Solution Structure of an Unusually Stable RNA Tetraplex Containing G- and U-Quartet Structures†

Chaejoon Cheong and Peter B. Moore*

Department of Chemistry, Yale University, 225 Prospect Street, New Haven, Connecticut 06511

Received February 6, 1992; Revised Manuscript Received June 5, 1992

ABSTRACT: A model for the solution structure of an RNA tetraplex, $(rUGGGGU)_4$, has been obtained by two-dimensional NMR spectroscopy and molecular dynamics. The molecule is parallel stranded and Hoogsteen base-paired in 50 mM KCl, and it is so stable that three of its six imino protons have exchange half-lives measured in days at 40 °C. The tetraplex is stabilized by base stacking and by the hydrogen bonds in four G quartets and at least one U quartet. This is the first indication of the existence of U-quartet structures of which we are aware.

It has been known since the 1960s that guanine readily forms planar, Hoogsteen-paired, quartets that are stabilized preferentially by K^+. Guanine mononucleotides will do it (Zimmerman, 1976; Pinnavaia et al., 1987; Borzo et al., 1980), and poly(rG) forms four-stranded helices that depend on the same G-quartet motif (Zimmerman et al., 1975). Interest in G-quartet structures has been stimulated lately by observations suggesting their involvement in the recombinational events required for immunoglobulin maturation (Sen & Gilbert, 1988, 1990), and in telomere function (Williamson et al., 1989; Sundquist & Klug, 1989).

Recently we observed that an RNA oligonucleotide derived from *Escherichia coli* 5S RNA, which contains a run of four G residues near its 3′-terminus, aggregates in the presence of K^+ or Na^+, and we were able to demonstrate that these aggregates are tetrameric and parallel stranded and depend on interactions involving G residues (Kim et al., 1991). Below we report the results of an NMR[1] analysis of a simpler, but similarly structured, tetrameric aggregate, the one formed by rUG_4U in KCl.

The NMR data confirm that the UG_4U aggregate is indeed a parallel-stranded tetramer and show that it is held together by Hoogsteen-paired G quartets. $(UG_4U)_4$ is so stable that it takes days for the imino protons of 3 of its 4 G quartets to exchange with solvent at 40 °C. Another surprising finding is that the U's at its ends form quartets stabilized by N3–O4 hydrogen bonds. The U quartet at the 3′ end of the molecule is stable enough to give an imino proton resonance that persists even at 65 °C. The proton and phosphorus NMR spectra of this molecule have been assigned. The "high-resolution" model for $(UG_4U)_4$ that has resulted from the analysis of its NOE's and coupling constants is presented and discussed. Its possible biological relevance is also considered.

MATERIALS AND METHODS

RNA Synthesis and Purification. rUG_4U was synthesized chemically by the Yale University School of Medicine Protein and Nucleic Acid Chemistry Facility on an Applied Biosystems 380B synthesizer. The sample was deprotected as described by Webster and Spicer (1991) and then purified by HPLC on a Nucleogen-DEAE 60-7 ion-exchange column at 50 °C, using 20 mM MES, pH 6.0, and 5 M urea and a LiCl gradient. Even under these strongly denaturing conditions, a considerable amount of the material came out in tetrameric form, which made purification difficult. Purified RNA was desalted by dialysis using Spectra/Por membrane tubing (MW cutoff = 1000) and 20 mM potassium phosphate buffer, pH 7. The sample volume was decreased to 400 μL by lyophilization, and it was then dialyzed against an appropriate NMR buffer (see below) for >24 h using a Microdialysis System 1200 MA (Bethesda Research Laboratories) and Spectra/Por membrane tubing (MW cutoff = 1000).

Preparation of NMR Sample. Two buffers were used for NMR spectroscopy: (i) 10 mM potassium phosphate, pH 6.7, 50 mM KCl, and 0.5 mM EDTA and (ii) 18 mM potassium phosphate, pH 5.2, 88 mM KCl, and 0.9 mM EDTA. Following dialysis, samples to be studied in H_2O were lyophilized and dissolved in 400 μL of degassed water consisting of 360 μL of H_2O and 40 μL of 99.9% D_2O (MSD Isotopes). Dioxane (0.1 mM) was used as a chemical shift standard (3.741 ppm at all temperatures). Nonexchangeable proton NMR spectroscopy was done on samples that were lyophilized three times from 99.9% D_2O and then dissolved in 99.996% D_2O (Cambridge Isotope Laboratories). The final RNA concentration (strand concentration) was 2.2 mM.

NMR Spectroscopy. Data were collected either on a Bruker AM-500 spectrometer (500 MHz for ^1H and 202 MHz for ^{31}P) or on the Yale Chemical Instrumentation Center 490-MHz spectrometer. The data were transferred to a VAX/VMS computer and processed using FTNMR/VMS (Hare Research, Inc.).

(A) 1-D NMR. 1-D spectra were taken in pH 6.7 buffer at various temperatures using a 1–1 spin-echo water suppression pulse sequence (Sklenar & Bax, 1987). The excitation maximum was placed on the imino proton resonances (11.4 ppm), and the sweep width was 8200 Hz.

(B) NOESY in D_2O. NOESY spectra in both pH 6.7 and pH 5.2 D_2O buffers were recorded with pure absorption phase in four quadrants (States et al., 1982). Five NOESY spectra were taken in pH 6.7 buffer, with 50-, 100-, 200-, and 300-ms mixing times at 40 °C, and with a 400-ms mixing time at 30 °C. One additional NOESY spectrum was taken in pH 5.2

† This work was supported by a grant from NIH (GM 41651) to P.B.M. and a Damon Runyon–Walter Winchell Cancer Research Fund Fellowship (DRG 1111) to C.C.

[1] Abbreviations: HPLC, high-pressure liquid chromatography; MES, 2-(*N*-morpholino)ethanesulfonic acid; EDTA, ethylenediaminetetraacetic acid; NMR, nuclear magnetic resonance; FID, free induction decay; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; DQF-COSY, double-quantum-filtered correlated spectroscopy; TOCSY, totally correlated spectroscopy.

0006-2960/92/0431-8406$03.00/0

buffer at 5 °C with a 250-ms mixing time. The residual HDO peak was preirradiated during repetition delays (1.8–1.9 s). A total of 210–220 FID's of 2K complex data points were collected for both real and imaginary parts in the t_1 dimension (States et al., 1982), with a 4630–4900-Hz sweep width. A total of 80–96 scans were averaged for each FID. Total acquisition times were 20–22 h. Data were zero-filled to 1K real points and apodized using skewed sine bells in both dimensions.

(*C*) *NOESY in H_2O.* NOESY spectra were recorded at pH 5.2 and pH 6.7 at 5 °C with a 200-ms mixing time. All three 90° pulses were replaced by 1–1 spin-echo pulses, and a 30-ms *z*-homospoil pulse was applied during mixing time (Sklenar & Bax, 1987). The H_2O peak were preirradiated during repetition delay (1.8 s) to allow increased receiver gain. The excitation maximum was set to 11.4 ppm. A total of 220 FID's were collected (States et al., 1982) with a 8200-Hz sweep width. Standard CYCLOPS phase cycling (Hoult & Richards, 1975) was used, and 128 scans were averaged for each 2K complex FID. The data were processed as described for NOESY spectra taken in D_2O.

