The Crystal Structure at 1.5 Å Resolution of an RNA Octamer Duplex Containing Tandem G·U Basepairs

Se Bok Jang,* Li-Wei Hung,† Mi Suk Jeong,* Elizabeth L. Holbrook,† Xiaoying Chen,‡ Douglas H. Turner,‡ and Stephen R. Holbrook†

*Korea Nanobiotechnology Center, Pusan National University, Jangjeon-dong, Keumjeong-gu, Busan, Korea; †Structural Biology Department, Physical Biosciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California; and ‡Department of Chemistry, University of Rochester, Rochester, New York

ABSTRACT The crystal structure of the RNA octamer, 5′-GGCGUGCC-3′ has been determined from x-ray diffraction data to 1.5 Å resolution. In the crystal, this oligonucleotide forms five self-complementary double-helices in the asymmetric unit. Tandem 5′GU/3′UG basepairs comprise an internal loop in the middle of each duplex. The NMR structure of this octameric RNA sequence is also known, allowing comparison of the variation among the five crystallographic duplexes and the solution structure. The G•U pairs in the five duplexes of the crystal form two direct hydrogen bonds and are stabilized by water molecules that bridge between the base of guanine (N2) and the sugar (O2′) of uracil. This contrasts with the NMR structure in which only one direct hydrogen bond is observed for the G•U pairs. The reduced stability of the r(CGUG)$_2$ motif relative to the r(GGUC)$_2$ motif may be explained by the lack of stacking of the uracil bases between the Watson-Crick and G•U pairs as observed in the crystal structure.

INTRODUCTION

Non-Watson-Crick pairs are a common element of RNA secondary structure and G•U pairs are the most common type. Single G•U pairs can be important for recognition by proteins and for formation of tertiary interactions. Examples include recognition of tRNA by its cognate synthetase (1) and the formation of tertiary structure in group I introns (2). Often, G•U pairs are found adjacent to each other, but certain combinations are more common than others (3,4). In particular, the 5′UG/3′GU motif is much more common than 5′GU/3′UG. The thermodynamics of adjacent G•U pairs are also sequence-dependent (3,5). For example, the motifs 5′GGUC/3′CUGG and 5′CGUG/3′GUGC are associated with free energy increments at 37°C of −4.1 and −1.5 kcal/mol, respectively, corresponding to more than a 50-fold difference in equilibrium constant for folding. It has been suggested that this difference in energetics is due to the electrostatics of stacking interactions (6). On the basis of NMR, molecular modeling, and chemical substitution experiments, it has also been suggested that these electrostatics lead to the G•U pairs in the 5′GGUC/3′CUGG motif having two hydrogen bonds (6), whereas those in the 5′CGUG/3′GUGC motif have one bifurcated hydrogen bond (7).

The crystal structures of several RNA oligonucleotides incorporating symmetric internal loops have been previously determined (8–14) and shown to have continuous basepairing with formation of U•G, G•A, U•C, U•U, and other non-Watson-Crick pairs. The helices containing these internal loops generally retain A-form geometry. The solution structures of r(GAGUGGUC)$_2$ and r(GGCGUGCC)$_2$ duplexes have been determined by NMR and restrained simulated annealing (7). The global geometry of both duplexes is close to *A*-form, with some distortions localized in the tandem G•U pair region. The striking observation was that in r(GGCGUGCC)$_2$ each G•U pair apparently has only one hydrogen bond instead of the two expected for a canonical wobble pair.

Here we present the crystal structure of the RNA octamer, r(GGCGUGCC) at 1.5 Å resolution. Basepairing between the central nucleotides within a double helix leads to the formation of a tandem GU/UG motif in which the G•U basepairs each have two hydrogen bonds. This contrasts with the NMR structure of the same duplex (7).

MATERIALS AND METHODS

Synthesis, crystallization, and data collection

The octaribonucleotide rGGCGUGCC (GU8) was synthesized chemically (7) and purified by anion exchange (DEAE) FPLC column chromatography with a linear salt gradient from 0.4 M to 2.0 M sodium acetate and pH gradient from 50 mM Tris-HCl, pH 6.8 to pH 7.3. Crystals of r(GGCGUGCC) were grown at room temperature from a solution of 0.2 M magnesium acetate, 0.1 M sodium cacodylate, pH 6.5, and 10% MPD to a size of 400 × 300 microns. The crystals were mounted in thin-walled quartz capillaries and data were collected at the Advanced Light Source synchrotron at Lawrence Berkeley National Laboratory, Berkeley, CA, to 1.5 Å resolution. These crystals belong to space group *P*1 with the following cell dimensions: $a = 26.33$ Å, $b = 43.99$ Å, $c = 54.97$ Å and $\alpha = 112.09°$, $\beta = 99.03°$, and $\gamma = 89.96°$. The R_{merge} for this data was 14.0% and the average F^2/ρ was 2.0 for all data to 1.5 Å and 3.3 for the data to 2.5 Å. The data were processed with the programs DENZO and SCALEPACK (15).

Submitted January 9, 2006, and accepted for publication February 28, 2006.

Address reprint requests to S. B. Jang, Tel.: 82-51-510-2523; E-mail: sbjang@pusan.ac.kr; or S. R. Holbrook, Tel.: 1-510-486-6059; E-mail: srholbrook@lbl.gov.

0006-3495/06/06/4530/08 $2.00

doi: 10.1529/biophysj.106.081018

Structure determination

The structure was determined by the molecular replacement methods AMORE (16) and EPMR (17). The NMR solution structure of r(GGCGUGCC)$_2$ (7) with 20 Å^2 temperature factors was used as the successful search model for the crystal structure of r(GGCGUGCC)$_2$. The search was carried out with data between 12.0 and 4.5 Å and four molecules were located with a correlation coefficient of 55.4 ($R = 44.0\%$) using AMORE. For the structure determination of the fifth molecule, the original (7) and refined models were tried as search models using AMORE, but all attempts failed. Subsequently, the structure of the fourth molecule refined by CNS was used as a search model for the fifth molecule (18). The positioning of the fifth duplex was by a six-parameter search, using data between 15.0 and 3.0 Å resolution and the solution with EPMR including all five duplexes, had a correlation coefficient of 69.2 ($R = 43.1\%$).

