96° (B_{II} phosphate), respectively, consistent with the observed downfield-shifted phosphorus resonance (Ebel et al., 1992; Chou et al., 1992a). In RNA, the equivalent dihedral angles are –74° and –106°, respectively, consistent with the normal position of the phosphorus resonance. Each G·A mismatch in RNA, however, can also form two additional hydrogen bonds compared to DNA, one from the adenine amino to the cross-strand guanosine 2′-oxygen (which is not present in DNA) and one from the G amino to the cross-strand adenine phosphate. On the other hand, the G·A mismatches in DNA have large intrastrand overlap with the closing Watson–Crick pairs, whereas in RNA, there is little intrastrand overlap at the interface between the G·A mismatches and the Watson–Crick pairs (Figure 10). Presumably, this difference is the result of B-form vs A-form stems. Interestingly, the unusual stacking geometry places the A5 six-membered ring below the G6H1′ (average proton to base plane distance is 2.7 Å, and displacement from the center of the six-membered ring is always <0.5 Å), which might explain the unusual chemical shift of this proton (Giessner-Prettre & Pullman, 1976). As a result of these differences, the H1′ resonance for the closing base-pair residue on the 3′ side of the G·A mismatches is broadened and probably shifted more upfield in RNA (tentatively assigned at 4.32 ppm) than in DNA (5.26 or 5.38 ppm). The poor stacking of closing G·C base pairs on G·A mismatches also rationalizes the slight dynamic character observed for G6 and C3.

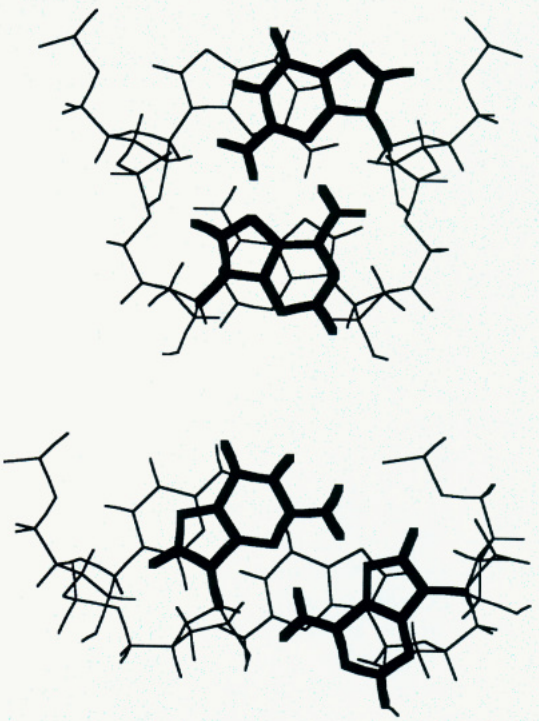

FIGURE 10: Base stacking between (top) both G·A mismatches and (bottom) the G·A mismatch and closing C·G base pair. Bases shown in bold are closer to the viewer. Note the extensive interstrand overlap between adenines and guanines in the top figure and weak overlap in the bottom figure.

Comparison of (rGGCGAGCC)$_2$ with the GCAA Hairpin. The structure of the hairpin rGGGCGCAAGCCUAUA reported by Heus and Pardi (1991b) contains a G·A mismatch which forms all the hydrogen bonds reported here for (rGGCGAGCC)$_2$. The weak stacking between the G·A mismatch and the closing C·G base pair are also similar in the two structures. Quantitative comparison on the other