(*D*) *DQF-COSY.* A ^{31}P-decoupled, double-quantum-filtered, COSY spectrum was recorded at 40 °C in pH 6.7 buffer to help assign proton resonances and to measure homonuclear scalar coupling constants. Broad-band ^{31}P decoupling was achieved using a GARP pulse sequence (Shaka et al., 1985). The residual HDO peak was preirradiated during repetition delays (1.7 s). TPPI (Marion & Wüthrich, 1983) with standard CYCLOPS phase cycling was used. A total of 720 FID's of 4K complex data points each were collected with a 3205-Hz sweep width. A total of 64 scans were averaged for each FID. A high-resolution (0.8 Hz/point in ω_2) spectrum was produced by zero-filling to 4K real points. The data were apodized by sine bells in both dimensions before Fourier transformation.

(*E*) *Heteronuclear COSY.* A proton-detected, ^{1}H–^{31}P heteronuclear COSY spectrum (Sklenar et al., 1986) was recorded at 40 °C in pH 6.7 buffer both for assignment purpose and to measure ^{1}H–^{31}P scalar coupling constants. A total of 128 FID's were collected (States et al., 1982) with spectral widths of 1000 Hz in ω_1 (^{31}P dimension) and 1370 Hz in ω_2 (^{1}H dimension). A total of 120 scans were averaged for each 2K complex FID, giving a digital resolution of 1.4 Hz/point in ω_2 (^{1}H dimension). Data were zero-filled to 1K real points and apodized by using sine bells in both dimensions.

(*F*) *TOCSY.* A homonuclear Hartmann–Hann transfer spectrum was obtained using MLEV-17 pulse sequences (Bax & Davis, 1985) and the TPPI method at 40 °C in pH 6.7 buffer. A total of 290 FID's were collected with 4000-Hz sweep width. A total of 80 scans were averaged for each 2K complex FID.

Structure Determination. The volumes of the cross-peaks in the NOESY spectra with mixing times 50, 100, and 200 ms were evaluated using the row/column approximation method (Holak et al., 1987). The differences between volumes evaluated by this method and by the integration method of FTNMR (Hare Research, Inc.) were smaller than the experimental uncertainties in cross-peak volumes. Volumes of symmetry-related cross-peaks were averaged. The experimental uncertainty of each cross-peak volume was estimated from the largest of either (i) half the volume difference of the peaks in the two regions, (ii) 10% of the averaged volume, or (iii) the background noise level.

The homo- and heteronuclear scalar coupling constants were estimated from ^{31}P-decoupled DQF-COSY and ^{31}P–^{1}H heteronuclear COSY spectra, as described by Verani et al. (1991).

Molecular dynamics and energy minimization calculations were done on a Silicon Graphics IRIS 4D or on a CONVEX computer using X-PLOR version 3.0 (Nilges et al., 1991; Brünger, 1990). Three-dimensional structures generated were displayed on the SGI using Insight II version 2.0.0 (Biosym Technologies, Inc.).

X-PLOR was run as follows for two-spin approximation calculations: (1) initial energy minimization (50 steps), (2) molecular dynamics for 2 ps at T = 2000 K (time step = 1 fs), (3) molecular dynamics for 1.5 ps to cool the system to 300 K (time step = 0.5 fs), and (4) final energy minimization, with the 4-fold rotational symmetry of the molecule forced (200 steps).

The two-spin approximation structures were refined further using the relaxation matrix routine in X-PLOR (Nilges et al., 1991). During molecular dynamics and energy minimization, the following relaxation potential energy term was used:

$$E_{\text{relax}} = K_{\text{R}} \sum_{\tau} \sum_{i} (\text{well})^2 [I_{i,\tau}(\text{calcd}), I_{i,\tau}(\text{obsd}), \Delta_{i,\tau}, n]$$

where K_{R} is the energy constant for the relaxation energy term and the sum runs for all spin pairs, i, and for three spectra with mixing times 50, 100, and 200 ms. $I_{i,\tau}$(obsd) is the estimated NOE cross-peak intensity of the ith spin pair in the spectra with τ-ms mixing time that has an uncertainty of $\Delta_{i,\tau}$. The well function was defined as

$$\text{well}(a,b,\Delta,n) = (b-\Delta)^n - a^n \qquad \text{if } a^n \le (b-\Delta)^n$$

$$\text{well}(a,b,\Delta,n) = 0 \qquad \text{if } (b-\Delta)^n \le a^n \le (b+\Delta)^n$$

$$\text{well}(a,b,\Delta,n) = a^n - (b+\Delta)^n \qquad \text{if } a^n \ge (b+\Delta)^n$$

where $n = {}^1/_6$. The input constraints were the same as those used in the two-spin approximation calculation, except that the NOE distance constraints were replaced by the NOE intensity constraints. An isotropic correlation time 3.0 ns was assumed for the molecule, but changes in correlation time made little difference (Gochin et al., 1990). All NOE cross-peak intensities were used for the calculation of the calibration factor between the observed and the calculated NOE intensities.

The protocol for the relaxation matrix refinement was the same as that used for the two-spin approximation, except that the molecular dynamics simulation was done at 600 K for 1 ps.

RESULTS

UG_4U Is Unusually Stable. Figure 1a shows the downfield spectrum of UG_4U at 50 °C in H_2O. There are 6 resonances in its aromatic region around 8 ppm, 4 singlets and 2 doublets. The singlets are guanine H8 resonances, and the doublets are uracil H6 resonances. The presence of one aromatic resonance in the molecule's spectrum for each base in its sequence shows that UG_4U either is monomeric under the conditions in question or is a multistranded structure in which the strands are equivalent.

The presence of imino proton resonances around 11 ppm in the molecule's spectrum indicates that UG_4U is hydrogen bonded under these conditions. Below 25 °C, a number of broad resonances become evident at around 9 ppm, which represent the amino protons of the molecule's G residues, and a broad imino proton resonance becomes evident at 10.7 ppm (data not shown). Since it is most unlikely that a single-

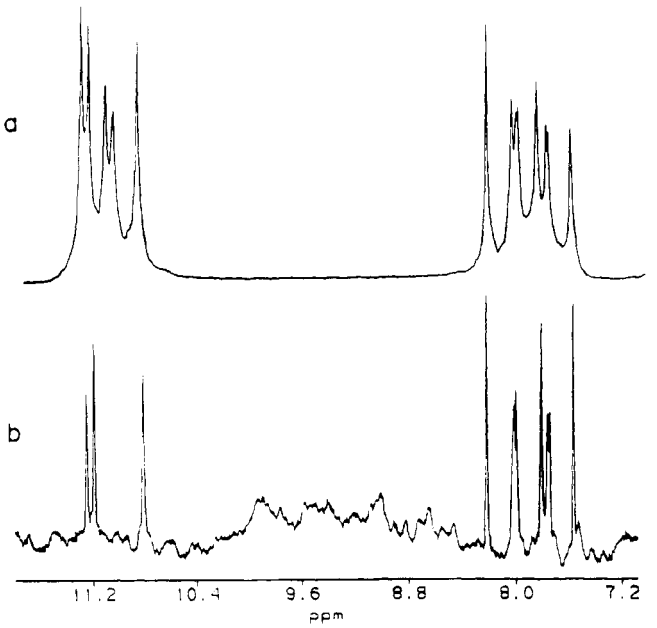

FIGURE 1: Downfield spectra of UG_4U. Spectrum a is that of UG_4U taken at 50 °C in H_2O in pH 6.7 buffer (800 scans) (see Materials and Methods). Spectrum b is that of the same molecule in the same buffer at 40 °C, but the solvent is D_2O, not H_2O. Spectrum b was taken after more than 24 h of incubation in D_2O at 40 °C. It is noisy because the number of FID's averaged was very small (80 scans).

