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)$_2$[†,‡]

Xiaoying Chen,[§] Jeffrey A. McDowell,[§,||] Ryszard Kierzek,[⊥] Thomas R. Krugh,[§] and Douglas H. Turner*[,§]

Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, and Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60−714 Poznan, Noskowskiego 12/14, Poland

Received December 23, 1999; Revised Manuscript Received April 3, 2000

ABSTRACT: G·U pairs occur frequently and have many important biological functions. The stability of symmetric tandem G·U motifs depends both on the adjacent Watson−Crick base pairs, e.g., 5′G > 5′C, and the sequence of the G·U pairs, i.e., 5′-UG-3′ > 5′-GU-3′, where an underline represents a nucleotide in a G·U pair [Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry* 34, 3204−3211]. In particular, at 37 °C, the motif 5′-CGUG-3′ is less stable by approximately 3 kcal/mol compared with other symmetric tandem G·U motifs with G-C as adjacent pairs: 5′-GGUC-3′, 5′-GUGC-3′, and 5′-CUGG-3′. The solution structures of r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$ duplexes have been determined by NMR and restrained simulated annealing. The global geometry of both duplexes is close to A-form, with some distortions localized in the tandem G·U pair region. The striking discovery is 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. In the one-hydrogen-bond model, the distance between GO6 and UH3 is too far to form a hydrogen bond. In addition, the temperature dependence of the imino proton resonances is also consistent with the different number of hydrogen bonds in the G·U pair. To test the NMR models, U or G in various G·U pairs were individually replaced by *N*3-methyluridine or isoguanosine, respectively, thus eliminating the possibility of hydrogen bonding between GO6 and UH3. The results of thermal melting studies on duplexes with these substitutions support the NMR models.

Knowing the elements of three-dimensional structures for RNA would greatly help to understand the molecular basis of RNA stability and function. A particularly interesting motif is the G·U pair. It is the most common base pair besides the Watson−Crick base pairs and accounts for about 50% of non-Watson−Crick base pairs and mismatches (*1*−*4*). The G·U pair was first hypothesized by Crick (*5*) to explain codon degeneracy. G·U pairs are also frequent in tRNA structures (*6*, *7*). They are often found in the middle of acceptor stems and are often important in the recognition of tRNA by its cognate synthetase (*8*−*10*). One example is the G3·U70 pair in the acceptor stem of tRNA^{Ala} from *Escherichia coli* (*11*, *12*). The acceptor identity is retained only if the G·U pair is replaced by other wobble pairs (*13*, *14*), and the single G·U pair is sufficient as the identity site (*15*). The G·U pair also has an important function in the P1 helix of group I introns, where a highly conserved G·U pair facilitates formation of the active tertiary structure by providing a guanosine amino group to form hydrogen bonds (*16*). When other possible pairs are substituted in this position, only C·A, presumably having similar geometry as a G·U pair, is tolerated in terms of the binding of substrate to ribozyme and fidelity of splicing (*17*). G·U pairs are also thought to have functions important in RNA editing (*18*). Clearly, the G·U pair has important biological functions.

The thermodynamic stabilities of symmetric tandem G·U motifs depend both on the adjacent Watson−Crick base pairs, i.e., 5′G > 5′C > 5′U ≥ 5′A, and on the sequence of the G·U pairs, i.e., 5′UG3′ > 5′GU3′, where here and henceforth an underline designates symmetric tandem G·U pairs (*19*, *20*). The thermodynamic data are particularly unusual for 5′-CGUG-3′. This motif is less stable by approximately 3 kcal/mol compared with other symmetric tandem G·U motifs with G-C as adjacent pairs: 5′-GGUC-3′, 5′-GUGC-3′, and 5′-CUGG-3′, although the only difference between them is either the swapping of the adjacent G-C pairs or the sequence of the G·U pairs (Table 1). Thermodynamic data for all duplexes containing 5′-GU-3′ do not fit into the nearest-neighbor model (*20*), suggesting structural differences between the duplexes. A survey of the available rRNA secondary structure database (*1*, *2*) reveals that 5′-CUGG-3′

[†] This work was supported by NIH Grants GM 22939 (D.H.T), 1R03 TW1068-01 (R.K. and D.H.T.), and GM 53826 (T.R.K.).