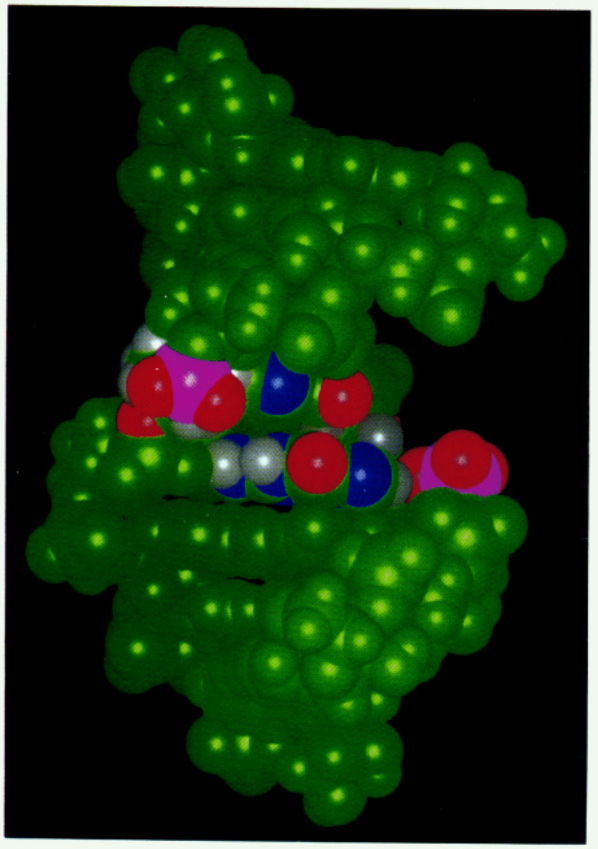

FIGURE 11: Space-filling major groove view (the guanosines from the G·A mismatches are highlighted). Note that Watson–Crick and Hoogsteen sites of the guanosines in the G·A mismatches are available and form a putative binding site for proteins or tertiary interactions.

side of the G·A mismatch is not possible since the sequences are different. Qualitatively, however, the intrastrand stacking between loop adenines observed in the GCAA hairpin is similar to the interstrand stacking between adenines in (rGGCGAGCC)$_2$.

Implications of the RNA Structure. Tandem G·A mismatches in the sequence 5′G–A3′ are common in 16S and 23S rRNA secondary structures (Gutell et al., 1985; Gutell & Fox, 1988) and at the end of one of the helices of hammerhead ribozymes (Forster & Symons, 1987a,b), suggesting a functional role. Striking features of the structure of (rGGCGAGCC)$_2$ are the availability for hydrogen bonding of the Watson–Crick positions of the A and G residues of the G·A mismatches in the minor and major grooves, respectively (Figures 9 and 11). The A and G residues are therefore in position to form tertiary or binding contacts with other nucleotides or with proteins. Another possibility is that tandem G·A mismatches could serve as a conformational switch between the structure observed for (rGGCGAGCC)$_2$ and some other G·A pairing orientation which is close in free energy.

ACKNOWLEDGMENT

We thank R. Kierzek for help with RNA synthesis. We thank I. Tinoco, Jr., for providing support (NIH Grant GM10840) for the part of this work performed at Berkeley and for critical reading of the manuscript. We also thank Peter Müller (Dr. Karl Thomae, GmbH) for useful discussions about molecular dynamics simulations and J. Jaeger, B. Wimberly, and G. Varani for assistance in acquiring ^{1}H–^{31}P HETCOR, ^{31}P-decoupled DQF-COSY, and ^{1}H–^{13}C HMQC spectra, respectively.

SUPPLEMENTARY MATERIAL AVAILABLE

Seven figures showing (1) the one-dimensional 500-MHz ^{1}H NMR spectrum (3.8–8.2 ppm), (2) ^{13}C-decoupled ^{1}H–^{13}C HMQC, (3) the H1′ to sugar region of the 100-ms mixing time NOESY spectrum, (4) the base to H1′ region of the 100-ms mixing time NOESY spectrum, (5) the base to sugar region of the 400-ms mixing time NOESY spectrum, (6) the ^{1}H–^{31}P HETCOR spectrum, and (7) superposition of the 11 final structures [full (rGGCGAGCC)$_2$ duplex shown] (8 pages). Ordering information is given on any current masthead page.

REFERENCES

Braunschweiler, L., Bodenhausen, G., & Ernst, R. R. (1983) *Mol. Phys. 48*, 535–560.

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

Carbonnaux, C., van der Marel, G. A., van Boom, J. H., Guschlbauer, W., & Fazakerley, G. V. (1991) *Biochemistry 30*, 5449–5458.

Cheong, C. Varani, G. , & Tinoco, I., Jr. (1990) *Nature 346*, 680–682.

Chou, S.-H., Flynn, P., & Reid, B. (1989) *Biochemistry 28*, 2422–2435.

Chou, S.-H., Cheng, J.-W., Fedoroff, O. Y., Chuprina, V. P., & Reid, B. R. (1992a) *J. Am. Chem. Soc. 114*, 3114–3115.

Chou, S.-H., Cheng, J. W., & Reid, B. R. (1992b) *J. Mol. Biol. 228*, 138–155.

Dickerson, R. E. (1989) *Nucleic Acids Res. 17*, 1797–1803.