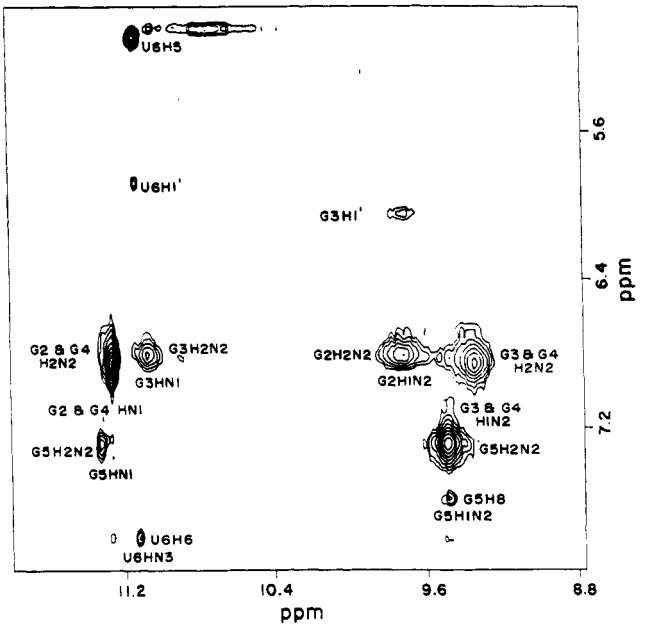

FIGURE 2: H_2O NOESY of UG_4U. A NOESY spectrum was taken of UG_4U in H_2O in pH 5.2 buffer at 5 °C (see Materials and Methods). The section of the spectrum shown is roughly its imino–aromatic region. The identifications of the resonances contributing to each column that includes a cross-peak are indicated underneath the most downfield of the peaks that appear in that column. The resonances contributing in the row dimension are specified beside each cross-peak.

stranded stem–loop of only 6 nucleotides would show 5 of its 6 imino protons, UG_4U must be multistranded.

The imino proton spectrum of UG_4U is the same from room temperature all the way up to 70 °C, showing that its hydrogen bonds are unusually stable. A further indication of its stability is the observation that 3 of its imino proton resonances can be detected in a sample that has been exchanged into D_2O several times by lyophilization from D_2O and incubated for 24 h in D_2O at 40 °C. As shown in Figure 1b, these imino peaks are as strong as the nonexchangeable aromatic proton peaks around 8 ppm. (The high noise levels in this spectrum reflect the small number of FID's averaged to product it.) The imino protons in question, which are those of G3, G4, and G5 (see below), require several days at 40 °C to equilibrate with solvent. We know of no Watson–Crick duplex as stable as this.

UG_4U Is Hoogsteen Base-Paired. NOESY spectra taken in H_2O reveal that 4 of the molecule's 5 imino proton resonances give NOE's to 2 amino proton resonances each, one around 9.5 ppm and a second around 6.8 ppm. The upfield connectivities are shown in Figure 2. Amino protons normally resonate around 6.5–7.5 ppm, and they are shifted downfield by ≈2 ppm when they are hydrogen bonded (Varani et al., 1991). Thus there are strong reasons for believing that the G–G interactions in the middle of the UG_4U structure involve hydrogen-bonded N2 amino protons.

Another clue as to the nature of the G–G interactions in UG_4U comes from the finding that the hydrogen-bonded amino proton of G5 (at 9.47 ppm) gives an NOE to what is apparently its own H8 (Figure 2). This cannot be an intrabase NOE because the distance from HN2 to H8 is too long (≈7.44 Å); it must be an interstrand NOE. This suggests that the N2 of at least one G must make hydrogen bonds to the N7 of its neighbor in the next strand, which in turn requires that the imino proton of the same base make a hydrogen bond to the O6 of that same neighbor. In short, the G's at the 5 position (G5), at least, must be Hoogsteen paired.

UG_4U Is a Tetraplex. The two G's in a single Hoogsteen pair are not equivalent, but they are equivalent in structures that contain rings of four Hoogsteen-paired G residues, each accepting hydrogen bonds from, and donating hydrogen bonds to, its neighbors on adjacent strands, as shown in Figure 3a. If the ring had only 3 G residues, the geometry would allow hydrogen bonds only through the imino protons but not through the amino protons (Zimmerman, 1975), which contradicts the NMR data. If the ring had more than 4 G residues, the hydrogen bond angles and distances would be less favorable than those in a quartet ring. Although the NMR data alone do not rule out the possibility of a structure with more than 4 strands, it is very likely that UG_4U is 4-stranded, which is the conclusion suggested by our gel results (Kim et al., 1991).

UG_4U Is Parallel Stranded. As shown in Figure 4, there are only two ways strands can associate in a tetrastranded complex that maintain strand equivalence. They can form a parallel-stranded structure that has a 4-fold rotational symmetry along its long axis. Alternatively, they can form a structure built of two pairs of antiparallel strands where the structure has a 2-fold rotational symmetry along its long axis and another 2-fold rotational symmetry perpendicular to its long axis.

If the UG_4U structure were antiparallel, the number of imino proton resonances the structure gives would have to be even, probably 4 or 6. Thus, the presence of 5 downfield resonances in the UG_4U spectrum indicates it is a parallel-stranded tetraplex.

UG_4U Is Stabilized by Novel U–U Interactions. Since the strands of the UG_4U complex are all equivalent, the only way a fifth imino proton resonance can appear is if the U residues at one end of the molecule hydrogen bond with each other. NOESY spectra taken in H_2O show that the fifth imino proton peak is that of the 3′-terminal U (see below). As mentioned earlier, an additional, weak imino proton resonance is evident at about 10.8 ppm at temperatures below 25 °C. It is not clear whether this peak is due to the formation of a looser,

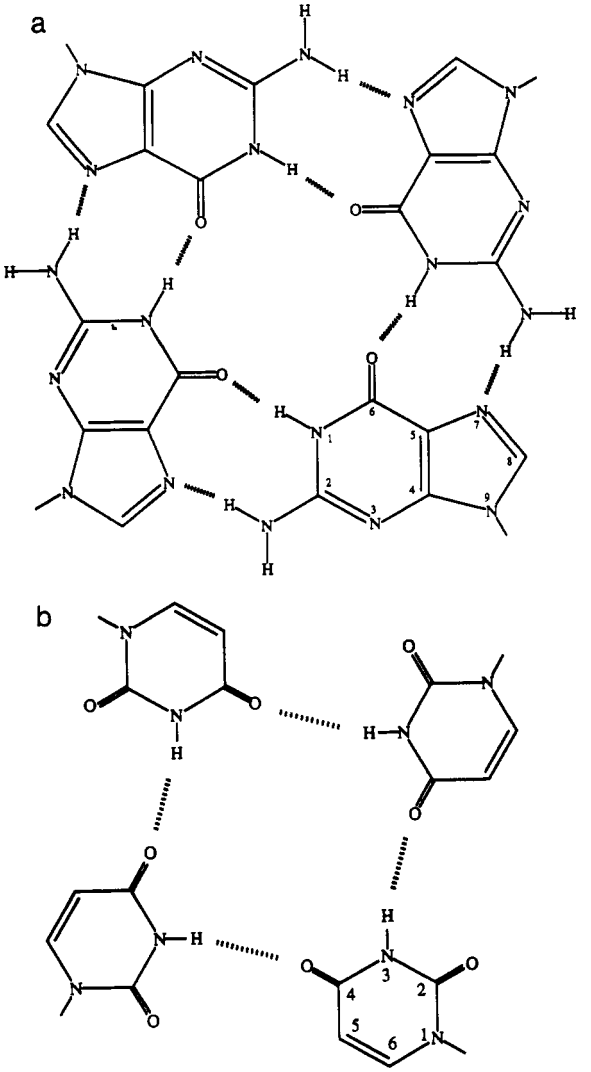

FIGURE 3: Base pairing in UG_4U. Panel a shows the base pairing expected in a Hoogsteen-paired G quartet. Panel b shows the hydrogen-bonding pattern proposed for the U quartets at both ends of UG_4U.

hydrogen-bonded U structure at the 5′ end like the one at the 3′ end, or whether it is due only to the protection of the 5′ U imino proton from exchange with solvent by base stacking.