Model refinement

The data between 20.0 and 1.5 Å resolution were used for all refinement cycles. Each cycle consisted of positional, followed by simulated annealing, and finally B-factor refinement using the CNS program of Brunger (18). Restraints were placed on bond lengths, bond angles, nonbonded contacts, temperature factors of neighboring atoms, planarity of the bases, and noncrystallographic symmetry. Difference Fourier and 2*Fo-Fc* electron density maps, as well as omit maps, were calculated at regular intervals to allow manual modification. The rebuilding of the model and addition of solvent were done using the O graphics program (19). Solvent molecules were added conservatively, with due regard for their environment including potential interactions with hydrogen-bonded partners. At the end of the refinement of r(GGCGUGCC)$_2$ the crystallographic *R*-factor was 22.6% and R_{free} was 28.5% with 614 bound water molecules as shown in Table 1. Electron density was well defined for all nonhydrogen atoms. The models exhibit good geometry with root mean-square deviation (RMSD) from ideal bond lengths and angles of 0.004 Å and 0.890°.

TABLE 1 Data and refinement statistics

Space group	*P*1
Cell dimensions	$a = 26.33$ Å
	$b = 44.0$ Å
	$c = 54.97$ Å
	$\alpha = 112.10°$
	$\beta = 99.03°$
	$\gamma = 89.96°$
Resolution (Å)	19.95–1.499
Refinement completeness (%)	84.2
Observed reflections	31,611
Unique reflections	3014
R_{merge} (%)*	12.1 (35.3)
Number of nonhydrogen RNA	1690
Molecules per asymmetry unit	5
Solvent molecules	613
R_{cryst} (%)† ($F \geq 1.0\ \sigma$) (28,887 reflections between 20 and 1.5 Å)	22.6
R_{free} (%)† (1513 reflections with 5% data)	28.5
Bond lengths (Å)	0.004
Bond angles (deg)	0.890
Average $\langle B \rangle$ factor (Å^2)	50.5

Numbers in parentheses indicate values for the highest resolution bin.

*$R_{merge} = \sum_h \sum_i |I(h,i) - \langle I(h) \rangle| / \sum_h \sum_i I(h,i)$, where $I(h,i)$ is the intensity of the i^{th} measurement of reflection h and $\langle I(h) \rangle$ is the mean value of $I(h,i)$ for all i measurements.

†$R_{cryst} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$, where F_{obs} denotes the observed structure factor amplitude, and F_{calc} denotes the structure factor amplitude calculated from the model; 5% of reflections were used to calculate R_{free} (data 20.0–1.5 Å used in the refinement).

RESULTS

In the crystal, two strands of the RNA octamer rGGCGUGCC form an *A*-type double-helix (GU8 duplex) incorporating tandem G•U basepairs as shown schematically in Fig. 1. A stereoview of a representative duplex with its computed helical axis is shown in Fig. 2 *a*, together with a space-filling illustration in Fig. 2 *b*. These duplexes stack head to head to produce pseudo-infinite helices throughout the crystal.

Electron densities were clear for all nonhydrogen atoms as shown for the basepairs in Fig. 3, *a* and *b*. The G•U basepairing geometry is illustrated in Fig. 4, *a–e*. In the crystal, the two G•U basepairs of each helix exhibit similar base-base hydrogen bonding patterns as illustrated in Fig. 4. Stereochemical analysis suggests that the guanine is protonated at the N1 position and the uracil is protonated at the N3 position, forming N1(G)-O2(U) and N3(U)-O6(G) hydrogen bonds in every case.

The G•U basepairs are also linked by bridging water molecules in the major and minor grooves. In all instances except one of the G•U basepairs of duplex *A* (Fig. 4 *a*), a bound water is found in the minor groove linking N2(G) and O2(U). This water is also bound to the uracil ribose O2′(U), thus stabilized by three hydrogen bonds, N2(G)-O(H_2O), O2′(U)-O(H_2O), and O2(U)-O(H_2O). This type of water-mediated basepairing has been previously observed in many crystal structures, including a duplex incorporating tandem C•A$^+$ basepairs (13) and r(CGACUUCGGUCC)$_2$ (8). The importance of this integral water in stabilizing the G•U basepairs is suggested by the extreme instability of tandem I•U basepairs that lack the exocyclic amino group of guanine (20). A bound water is also observed in the major groove of 7 of the 10 G•U basepairs in the crystal linking O6(G) and O4(U). This major groove water is not seen in either G•U basepair of duplex *D*.

In the NMR structure (7) shown in Fig. 4 *f*, a bifurcated hydrogen bond is seen between O2(U) and N1(G) and N2(G), implying that the minor groove water may not be binding in the same manner. The two G•U basepairs are symmetrically related and show only this single base-base bifurcated hydrogen bond in contrast to the examples observed in the crystal structure.

FIGURE 1 Schematic diagram of *r*(GGCGUGCC)$_2$, as found in the crystal structure. The self-complementary sequence forms a duplex with Watson-Crick basepairs indicated by solid lines and noncanonical G•U basepairs by double dots. Residue annotation for the two strands is shown. The numbering of the duplex residues is 100–500 plus these annotations for duplexes *A–E*.