Ebel, S., Lane, A. N., & Brown, T. (1992) *Biochemistry 31*, 12083–12086.

Ernst, R. R., Bodenhausen, G., & Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, New York.

Fazakerley, G. V., Quignard, E., Woisard, A., Guschlbauer, W., van der Marel, G. A., van Boom, J. H., Jones, M., & Radman, M. (1986) *EMBO J. 5*, 3697–3703.

Forster, A. C., & Symons, R. H. (1987a) *Cell 49*, 211–220.

Forster, A. C., & Symons, R. H. (1987b) *Cell 50*, 9–16.

Gao, X., & Patel, D. J. (1988) *J. Am. Chem. Soc. 110*, 5178–5182.

Giessner-Prettre, C., Pullman, B., Borer, P. N., Kan, L.-S., & Ts'O, P. O. P. (1976) *Biopolymers 15*, 2277–2286.

Gorenstein, D. (1984) in *^{31}P NMR, Principles and Applications*, Academic Press, New York.

Gutell, R. R., & Fox, G. E. (1988) *Nucleic Acids Res. 16* (Suppl.), r175–r269.

Gutell, R. R., Weiser, B., Woese, C. R., & Noller, H. F. (1985) *Prog. Nucleic Acid Res. Mol. Biol. 32*, 155–216.

Haasnoot, C. A. G., den Hartog, J. H. J., de Rooij, J. F. M., van Boom, J. H., & Altona, C. (1980) *Nucleic Acids Res. 8*, 169–181.

Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnedijk, T. J. J., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1983) *J. Biomol. Struct. Dyn. 1*, 115–129.

He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., & Turner, D. H. (1991) *Biochemistry 30*, 11124–11132.

Heus, H. A., & Pardi, A. (1991a) *J. Mol. Biol. 217*, 113–124.

Heus, H. A., & Pardi, A. (1991b) *Science 253*, 191–193.

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.

Kan, L.-S., Chandrasegaran, S., Pullford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.

Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. (1989) *Biochemistry 28*, 5647–5657.

Lane, A., Martin, S. R., Ebel, S., & Brown, T. (1992) *Biochemistry 31*, 12087–12095.

Li, Y., Zon, G., & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.
Macura, S., & Ernst, R. R. (1979) *J. Magn. Reson. 46*, 269–282.
Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711–736.
Nikonowicz, E. P., & Gorenstein, D. G. (1992) *J. Am. Chem. Soc. 114*, 7494–7503.
Noller, H. F. (1984) *Annu. Rev. Biochem. 53*, 119–162.
Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1987) *Eur. J. Biochem. 170*, 225–239.
Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 264–268.
Saenger, W. (1989) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
SantaLucia, J., Jr. (1991) Ph.D. Thesis, University of Rochester, Rochester, NY.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813–8819.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991a) *J. Am. Chem. Soc. 113*, 4313–4322.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991b) *Biochemistry 30*, 8242–8251.
SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1992) *Science 256*, 217–219.
Scheek, R. R., Boelens, R., Russo, N., van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371–1376.
Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469.
Sklenar, V., & Feigon, J. (1990) *Nature 345*, 836–838.
Sklenar, V., Miyoshiro, H., Zon, G., & Bax, A. (1986) *FEBS Lett. 208*, 94–98.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Traub, W., & Sussman, J. L. (1982) *Nucleic Acids Res. 10*, 2701–2708.
Usman, N., Ogilvie, K. K., Jiang, M.-V., & Cedergren, R. (1987) *J. Am. Chem. Soc. 109*, 7845–7854.
Varani, G., & Tinoco, I., Jr. (1991a) *J. Am. Chem. Soc. 113*, 9349–9354.
Varani, G., & Tinoco, I., Jr. (1991b) *Q. Rev. Biophys. 24*, 479–532.
Varani, G., Wimberly, B., & Tinoco, I., Jr. (1989) *Biochemistry 28*, 7760–7772.
Wang, A. C., Kim, S. G., Flynn, P. F., Chou, S.-H., Orban, J., & Reid, B. R. (1992) *Biochemistry 31*, 3940–3946.
Wemmer, D. E., & Reid, B. R. (1985) *Annu. Rev. Phys. Chem. 36*, 105–137.
Wimberly, B., Varani, G., & Tinoco, I., Jr. (1993) *Biochemistry 32*, 1078–1087.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.