The imino proton of U6, the one that is strongly hydrogen bonded, gives NOE's to its own H1′, H5, H6 resonances (Figure 2). The weak NOE from the imino proton to H1′ might be an intranucleotide NOE (≈4.49 Å). But the NOE's from the imino proton to H5 and H6 must be internucleotide; the NOE to H5 is too strong to be intranucleotide (≈4.22 Å), and intranucleotide distance from the imino proton to H6 is out of NOE range (>5 Å). NOE's of this kind are not seen in AU base pairs. Thus these NOE's are indicative of the presence of N3–O4 hydrogen bonds of the kind shown schematically in Figure 3b.

Assignments: *Exchangeable Proton Resonances.* All exchangeable and nonexchangeable proton resonances, as well as the ^{31}P resonances of $(UG_4U)_4$, were assigned using NOESY experiments and several different kinds of homo- and heteronuclear COSY experiments. The assignment strategy used has been described by Wüthrich (1986), Cheong (1989), and Varani et al. (1991).

The imino proton resonances of $(UG_4U)_4$ are difficult to assign because they are so close together in chemical shift. Nevertheless, all imino and amino proton resonances could be assigned by NOESY in H_2O at pH 5.2 except for an ambiguity

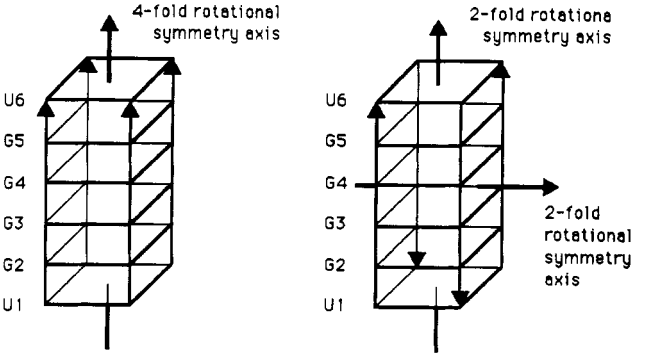

FIGURE 4: Possible strand alignments in UG_4U. The two ways in which strands can be arranged in UG_4U consistent with the condition that all strands be equivalent are shown. The symmetry axes of both structures are indicated.

involving the imino and the amino protons of G3 and G4, which could not be distinguished. The chemical shifts of exchangeable proton resonances are reported in Table I.

Two of the five strong imino proton peaks in the $(UG_4U)_4$ spectrum are not seen in D_2O. As shown in Figure 2, one of them (11.12 ppm) gives a strong NOE to the H5 of U6 and weaker NOE's to its H6 and H1′. [The assignments of these nonexchangeable proton resonances were established independently (see below).] Furthermore, this imino proton resonance does not give strong NOE's to amino proton resonances (around 9.5 ppm), unlike the other imino proton resonances. For these reasons, the 11.12 ppm resonance was assigned to U6's imino proton. The broad peak (10.78 ppm) which disappears above 25 °C was assigned to U1's imino proton. The remaining 4 imino proton resonances, all of which give NOE's to pairs of amino proton resonances, were assigned to G quartets. Resonances around 9.5 ppm were assigned to H-bonded G amino proton resonances, and resonances around 6.8 ppm were assigned to non-H-bonded G amino proton resonances (Varani et al., 1991).

Once the imino, H-bonded amino, and non-H-bonded amino proton resonances belonging to each G quartet were grouped together, internucleotide NOE's provided the clues necessary to identify each quartet. The NOE cross-peak between U6's imino proton and G5's H-bonded amino proton identified G5 quartet spins (1 imino and 2 amino protons). The second imino proton resonance (11.06 ppm), which disappears in D_2O, was identified as that of G2 because its amino proton gives an NOE to the H8 of G2. The remaining two imino proton resonances, which are those of G3 and G4, give NOE's to each other but not to other bases, which is why they cannot be distinguished.

Assignments: *Nonexchangeable Proton Resonances.* NOESY, DQF-COSY, TOCSY, and ^1H–^{31}P heteronuclear COSY experiments were done in pH 6.7 D_2O buffer at 40 °C to assign nonexchangeable proton resonances and ^{31}P resonances. Additional NOESY experiments were done at different temperatures (30 and 5 °C) and pH's (pH 5.2). Virtually, no change in the chemical shifts of nonexchangeable proton resonances was observed within ±0.05 ppm.

Figure 5 shows the aromatic–anomeric region of the NOESY spectrum of UG_4U. Its sequential assignment was straightforward. All the sugar proton spins in each sugar systems were identified from scalar couplings in DQF-COSY and TOCSY spectra. These assignments were confirmed by the detection of intranucleotide NOE's from H8/H6 to sugar protons and by intranucleotide NOE's among sugar protons. The identification of each nucleotide spin systems was achieved using several independent, internucleotide sequential NOE

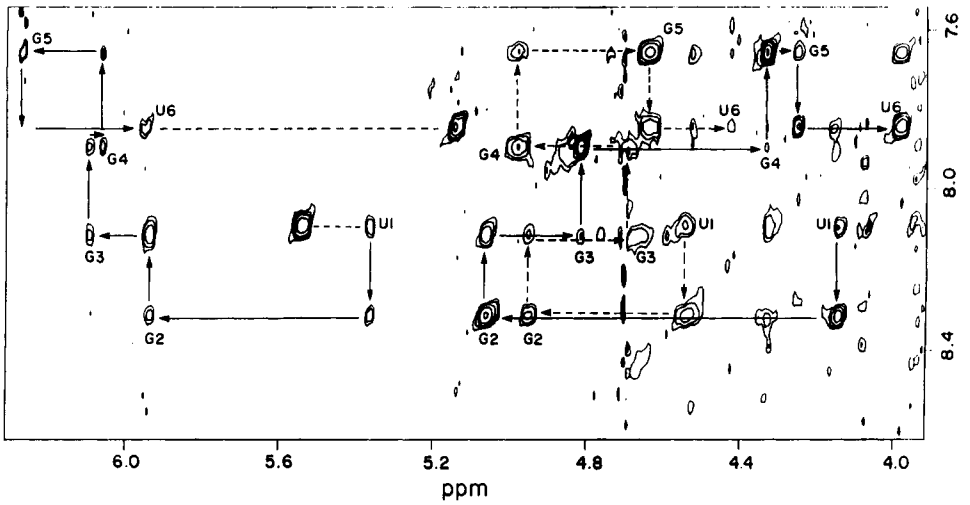

FIGURE 5: D_2O NOESY spectrum of UG_4U. The NOESY spectrum shown was taken in pH 6.7 buffer at 40 °C. The region shown is the aromatic–anomeric/2′,3′ region. Aromatic–anomeric (solid arrows), aromatic-2′ (solid arrows), and aromatic-3′ (broken arrows) connectivities are traced out. Only intranucleotide cross-peaks are labeled.