[‡] The coordinates and restraints reported in this paper have been deposited with the Brookhaven Protein Data Bank: PDB ID 1EKA and RCSB ID RCSB010668 for r(GAGUGCUC)$_2$; PDB ID 1EKD and RCSB ID RCSB010670 for r(GGCGUGCC)$_2$.

* To whom correspondence should be addressed: tel (716) 275-3207; fax (716) 473-6889; e-mail Turner@chemistry.rochester.edu.

[§] University of Rochester.

[||] Present address: Abbott Laboratories, Building AP10/2, Department 42T, 100 Abbott Park Rd., Abbott Park, IL 60064-6100.

[⊥] Polish Academy of Sciences.

10.1021/bi992938e CCC: $19.00
Published on Web 07/06/2000

Table 1: Free Energy Increments (kcal/mol) at 37 °C for Symmetric Tandem G·U Motifs in RNA Duplexes[a]

adjacent base pairs	G·U pairs 5′UG3′ 3′GU5′	G·U pairs 5′GU3′ 3′UG5′
		−4.1[d] (AUGGUCAU)
5′G	−4.9[b] (AUGUGCAU)	−4.7[d] (CGGGUCCG)
3′C	−4.8[c] (GAGUGCUC)	−4.2[e] (GAGGUCUC)
		−3.4[d] (CUGGUCAG)
5′C	−4.2[b] (GCUGGC)	−1.3[d] (GUCGUGAC)
3′G	−4.1[c] (GGCUGGCC)	−0.8[d] (GGCGUGCC)
5′U	−2.9[d] (CGUUGACG)	−0.3[d] (CCUGUAGG)
3′A	−2.3[c] (GGUUGACC)	
5′A	−2.1[d] (GGAUGUCC)	−0.1[f](GGAGUUCC)
3′U	−1.6[f] (CCAUGUGG)	0.5[f] (CCAGUUGG)

[a] Increments (given in kilocalories per mole) are defined as in the following example for 5′GUGC3′: $\Delta G°_{37}$(5′GUGC3′) = $\Delta G°_{37}$(5′GAGUGCUC3′) − $\Delta G°_{37}$(5′GAGCUC3′) + $\Delta G°_{37}$(5′GC3′). Values of free energy increments are listed next to the sequence containing the tandem G·U motifs. Data for duplexes are from cited references, and values for the last term are from ref 75. [b] Reference 77. [c] Reference 57. [d] Reference 20. [e] Reference 19. [f] Reference 58.

Chart 1

I **II**

and 5′-GUGC-3′ occur 33 and 24 times, respectively, while 5′-CGUG-3′ and 5′-GGUC-3′ occur only three and two times, respectively. To determine if context affects the structures of tandem G·U pairs, NMR structures have been determined for duplexes r(GAGUGCUC)$_2$ (Chart 1, structure I) and r(GGCGUGCC)$_2$ (Chart 1, structure II). To further investigate the origin of stability differences in G·U pairs, *N*3-methyluridine (mU) or isoguanosine (iG) were individually substituted for U or G, respectively, in G·U pairs in various RNA duplexes. The results of the NMR and substitution studies suggest a model in which the relatively less stable G·U pairs have only one hydrogen bond each.

MATERIALS AND METHODS

RNA Synthesis and Purification. Protected isoguanosine (iG) was synthesized from 2,6-diaminopurine riboside with some modifications of the original procedures (*21, 22*). To convert 2,6-diaminopurine riboside into isoguanosine, the substrate was treated with lithium nitrite in the presence of acetic acid. Substitution of the sodium nitrite in the original procedure by lithium salt simplified purification because the lithium salt is methanol-soluble. Isoguanosine was obtained in 67% yield. The 6-amino group of isoguanosine was protected with dimethylchloromethyleneammonium chloride (*22*, *23*) followed by protection of the 5′-hydroxyl with dimethoxytrityl chloride. After column purification, the 5′-*O*-dimethoxytrityl-6-[*N*-[(dimethylamino)ethylidene]amino]-isoguanosine was obtained in 84% yield for the last two reactions. After treatment with *tert*-butyldimethylsilyl chloride in the presence of imidazole, the 2′-silylated isomer of protected isoguanosine was obtained in 45% yield. This last derivative was transformed into the 3′-phosphoramidite with 60% yield (*24*).