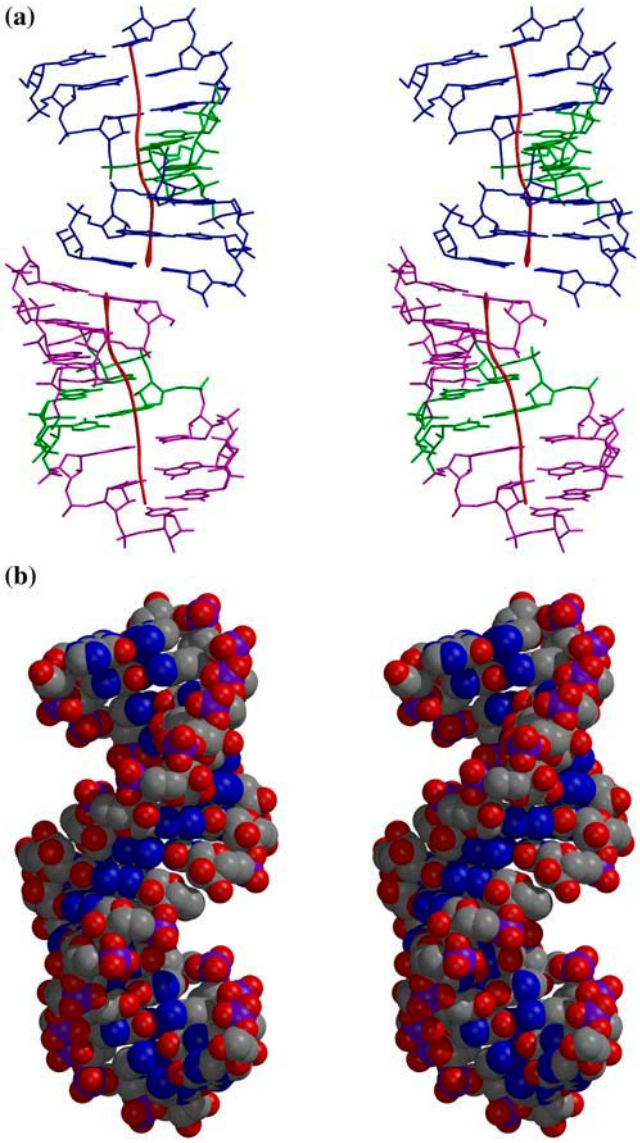

FIGURE 2 Stereoviews of the GU8 duplexes. (*a*) Stick representation of two stacked GU8 duplexes (Watson-Crick pairs in *blue* and *magenta*, G•U pairs in *green*) found in the crystallographic asymmetric unit. The curvature of the helical axis as calculated by the CURVES program is depicted by a red line. Bound waters are not shown in this illustration. (*b*) Space-filling representation of the GU8 double helices shown in panel *a*, using the same viewpoint. Atom types are color-coded as follows: carbon, gray: nitrogen, blue; oxygen, red; and phosphorus, pink.

The five independent duplexes can be compared in several ways including least-squares superposition, torsion angles, and helical parameters. Global helical parameters for the central six basepairs of each of the five independent duplexes, the NMR solution structure (7) (PDB code 1EKD), and canonical A-form RNA, calculated with the CURVES program (22) are compared in Table 2. The overall curvature of the NMR duplex model is much greater than that of the crystal structure duplexes.

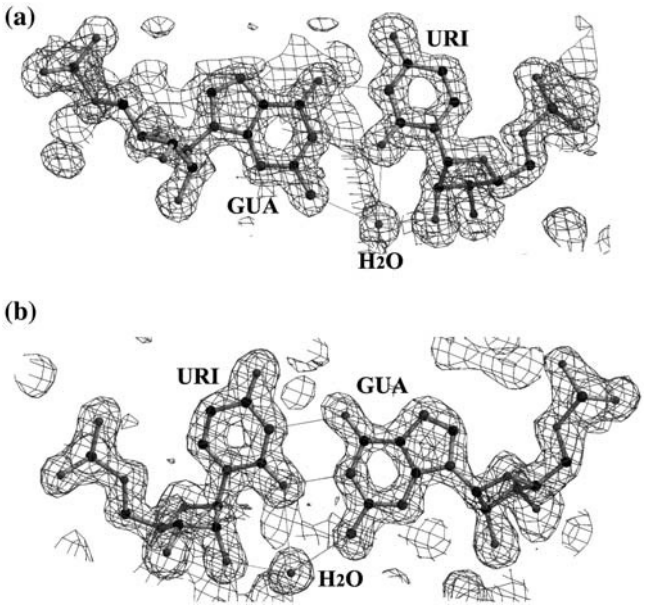

FIGURE 3 (*a*,*b*) Views of two consecutive G•U pairs with an integral bound water molecule in the minor groove (duplex *D* shown). The 2*Fo-Fc* electron density contoured at the 1.0σ level is superimposed on the nucleotide structure.

The result of pairwise least-square superposition of the duplexes is given in Table 3. From this table it is clear that helices A, B, and C of the GU8 crystal structure are generally more similar to each other (RMSD = 0.68–0.94 Å) than to the NMR model, 1EKD (RMSD = 1.33–1.72 Å). Helices D and E of the GU8 crystal are similar to each other (RMSD = 0.57 Å) and to the NMR structure (RMSD = 0.85–0.86 Å), but differ more from helices A, B, and C of the crystal (RMSD = 0.92–1.70 Å).

A more detailed comparison of the five duplexes observed in the crystals and the solution (NMR) structure can be made by examination of their variable torsion angles. Generally, the Watson-Crick regions of the octamers show standard torsion angles with the exception of the 3′-terminal cytosines. The major question is what changes in the dihedral angles are necessary to incorporate the tandem G•U basepairs into the double helix. Examination shows that the ζ (O3′-P) torsion between the G•U and U•G basepairs differs from the standard *gauche* angle ($\sim -60°$) occurring in the Watson-Crick regions. In the crystal structure, this conformational angle varies between $-77°$ and $-94°$ with an average value of -86 (SD 4.7°). In the NMR structure, 1EKD, this angle is $-96.5°$.