Table I: Chemical Shifts of Proton[a] and Phosphorus[b] Resonances for $(rUG_4U)_4$ in ppm

	exchangeable protons[c]			nonexchangeable protons[d]							
	imino	amino[e]	amino[f]	H5	H6/H8	H1′	H2′	H3′	H4′	H5′/H5″[g]	$^{31}P^d$
U1	10.78	na[h]	na	5.46	8.01	5.29	4.07	4.47	4.23	4.00/3.87	
											−4.68
G2	11.06	9.72	6.80	na	8.23	5.85	4.98	4.86	4.58	4.43/4.26	
											−3.32
G3[i]	10.88	9.36	6.83	na	8.03	6.01	4.74	4.61	4.69	4.55/4.50[j]	
											−4.28
G4[i]	11.26	9.34	6.85	na	7.82	5.96	4.25	4.90	4.56	4.80/4.22	
											−4.10
G5	11.31	9.47	7.30	na	7.58	6.19	4.17	4.56	4.44	4.66/4.15	
											−3.17
U6	11.12	na	na	5.06	7.76	5.87	3.91	4.35	4.25	4.08[j]	

[a] Relative to dioxane (3.741 ppm). [b] Relative to TMP (0.0 ppm). [c] Measured in 18 mM phosphate, pH 5.2, 88 mM Kcl, and 0.9 mM EDTA at 5 °C. [d] Measured in 10 mM phosphate, pH 6.7, 50 mM KCl, and 0.5 mM EDTA at 40 °C. [e] Hydrogen-bonded amino protons (H1N2). [f] Non-hydrogen-bonded amino protons (H2N2). [g] H5′ and H5″ proton resonances are not stereospecifically assigned. [h] Not applicable. [i] The exchangeable proton resonances of G3 and G4 could be switched (see the text). [j] Assignments are not certain.

Table II: Homo- and Heteronuclear Scalar Coupling Constants in Hz for $(rUG_4U)_4$ at 40 °C in 10 mM Phosphate, pH 6.7, 50 mM KCl, and 0.5 mM EDTA[a]

	1′2′	2′3′	3′4′	4′5′/4′5″	5′5″	3′P	P5′/P5″	P4′[b]
U1	6	5	8.5	4/4.5	12	11.5	na[c]	na
G2	8	8 ± 3	4	<3	10	9	14.5 ± 3	<3
G3	3	on water	11 ± 3	<3	diagonal	7	2	<3
G4	<2	4.5	10	<3	10	9.5	6.5	7
G5	<2	4.5	10	5	10	9	4.5	2
U6	7.5	5	3	(<3)[d]	(diagonal)[d]	na		5

[a] Uncertainties are ±2 Hz, unless indicated. [b] Four-bond coupling. [c] Not applicable. [d] Not certain.

pathways, [H8/H6(i) → H1′(i), H2′(i), or H3′(i) → H8/H6(i + 1)] and [H1′(i) → H2′(i) → H1′(i + 1)]. These assignments were further confirmed by heteronuclear, sequential scalar couplings [H3′(i) → P(i + 1) → H5′(i + 1), H5″(i + 1), or H4′(i + 1)]. ^1H–^{31}P scalar couplings also provided a way to assign phosphorus resonances and some H5′ and H5″ resonances.

Additional evidence for the parallel strandedness of the UG_4U complex comes from the base orientations with respect to their sugars. The H8/H6 to H1′ NOE cross-peaks of the UG_4U are all much weaker than the U1 and U6 H5/H6 cross-peaks (Figure 5), which is the characteristic of bases that are in anti conformation (Wüthrich, 1986). Since it is impossible to build an antiparallel G quartet structure that does not have alternating syn and anti conformations in each G plane (Williamson et al., 1989; Sundquist & Klug, 1989), again we conclude that UG_4U must be forming a parallel-stranded structure.

UG_4U has only 44 nonexchangeable protons and 5 phosphorus atoms. All of their resonances were identified except one H5″ resonance, which could not be resolved due to spectral overlap to H5′ resonance of the same sugar (U6). It was not possible to assign ribose H5′ and H5″ protons stereospecifically. Chemical shifts are reported in Table I.

Analysis of NMR Data. (*A*) *Base Stacking.* The aromatic–anomeric sequential NOE connectivity (Figure 5) shows that the bases are well stacked from U1 down to U6, but the stacking between G5 and U6 is unusual. An NOE from G5 H8 to U6 H2′ is observed, but the expected NOE from G5 H1′ to U6 H6 is not (Figure 5).

(*B*) *Backbone Torsion Angles.* The scalar coupling constants of UG_4U were measured from homo- and heteronu-

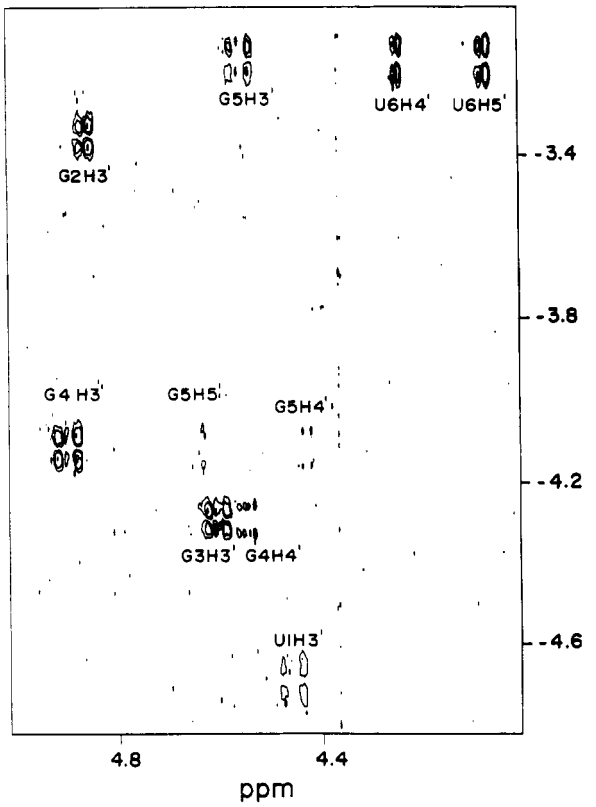

FIGURE 6: ^{31}P–^{1}H heteronuclear COSY spectrum of UG_4U. A ^{31}P–^{1}H heteronuclear COSY spectrum was collected on UG_4U at 40 °C as described in Materials and Methods. Assignments are provided for all cross-peaks. Note that only 5 phosphorous resonances are expected because the molecule examined had no terminal phosphates.

clear correlated spectra. They are reported in Table II. Backbone torsion angles β, γ, and ϵ were estimated from these scalar couplings using the semiempirical Karplus equation (Altona, 1982). β torsion angles were estimated from ^{31}P–H5′/H5″ scalar coupling constants, $J_{P5'}$ and $J_{P5''}$. γ torsion angles were estimated from $J_{4'5'}$ and $J_{4'5''}$, and ϵ torsion angles, from $J_{3'P}$. Several ranges of angles had to be considered in each case because any value of a coupling constant corresponds to several angles in the Karplus equation, and stereospecific assignments of H5′ and H5″ were not possible. For G4, G5, and U6, four-bond scalar couplings $J_{P4'}$ (≈2–7 Hz) were observed (see Figure 6 and Table II). This is possible only when all the atoms in the coupling path (P–O5′–C5′–C4′–H4′) are in all-trans planar "W"-type geometry (Sarma et al., 1973), or β is trans and γ is gauche$^-$.

Sugar ring conformations of UG_4U were obtained from $J_{1'2'}$ scalar coupling constants (de Leeuw & Altona, 1982). The sugars of G4 and G5 are ≥85% N-type or C3′-endo, which is characteristic of A-form RNA. The sugar of G3 is close to $^3T_4$. Its pseudorotation phase angle P is ≈30 ± 15°, which is not far away from C3′-endo ($P \approx 18°$). The U1 ribose seems to be in dynamic exchange between C2′-endo and C3′-endo (≈30% N-type). A small U1 H1′–H2′ cross-peak is visible in DQF-COSY spectra, and the H3′–H5′ coupling is smaller than expected for a ribose in the C3′-endo conformation. The ribose conformations of G2 and U6 are ≈100% S-type, however, which is most unusual for RNA.

^{31}P chemical shifts are sensitive to the backbone angles α and ζ; ^{31}P resonates downfield when ζ/α are not in the usual guache$^-$/gauche$^-$ conformation (Gorenstein, 1984; Roongta et al., 1990). As shown in Table I, ^{31}P chemical shifts of UG_4U range between ≈–4.7 and ≈–3.1 ppm. Only one pair of backbone angles ζ(U1)/α(G2) whose ^{31}P resonates at –4.68 ppm were assigned to gauche$^-$/gauche$^-$ conformation.

Model Building. The three-dimensional model of $(UG_4U)_4$ described below was obtained as follows. A model for $(rG_6)_4$ was constructed using the structure proposed for poly(rG) on the basis of fiber diffraction data (Zimmerman, 1976). In this model, G quartets are planar, and the geometry of the quartets was held planar throughout the model building process. The G quartets at the ends of the structure were replaced by U's using Insight II, and the G2 and U6 ribose conformations were changed to C2′-endo using X-PLOR, keeping all other parts of the structure unchanged. Finally, the model was refined by molecular dynamics and energy minimization using X-PLOR, first by carrying out a two-spin approximation calculation and then by using a relaxation matrix calculation that took into account all of the NMR data.

A total of 428 distance constraints (≈18 distance constraints/nucleotide) were used: 212 derived from intranucleotide NOEs, 112 from internucleotide NOEs, 36 from hydrogen bonds, 8 from exchangeable proton NOEs, and 60 derived from backbone torsion angle information. Exchangeable proton NOE distance constraints were used only qualitatively during the two-spin phase of model building. Pairs of exchangeable protons for which an NOE cross-peak was seen were assigned distances of 3.5 ± 1.5 Å. In addition, 104 backbone and glycosidic torsion angles, sugar ring torsion angles to hold G2 and U6 in the C2′-endo conformation, and some additional dihedral angles to make quartet planes planar were entered as dihedral constraints.

A total of 8 structures were generated using different seeds during the refinement. The final structures were identical to each other within the experimental errors except some backbone angles connecting G5 and U6 (see below); their rms deviations were 0.71 ± 0.16 Å for all atoms.

Assessment of the Final Model. The model shown in Figure 7 was tested in several ways. First, initial models were obtained by distance geometry methods, rather than the model building method just described. The distance geometry routine in XPLOR was used to compute conformations for a single strand of the tetraplex using intrastrand NOE distance constraints as the input data. The single-strand models that resulted were then replicated in a 4-fold manner to generate tetraplexes, which were refined in the usual way. The models that resulted were indistinguishable from the one shown in Figure 7. Second, a starting model was obtained as described in the previous section with its U6 residues in the B conformation. Again the refined models that resulted were the same as shown in Figure 7. Finally, a range of correlation times were tested as inputs for the full matrix relaxation phase of refinement. Correlation times of 1, 3, and 5 ns all gave the same final result. When the correlation time was set at 30 ns, significant variations began to appear, and when the correlation time was set at 0.3 ns, a quite different result was obtained. (The correlation time of a rigid sphere whose diameter is the same as the height of the molecule or its diameter has a correlation time of about 2.8 ns.) Thus the final model is quite insensitive to the correlation time chosen during final refinement.

The quality of structures obtained from NMR data should be gauged by two different criteria: (i) faithfulness to the experimental data and (ii) fidelity with respect to previously

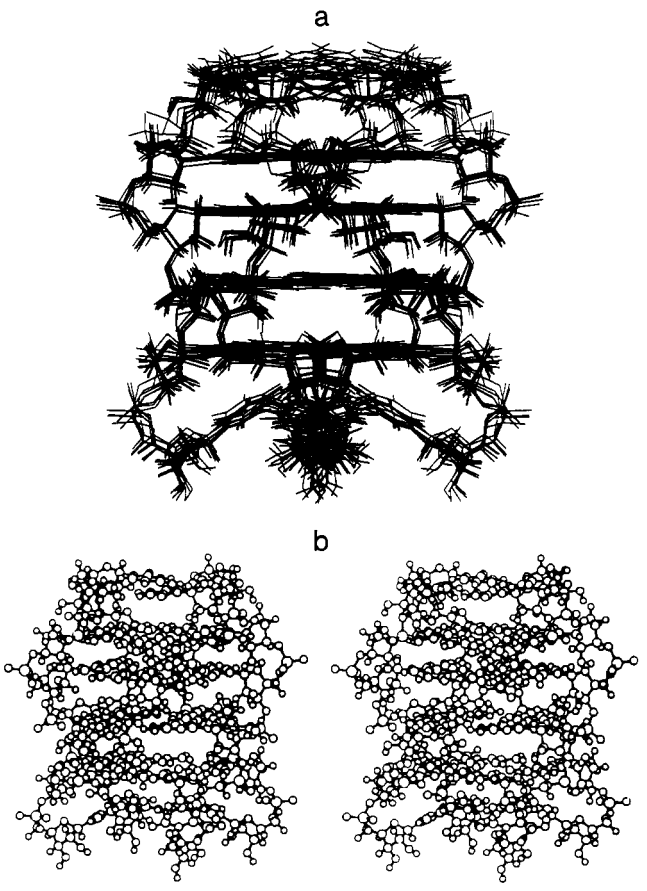

FIGURE 7: Fully refined model for UG_4U. In (a), 8 final structures are superimposed, and in (b), one of them is shown as a stereopair in a ball-and-stick representation.

established empirical data. The first can be addressed using a residual, or *R*-factor:

$$R^n = \sum_\tau \sum_i (\text{well})[I_{i,\tau}(\text{calcd}), I_{i,\tau}(\text{obsd}), \Delta_{i,\tau}, n] / \sum_i I_i^n(\text{obsd})$$

(see Materials and Methods for the definition of the well function). When $n = 1$, this *R*-factor corresponds to the well-known crystallographic *R*-factor. When $n = {}^1/_6$, *R* corresponds to the residual suggested for NMR structures by James and co-workers (Borgias et al., 1990). The consistency of the structure with the empirical data used for model building can be judged on the basis of the deviations from ideality of its bond lengths, bond angles, and base planarities.