Fig. 5, *a–c*, shows the basepair stacking between the tandem G•U pairs and their adjacent Watson-Crick pairs. Although both Watson-Crick to G•U steps show extensive cross-strand purine-purine (guanine-guanine) stacking with little or no overlap between the uracil and cytosine, there is good same-strand stacking between the tandem G•U pairs. This is in contrast to the situation for r(GAGUGCUC)$_2$ duplexes (7) in which the Watson-Crick to G•U pair steps have good

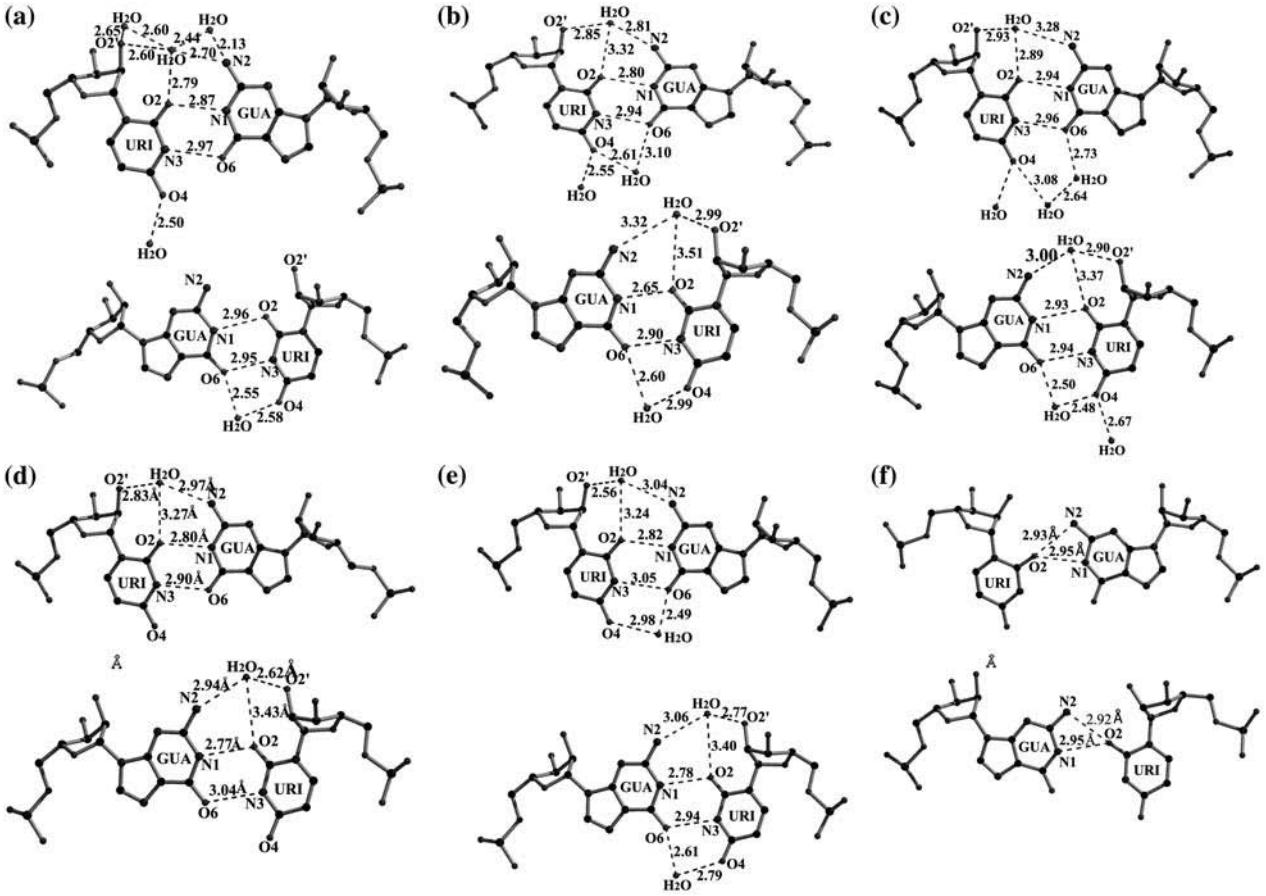

FIGURE 4 (*a–e*) The 10 independent G•U basepairs observed in the asymmetric unit of the GU8 crystal. Hydrogen bonds are indicated by dashed lines (with distances shown in Angstroms). An integral water molecule linking the Guanine amino group (N2) to the uracil O2′ and O2 is found in most, but not all, examples. (*f*) The G•U basepairs observed in the NMR structure GU8 (1EKD). Bifurcated hydrogen bonds are indicated by two dashed lines originating from a common atom.

same-strand stacking, but there is cross-strand purine-purine stacking between the tandem G•U pairs with no uracil overlap.

The packing of the five RNA duplexes in the crystallographic asymmetric unit is shown from two views in Fig. 6. The duplexes are labeled *A–E* for reference. A summary of intermolecular interactions in the crystal is given in Table 4. Gagnon and Steinberg (23) have identified a tertiary interaction motif in which the minor groove of a G•U basepair from one helix is packed against the major groove of Watson-Crick pair from another helix. We also see this interaction in the packing of GU8 duplexes in the crystal. For example, the tandem G•U pairs of duplex *A* interact with tandem G•C pairs of duplex *D* (Table 4). As can be seen from the table, these interduplex interactions involve hydrogen bonding between the N2 and O2′ of the G•U pairs and the O3′, O2′, and O1P atoms of G•C pairs. This type of interhelical interaction involving G•U and G•C basepairs may be a general mode of helical packing in functional or genomic RNA.