R-factors and deviations from ideality of the starting X-ray fiber diffraction structure, of the two-spin approximation structures, and of the final refined structures are reported in Table III. The *R*-factors of the final structures are excellent when compared to those of NMR structures of other proteins or DNAs [e.g., see Nilges et al. (1991) and Gochin and James (1990)]. The deviations from ideality for the final structures are also satisfactory.

Structural Features of $(UG_4U)_4$. The final 8 structures are shown superimposed in Figure 7a, and a ball-and-stick model of one of the structures is also shown in stereo in Figure 7b. The overall structure is very compact. The 3′-end U-quartet plane is buckled, but the conformation of this part of the molecule is well-defined by the data since the superposition is very good in this part of the model. Indeed, the conformation of this part of the molecule is fixed by several internucleotide NOE's: [G5 H8 to U6 H5], [G5 H8 to U6 H2′], [G5 H2′ to U6 H5], and [G5 H2′ to U6 H6]. Furthermore, the G5H1′–U6H6 NOE that one normally expects to see in RNAs is not observed. Thus the conformation of this part of the molecule is quite unusual. The U's in that plane are stabilized by hydrogen-bonding and stacking interactions with G5 bases.

The 5′-end U1 nucleotides are less well constrained by the data. Since the imino proton resonance of this uracil is evident only when the temperature is below 25 °C (data not shown), and its chemical shift is downfield of free imino proton resonances, it is likely that a quartet structure similar to the one at the 3′ end forms at the 5′ end, below room temperature.

Average backbone torsion angles of the final structures are reported in Table IV. Most of them are close to backbone angles characteristic of A-form RNA. Non-A-like backbone angles are localized in G2pG3 and G5pU6, and they are not artifacts of molecular dynamics or energy minimization. The NMR data strongly indicate that there are unusual features in G2pG3 and G5pU6; the two most down-shifted ^{31}P resonances are of G2pG3 and G5pU6 (Table I), $J_{1'2'}$ scalar couplings of G2 and U6 are large (7.5–8 Hz), and there are unusual intranucleotide NOE's between G5 and U6 (see above). It is interesting to note that, of the five quartets (G2, G3, G4, G5, and U6), the two end quartets (G2 and U6) have unusual C2′-endo sugar conformation, and the first and the last ^{31}P's chemical shifts are downshifted as much as ≈1–1.5 ppm.

Although the position of the U6 base is well-defined, the backbone connecting G5 and U6 is not. Three sets of backbone angles in G5pU6 (ϵ, ζ of G5, and α, β of U6) are observed in the final structures: (i) $\epsilon^-\zeta^-\alpha^-\beta^t$, (ii) $\epsilon^+\zeta^+\alpha^+\beta^-$, and (iii) $\epsilon^t\zeta^-\alpha^t\beta^-$. A $\beta^t\gamma^+$ constraint was imposed on U6 during refinement because of the four-bond scalar coupling $J_{P4'}$ observed for that nucleotide. Nevertheless, the final structures in groups ii and iii do not satisfy it. Therefore, the U6 β backbone angle is not as well-defined in our models as it was by our NMR data. All other experimental backbone angle constraints are satisfied in all final structures.

The ϵ^+ conformation which is seen at G5 in group ii structures is supposed to be forbidden for nucleic acids due to steric interactions involving C2′ and 3′-phosphate (Sundaralingam, 1969; Altona, 1982). But in this unusual backbone context ($\zeta^+\alpha^+$), no steric hinderance was observed and the conformational energy of this group was as low as that of other groups.

Although a $\zeta^-\alpha^-$ constraint was imposed only on U1pG2, G3pG4 and G4pG5 were also in $\zeta^-\alpha^-$ conformation. But G2pG3 has an unusual $\zeta^-\alpha^+$ conformation, and G5pU6 is in a $\zeta^+\alpha^+$ or $\zeta^-\alpha^t$ conformation. β, γ, and ϵ torsion angles are in the usual $\beta^t\gamma^+\epsilon^t$ conformation for all residues except U6 (some in β^- conformation), G3 (all in γ^t conformation), and G5 (some in ϵ^- and some in ϵ^+). The glycosidic conformation is anti for all residues except U6, which is in a high anti conformation.

The spacings between quartet planes are not uniform, as shown in Figure 7a. The spacings between G2 and G3, and between G4 and G5, are smaller (≈3.6 ± 0.3 Å) than those between G3 and G4 (≈4.4 ± 0.2 Å). This feature of the structure is not an artifact of refinement; a number of internucleotide NOE's in this part of the molecule have intensities that differ by as much as a factor of 10 compared to their counterparts in other parts of the molecule.

DISCUSSION

Despite the fact that RNA plays important roles in many aspects of gene expression, only a handful RNA structures are known at atomic resolution. tRNAs [see Saenger (1984)

Table III: Average R-Factors and Deviations from Ideality

	$R^{1/6}$ [a]	R[a]	Δ(bond) (Å)	Δ(angle) (deg)	Δ(impr)[b] (deg)
X-ray fiber	0.108	0.62	0.022	3.96	0.26
two-spin	0.051 ± 0.001	0.40 ± 0.01	0.027 ± 0.001	5.84 ± 0.11	1.50 ± 0.03
relaxation	0.036 ± 0.001	0.16 ± 0.01	0.009 ± 0.001	3.64 ± 0.07	0.54 ± 0.09

[a] When a parabolic potential was used, $R^{1/6} = 0.164$ and $R = 0.78$ for the initial X-ray fiber-derived structure; $R^{1/6} = 0.096 \pm 0.002$ and $R = 0.55 \pm 0.01$ for two-spin structures; and $R^{1/6} = 0.086 \pm 0.002$ and $R = 0.32 \pm 0.01$ for relaxation structures. [b] An improper angle is a dihedral angle defined in X-PLOR that determines base planarity (Brünger, 1990; Nilges et al., 1991).

Table IV: Backbone Torsion Angles of Final Structures[a]

	α (deg)	β (deg)	γ (deg)	δ (deg)	ε (deg)	ζ (deg)	χ (deg)
U1				99 ± 13	−152 ± 5	−75 ± 7	−153 ± 4
G2	−66 ± 6	164 ± 2	67 ± 3	**145 ± 6**	−164 ± 11	**−166 ± 3**	−122 ± 7
G3	**120 ± 11**	−160 ± 8	**176 ± 1**	87 ± 3	−166 ± 7	−80 ± 2	−170 ± 2
G4	−81 ± 6	−175 ± 8	63 ± 1	92 ± 5	−166 ± 5	−76 ± 5	−159 ± 4
G5[b]	−78 ± 10	−177 ± 6	62 ± 1	101 ± 6	**−109 ± 1**	−92 ± 6	−155 ± 6
					69 ± 1	**89 ± 2**	
					−171 ± 9	**−49 ± 9**	
U6[b]	**−48 ± 6**	136 ± 5	42 ± 2	**141 ± 6**			**−73 ± 1**
	95 ± 2	**−98 ± 4**	77 ± 2				**−47 ± 2**
	−178 ± 1	**−111 ± 1**	67 ± 3				**−51 ± 1**
A-RNA[c]	−68	178	54	84	−153	−71	−158

[a] Non-A-like backbone angles are reported in boldface. Error limits are standard deviations of final structures. [b] Three sets of angles are reported for G5 and U6 because three different solutions for the structure emerged from the analysis (see text). [c] A-form values from fiber diffraction studies (Saenger, 1984).

and references therein], a Watson–Crick duplex (Dock-Bregeon et al., 1988), and an RNA duplex involving G-C and G-U mismatches (Holbrook et al., 1991) have been solved by X-ray crystallography. In addition, a symmetric 6 bp RNA duplex (Happ et al., 1988) and two RNA hairpin structures have been solved by 2-D NMR methods (Cheong et al., 1990; Heus & Pardi, 1991). The two RNA hairpins include the most frequently occurring tetraloop sequences in biology (Uhlenbeck, 1990) and seem to have considerable importance (Tuerk et al., 1988; Uhlenbeck, 1990). Here, we provide a high-resolution structure for an RNA molecule that is phenomenal in its thermal stability and very unusual in its strand association. The size of the molecule is 24 nucleotides (over 13 kDa), but its NMR spectrum is much simpler due to its 4-fold symmetry.