TABLE 2 Helical parameters of GU8—comparison to the average NMR structure (1EKD) and canonical A-RNA with terminal basepairs excluded

	Average helical twist (°)	Average rise/residue (Å)	Helical curvature (Å)
GU8			
a)	31.9	3.24	7.9
b)	31.9	3.20	14.8
c)	31.0	3.25	6.4
d)	30.4	3.26	13.8
e)	30.2	3.37	11.4
NMR	29.9	3.24	33.8
A-RNA	32.70	2.81	0.0

Two putative metal ions are localized in the asymmetric unit of the GU8 crystal structure as shown in Fig. 7, both involving the G•U pairs. As shown in Fig. 7 *a*, a large peak is found in the major groove between the tandem G•U pairs of duplex *C* in a more or less octahedral coordination directly to the nucleotide bases. The distances of ~2.4 Å (Table 4) are intermediate between that expected for magnesium and that expected for water or sodium. The occupancy is comparable to that of a water molecule as evaluated by the refined B-factor. This electron density peak is not found in the other

TABLE 3 Least-squares comparison of GU8 (crystal) and NMR (solution) duplex structures excluding terminal residues

		GU8				
		A	B	C	D	E
GU8	A	—	**0.58/0.68** 2.26 G4(O1P)	**0.88/0.94** 2.26 G4(O1P)	**0.99/1.12** 3.01 G2(O1P)	**0.85/0.92** 2.03 C11(O1P)
	B		—	**0.65/0.70** 1.15 G10(P)	**0.20/1.32** 2.79 G2(O1P)	**1.10/1.19** 2.66 C11(O1P)
	C			—	**1.64/1.70** 2.98 G2(O1P)	**1.51/1.56** 2.74 C11(O1P)
	D				—	**0.52/0.57** 1.70 G12(O5′)
	E					—
NMR (1EKD)		**/1.33**	**/1.50**	**/1.72**	**/0.85**	**/0.86**

The average/root mean-square deviations are in Angstroms (bold). The maximum deviations are shown immediately below in Angstroms. For the NMR structure, only the root mean-square deviations are shown after a slash.

four duplexes. The other position has a fully hydrated metal ion, presumably magnesium, involved in intermolecular bridging between duplexes *B* and *E* (Fig. 7 *b*).

The root mean-square deviations (RMSDs) of just the two central tandem G•U/U•G mismatches of the x-ray and NMR (1EKD) solution structures are 1.19 and 1.17 Å, respectively (Fig. 8, *a* and *b*). The RMSD of the x-ray and the NMR whole structures is 1.88 Å. The region of G•U/U•G of the two central tandem pairs adopt a different orientation with conformational changes by dynamic structure in NMR solution.

The root mean-square deviations of the GGC/GCC sequences in this G•U structure and the previous x-ray structure of $(GGC\underline{GA}GCC)_2$ (1SA9) are 0.97 and 0.86 Å, respectively (14). The RMSD of the $(GGC\underline{GU}GCC)_2$ and the $(GGC\underline{GA}GCC)_2$ is 2.40 Å. From the superposition of two structures, the residues with RMSDs exceeding ~1.48 Å are located at the central tandem pairs region. These results indicate that the big conformational changes between the isolated GGC/GCC and the whole structure are mostly the two central tandem mismatches region.

DISCUSSION

The basepairs formed between guanine and uracil, G•U pairs, are of particular interest due to their thermodynamic stability, widespread occurrence in biological molecules, and polymorphism. The crystal structure of r(GGC<u>GU</u>GCC) provides five independent examples of self-complementary duplexes containing a total of 10 G•U pairs.

The solution structure of the same sequence has been solved by NMR methods (7) and proposed to have only a single hydrogen bond (three-centered, O2(U)-N2(G), N1(G)) for each G•U pair (Fig. 4 *f*). A distance of 3.4 Å was reported

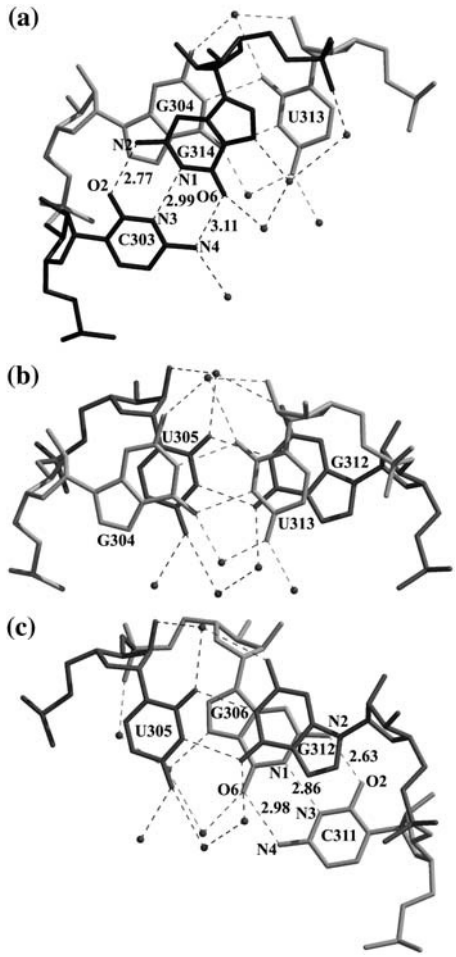

FIGURE 5 Base stacking between GU/UG pairs and adjacent C•G and G•C Watson-Crick pairs. (*a*) Stacking between G•U pair and C•G Watson-Crick pair. (*b*) Stacking between G•U and U•G pairs. (*c*) Stacking between U•G pair and G•C Watson-Crick pair. Bound water molecules are shown as black spheres.

between the imino hydrogen at N3(U) and O6(G), which is too long for a hydrogen bond.