Although we do not know of any biological example of an RNA quartet structure, it would not be surprising if they were found in the future, given their remarkable stability. The formation of a parallel-stranded RNA tetraplex structure might occur in the last stages of the folding of a large RNA, where 4 strands having short G runs were brought close together. The formation of quartet structures would tend to lock any structure that contained them into place. Preliminary gel mobility shift experiments show that there may be a protein in yeast cell extracts that can recognize $(rUG_4U)_4$ (Joon Kim, personal communication).

The stability of the tetraplex structure comes from the fact that each base makes 4 hydrogen bonds to its neighbors whereas there are only 2 or 3 hydrogen bonds per base in a normal Watson–Crick base pair. Another contribution to stability undoubtedly comes from favorable interactions between monovalent cations and guanines in the middle of each quartet. There are good reasons to believe that monovalent cations are important in the tetraplex formation; certain types of cations are required for tetraplex formation [e.g., see Sen and Gilbert (1990) and Kim et al. (1991)]. Furthermore, chemical shift and line width analyses of ^{23}Na spectra suggest that the cations bind to 5′-GMP when the nucleotide concentration exceeds a critical point, but not to 5′-AMP, 5′-ATP, or 2′-GMP (Borzo et al., 1980).

A feature of the model that puzzled us for a long time was the unevenness of the spacings between adjacent quartet planes. While this paper was in press, a crystallographic structure was published of an antiparallel DNA G quartet that may explain why our model looks the way it does [Kang et al., 1992; see also Smith and Feigon (1992)]. The DNA structure, which has 4 G quartets, includes just one potassium ion, which is located between the second and third quartets. Its presence makes these quartets buckle apart. It is likely that our molecule too has a single, tightly bound potassium ion and that its presence in the space between the G3 and G4 planes accounts for their wide spacing. When we refined our model with the constraints on G quartet geometry released, its residuals did not improve significantly, and the geometries of the quartets did not change in a useful way. It is clear that we would have to determine our NMR data with higher accuracy that we have been able to so far if we want to define the quartet geometries better.

The existence of U quartets in UG_4U suggests that four-stranded poly(U) structures might be possible, but they have not been seen, as far as we know. A four-stranded poly(U) helix would have short distances between the phosphate groups of its strands, and the two hydrogen bonds per base that would form might not be enough to hold them together. It is probably wiser to think of the U quartets in $(UG_4U)_4$ as a "finishing" structure, which is to say a hydrogen-bonded structure that forms because it is a low-energy structure in the special context provided at the end of a tetra-G stack, not because it is an intrinsically stable structure.

ACKNOWLEDGMENT

We thank Joseph Kim for helpful discussions, Michael Nilges for his help with X-PLOR, Gregory Kellogg for writing some NMR programs, and Kenneth Harris for purifying RNA.

REFERENCES

Altona, C. (1982) *Recl. Trav. Chim. Pays-Bas 101*, 413–433.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Borgias, B. A., Gochin, M., Kerwood, D. J., & James, T. L. (1990) *Prog. Nucl. Magn. Reson. Spectrosc. 22*, 83–100.

Borzo, M., Detellier, C., Laszlo, P., & Paris, A. (1980) *J. Am. Chem. Soc. 102*, 1124–1134.

Brünger, A. T. *X-PLOR Version 3.0, User Manual*, Yale University, New Haven, CT, (in preparation).

Cheong, C. (1989) Ph.D. Thesis, University of California, Berkeley.

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

De Leeuw, F. A. A. M., & Altona, C. (1982) *J. Chem. Soc., Perkin Trans. 2*, 375–384.

Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Moras, D., deBear, J. S., Gough, G. R., Gilham, P. T., & Johnson, J. E. (1988) *Nature 335*, 375–378.

Gochin, M., & James, T. L. (1990) *Biochemistry 29*, 11172–11180.

Gochin, M., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 11161–11171.

Gorenstein, D. G. (1984) *Phosphorus-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) Academic Press, New York.

Happ, C. S., Happ, E., Nilges, M., Gronenborn, A. M., & Clore, G. M. (1988) *Biochemistry 27*, 1735–1743.

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

Holak, T. A., Scarsdale, J. N., & Prestegrad, J. H. (1987) *J. Magn. Reson. 74*, 546–549.

Holbrook, S. R., Cheong, C., Tinoco, I., Jr., & Kim, S.-H. (1991) *Nature 353*, 579–581.

Hoult, D. I., & Richards, R. E. (1975) *Proc. R. Soc. London, Ser. A 344*, 311.

Kang, C., Zhang, X., Ratliff, R., Moyzis, R., & Rich, A. (1992) *Nature 356*, 126–131.

Kim, J., Cheong, C., & Moore, P. B. (1991) *Nature 351*, 331–332.

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

Nilges, M., Habazettl, J., Brünger, A. T., & Holak, T. A. (1991) *J. Mol. Biol. 219*, 499–510.

Pinnavaia, T. J., Marshall, C. L., Mettler, C. M., Fisk, C. L., Miles, H. T., & Becker, E. D. (1978) *J. Am. Chem. Soc. 100*, 3625–3627.

Roongta, V. A., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245–5258.

Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sarma, R. H., Mynott, R. J., Wood, D. J., & Hruska, F. E. (1973) *J. Am. Chem. Soc. 95*, 6457–6459.

Sen, D., & Gilbert, W. (1988) *Nature 334*, 364–366.

Sen, D., & Gilbert, W. (1990) *Nature 344*, 410–414.

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

Sklenár, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469–479.

Sklenár, V., Miyoshiro, H., Zon, G., & Bax, A. (1986) *FEBS Lett. 208*, 94–98.

Smith, F. W., & Feigon, J. (1992) *Nature 356*, 164–168.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Sundaralingam, M. (1969) *Biopolymers 7*, 821–860.

Sundquist, W. I., & Klug, A. (1989) *Nature 342*, 825–829.

Tuerk, C., et al. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364–1368.

Uhlenbeck, O. C. (1990) *Nature 346*, 613.

Varani, G., Cheong, C., & Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280–3289.

Webster, K., & Spicer, E. K. (1990) *J. Biol. Chem. 265*, 19007–19014.

Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell 59*, 871–880.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Zimmerman, S. B. (1976) *J. Mol. Biol. 106*, 663–672.

Zimmermann, S. B., Cohen, G. H., & Davies, D. R. (1975) *J. Mol. Biol. 92*, 181–192.

Registry No. $r(UG_4U)_4$, 142928-36-7; guanine, 73-40-5; uracil, 66-22-8.