Substitution of a methyl group for the imino hydrogen at the N3 of each U made duplex formation less favorable by 2.6 kcal/mol at 37°C, which was a smaller effect than the average destabilization by 6.0 kcal/mol observed for duplexes with tandem G•U pairs having two hydrogen bonds (7). Thus, thermodynamic measurements were consistent with the interpretation of the NMR data.

In the crystal structure, two hydrogen bonds, O2(U)-N1(G) and N3(U)-O6(G), are formed for all G•U basepairs, with average distances of 2.83 Å (SD 0.09 Å) and 2.96 Å (SD 0.05 Å), respectively. Thus, in the crystal, the N3(U)-O6(G) hydrogen bond, although longer and therefore weaker than the O2(U)-N1(G) bond, is clearly formed. A bridging water molecule with three ligands is present in the minor groove in 9 of the 10 examples, and in the major groove a bridging water is present with two ligands in 7 of 10 cases (see Fig. 4),

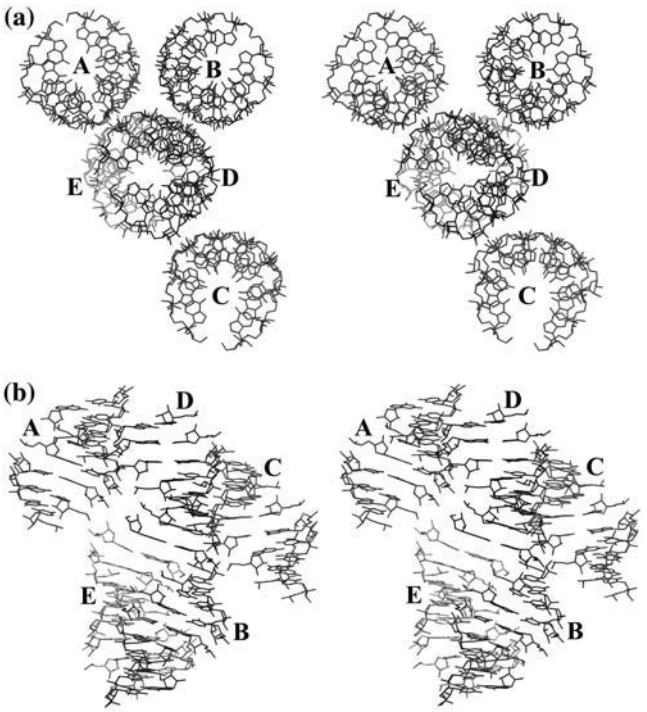

FIGURE 6 Two views of the GU8 double-helix structure. (*a*) Structural packing of the five self-complementary double helices in the asymmetric unit viewed from the top. (*b*) The same representation rotated by 90°. The five molecules in the asymmetric unit are labeled as *A*, *B*, *C*, *D*, and *E*.

TABLE 4 The distances of intermolecular interactions and G•U/U•G basepair interaction with a ligand

Interactions	Distance (Å)
Base N2 (A112Gua) - Sugar O3′ (D402Gua)	3.46
Sugar O2′ (A113Uri) - Phosphate O1P (D403Cyt)	2.42
Phosphate O1P (D412Gua) - Sugar O2′ (C315Cyt)	2.71
Sugar O2′ (C304Gua) - Sugar O2′ (D411Cyt)	3.72
Sugar O4′ (C304Gua) - Sugar O2′ (D402Cyt)	3.62
Base N2 (B204Gua) - Sugar O2′ (E503Cyt)	3.04
Base N2 (B204Gua) - Sugar O3′ (E503Cyt)	3.11
Sugar O2′ (B214Gua) - Phosphate O1P (E504Gua)	2.63
Phosphate O1P (B206Gua) – Sugar O2′ (E516Cyt)	3.19
Ligand (Mg^{2+})-Base	
Base O2 (C305Uri) - Mg^{2+}	2.37
Base O2 (C313Uri) - Mg^{2+}	2.40
Base N1 (C304Gua) -Mg^{2+}	2.42
Base O6 (C304Gua) -Mg^{2+}	2.45
Base N1 (C312Gua) -Mg^{2+}	2.46
Base O6 (C312Gua) -Mg^{2+}	2.43
Base N3 (C305Uri) - Mg^{2+}	2.54
Base N3 (C313Uri) - Mg^{2+}	2.51
Mg^{2+}-Solvent	
Mg^{2+} - S23	2.09
Mg^{2+} - S103	2.17
Mg^{2+} - S286	2.01
Mg^{2+} - S27	2.03
Mg^{2+} - S33	2.14
Mg^{2+} - S295	2.08

The hydrated Mg^{2+} ion interacts with G-C/C-G Watson-Crick pairs and six water molecules.

perhaps resulting in greater stabilization of the O2(U)-N1(G) hydrogen bond compared to the O6(G)-N3(U) hydrogen bond.

The NMR and crystal structures clearly differ in the hydrogen bonding within the G•U pairs. There are several possible reasons for this. There are no NMR data that specifically forbid an N3(U)-O6(G) hydrogen bond and without employing molecular dynamics it was possible to locate a reasonable A-form RNA-like structure for r(GGCGUGCC)$_2$ that contained G•U pairs with two hydrogen bonds and that did not have major violations with the NMR data (7). Thus, the structure is dependent on the force field used for modeling with NMR restraints. Modeling that used a force constant of 100 kcal/(mol A2) to force two hydrogen bonds still gave the one hydrogen-bond structure, however, so the energy calculations strongly favor the one hydrogen-bond model (7). The imino resonances for the G•U pairs were broad and both gave unusually strong cross-peaks to water, which suggest a dynamic structure. Moreover, the modeling was done in the absence of water. It is likely, however, that the energetics of binding water are different in solution and crystal phases due to different entropy changes. The crystal structure is definitive at this resolution. It is nevertheless possible that the average solution structure differs slightly from the crystal structures. A molecular dynamics study of tandem G•U pairs suggests a range of conformations in solution (24). Moreover, crystal structures can be affected by packing interactions. Another possibility for the structural differences is that Mg^{2+} is present in the crystals but not in the NMR samples. The Mg^{2+} coordination shown in Fig. 7 *a*, which bridges the O2 and O6 atoms of the tandem G•U pairs, although only observed for duplex *C*, may play a role in stabilization of the second hydrogen bond. The available data, therefore, suggest that tandem G•U pairs are structurally pliable.

If the second hydrogen bond is at least partially formed between G•U pairs with this flanking sequence (..CGUG..), then how can we explain the lower stability of this sequence with respect to other symmetric tandem G•U motifs (7) and the relatively rare occurrence of this motif in biological RNAs?

The stacking diagrams of Fig. 5 suggest one possible explanation. While the self-complementary duplex of sequence r(GAGUGCUC)$_2$ has poor base stacking between the tandem G-U pairs (7), the GU8 sequence has poor stacking between the G•U pairs and their adjacent Watson-Crick pairs, with good stacking between the G•U pairs themselves. Thus, the two basepair steps with reduced stacking in GU8 may result in less stability and occurrence than other sequences with only one step with reduced overlap. Overlap of bases is probably not the only determinant of stability, however. For example, the NMR structure of the (GGUC)$_2$ motif in the (GAGGUCUC)$_2$ duplex (6) shows overlap similar to that of the (CGUG)$_2$ motif in (GGCGUGCC)$_2$ (Fig. 5), but is 3.8 kcal/mol more stable at 37°C (6).

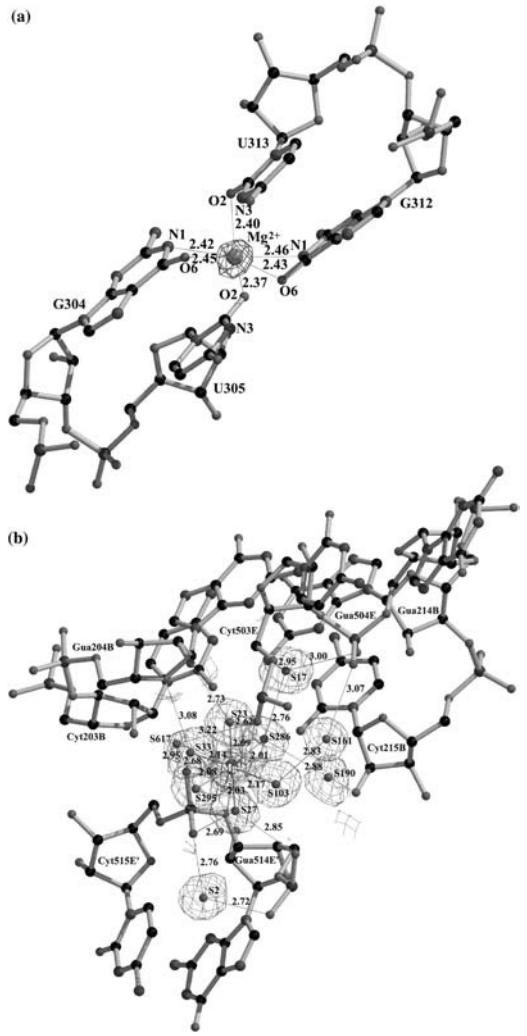

FIGURE 7 Ball-and-stick diagram of the coordination environments of the two bound Mg^{2+} ions. (*a*) Electron density (2*Fo-Fc*) contoured at 1σ for the bound Mg^{2+}. The Mg^{2+} ion is in octahedral coordination with the bases of GU/UG pairs in the *C* duplex. (*b*) Difference electron density (*Fo-Fc*) contoured at 3.0σ in which the Mg^{2+} ion is omitted from the structure factor and phase calculation. Electron density (2*Fo-Fc*) is contoured at 1σ for bound water molecules. The bound Mg^{2+} ion is in octahedral coordination with six water molecules (bridges between duplexes *B* and *E*).

In summary, we have determined the crystal structure of a self-complementary octameric RNA sequence that forms five unique duplexes in the crystal and compared it with the solution structure determined by NMR methods. Among the five unique duplexes observed in this crystal structure, we have found all G•U pairs stabilized by two base-base hydrogen bonds, O2(U)-N1(G) and N3(U)-O6(G), as well as either one or two bridging water molecules linking O2(U)-N2(G) and sometimes O4(U)-O6(G) hydrogen bonds. Based on the difference in stacking between duplexes with the (..CGUG..) motif compared to the (..GUGC..) motif, we propose that the difference in thermodynamic stability may be due to poor stacking between the Watson-Crick and G•U basepairs as observed in the GU8 crystal structure.

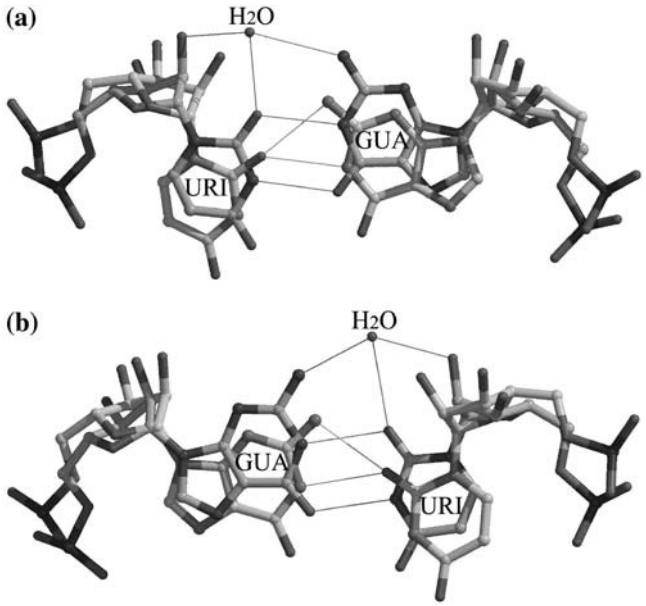

FIGURE 8 (*a*,*b*) Views of superposition of tandem G•U/U•G mismatches of x-ray and NMR (1EKD) structures. The symmetric G•U/U•G pairs are incorporating with a bound water molecule and the hydrogen bonds are indicated by solid lines.

The refined coordinates of the GU8 duplex have been deposited in the NDB Nucleic Acids Database (http://ndbserver.rutgers.edu) (25) as NDB ID code No. AR0067 and in the Protein DataBank as PDB ID code No. 2G3S.

The authors acknowledge the significant contributions of Hanna Walukievicz and Ramona Pufan in purification and crystallization of this RNA.

This work was supported by National Institutes of Health grant No. GM 4921501 to S.R.H., grant No. GM 22939 to D.H.T., and by Korea Research Foundation grant No. KRF-2005-041-E00510 to S.B.J. Facilities and equipment were provided through support of the Office of Energy Research, Office of Health and Environmental Research, and Health Effects Research Division of the U.S. Department of Energy.

REFERENCES

1. Gabriel, K., J. Schneider, and W. H. McClain. 1996. Functional evidence for indirect recognition of G•U in tRNA(Ala) by alanyl-tRNA synthetase. *Science*. 271:195–197.
2. Strobel, S. A., and T. R. Cech. 1995. Minor groove recognition of the conserved G•U pair at the *Tetrahymena* ribozyme reaction site. *Science*. 267:675–679.
3. He, L., R. Kierzek, J. SantaLucia, Jr., A. E. Walter, and D. H. Turner. 1991. Nearest-neighbor parameters for G•U mismatches: 5′GU3′/3′UG5′ is destabilizing CGUG/GUGC, UGUA/AUGU, and AGUU/UUGA but stabilizing in GGUC/CUGG. *Biochemistry*. 30:11124–11132.
4. Gautheret, D., D. Konings, and R. R. Gutell. 1995. G•U basepairing motifs in ribosomal RNA. *RNA*. 1:807–814.
5. Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288:911–940.
6. McDowell, J. A., and D. H. Turner. 1996. Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: solution structure of $(rGAGGUCUC)_2$ by two-dimensional NMR and simulated annealing. *Biochemistry*. 35:14077–14089.

7. Chen, X., J. A. MacDowell, R. Kierzek, T. R. Krugh, and D. H. Turner. 2000. Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for G•U pairs: one hydrogen bond for each G•U pair in r(GGCGUGCC)$_2$ and two for each G•U pair in r(GAGUGCUC)$_2$. *Biochemistry*. 39: 8970–8982.
8. Holbrook, S. R., C. Cheong, I. Tinoco, Jr., and S. H. Kim. 1991. Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick basepairs. *Nature*. 353:579–581.
9. Cruse, W. B. T., P. Saludijan, E. Biala, P. Strazewski, T. Prange, and O. Kennard. 1994. Structure of a mispaired RNA double helix at 1.6 Å resolution and implications for the prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA*. 91:4160–4164.
10. Baeyens, K. J., H. L. DeBondt, and S. R. Holbrook. 1995. Structure of an RNA double helix including uracil-uracil basepairs in an internal loop. *Nature Struct. Biol.* 2:56–62.
11. Baeyens, K. J., H. L. De Bondt, A. Pardi, and S. R. Holbrook. 1996. A curved RNA helix incorporating an internal loop with G•A and A•A non-Watson-Crick basepairing. *Proc. Natl. Acad. Sci. USA*. 93:12851–12855.
12. Lietzke, S. E., C. L. Barnes, J. A. Berglund, and C. E. Kundrot. 1996. The structure of an RNA dodecamer shows how tandem U•U basepairs increase the range of stable RNA structures and the diversity of recognition sites. *Structure*. 4:917–930.
13. Jang, S. B., L.-W. Hung, Y.-I. Chi, E. L. Holbrook, R. Carter, and S. R. Holbrook. 1998. Structure of an RNA internal loop consisting of tandem C•A$^+$ basepairs. *Biochemistry*. 37:11726–11731.
14. Jang, S. B., K. J. Baeyens, M. S. Jeong, J. SantaLucia, Jr., D. H. Turner, and S. R. Holbrook. 2004. Structures of two RNA octamers containing tandem G•A basepairs. *Acta Crystallogr*. D60:829–835.
15. Otwinowski, Z., and W. Minor. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 276:307–326.
16. Navaza, J. 1994. AMORE: an automated package for molecular replacement. *Acta Crystallogr*. A50:157–163.
17. Kissinger, C. R., D. K. Gehlhaar, and D. B. Fogel. 1999. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr. D Biol. Crystallogr*. 55:484–491.
18. Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr*. 54:905–921.
19. Jones, T. A., J.-Y. Zou, S. W. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr*. A47: 110–119.
20. Serra, M. J., P. E. Smolter, and E. Westhof. 2004. Pronounced instability of tandem I•U basepairs in RNA. *Nucleic Acids Res*. 32: 1824–1828.
21. Reference deleted in proof.
22. Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: convention and principles. *J. Biomol. Struct. Dyn*. 6: 655–667.
23. Gagnon, M. G., and S. V. Steinberg. 2002. GU receptors of double helices mediate tRNA movement in the ribosome. *RNA*. 8:873–877.
24. Pan, Y. P., U. D. Priyakuman, and A. D. MacKerell. 2005. Conformational determinants of tandem GU mismatches in RNA: insights from molecular dynamics simulations and quantum mechanical calculations. *Biochemistry*. 44:1433–1443.
25. Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The Nucleic Acid Database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J*. 63: 751–759.