Phosphoramidite of *N*3-methyluridine (mU) was purchased from ChemGenes. Regular RNA phosphoramidites were purchased from GLEN Research. Oligonucleotides were synthesized by β-cyanoethyl phosphoramidite chemistry (*24*, *25*) with an Applied Biosystems DNA/RNA synthesizer, model 392. Removal from the CPG support and amino deprotection were done by treatment with ethanolic ammonia. Silyl deprotection was accomplished by incubation with freshly made 1 M triethylaminohydrogen fluoride at 55 °C for ∼50 h. Samples were desalted by dissolving in 5 mM ammonium bicarbonate at pH 7 or deionized water at pH 4 and passing through a C-18 column (Waters). Oligonucleotides were purified by TLC on Si500F plates (Baker) with the solvent 55:35:10 1-propanol/ammonia/water. The duplex r(GGCGUGCC)$_2$ was purified further by HPLC on a C-18 reverse phase column (Supercosil) using 50% acetonitrile as mobile phase. Purity for all the duplexes was >98% as determined by HPLC with a C-8 reverse-phase column (Hamilton) and 50% methanol as mobile phase or with a C-18 reverse-phase column (Supercosil) and 50% acetonitrile as mobile phase.

NMR Spectroscopy. Oligonucleotides were dialyzed against 0.1 mM Na_2EDTA for 24 h and then water for 24 h. After this, oligomers were dissolved in 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM Na_2EDTA at pH 7.0 and lyophilized to dryness twice with 99.96% D_2O and one final time with 99.996% D_2O. The oligomers were then dissolved under dry argon in 99.996% D_2O for nonexchangeable proton and phosphorus spectra. The final strand concentrations were approximately 3.5 and 2.5 mM for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$, respectively. For exchangeable proton spectra, the oligomers were dissolved in 90% H_2O/10% D_2O and the final strand concentrations were approximately 1.5 and 0.5 mM, respectively.

NMR spectra were collected on Varian VXR-500S and Varian INOVA-500 spectrometers and processed with Varian and Felix software (MSI, Inc). Two-dimensional spectra were recorded in phase-sensitive mode by the States−Haberkorn method (*26*). Proton and phosphorus chemical shifts were referenced to the internal EDTA and the phosphate in the buffer, respectively.

The one-dimensional exchangeable proton spectra were acquired with a binomial 1:3:3:1 pulse sequence (*27*) for H_2O peak suppression. The sweep width was 12 000 Hz. Pulse delays were calculated to achieve a signal-to-noise maximum at 12 ppm, which is in the imino proton resonance region. One-dimensional nuclear Overhauser effect (NOE) experiments were performed by irradiation for 1−2 s at low level power. To minimize spillover effects, the off-resonance irradiation field of the control spectrum was offset from the resonance frequency potentially exhibiting spillover by the difference of the on-resonance saturation frequency and the resonance frequency potentially exhibiting spillover.

NOESY spectra in H_2O were acquired at 150 ms mixing times at 0 or 5 °C. A total of 200 free induction decays (FIDs) were acquired with 1K complex points and 10 000 Hz spectral width. For each FID, 32−64 scans were averaged. The read pulse was a symmetrically shifted soft pulse with excitation maximum at the imino proton resonance region (*28*). Data were apodized with phase-shifted sine-bell squared functions.

NOESY spectra in D_2O were acquired at 60, 100, 150, 250, 350, and 400 ms mixing times at 28 or 32 °C. A total of 256−512 FIDs were acquired with 2K complex points

and 5000 Hz spectral width. For each FID, 32−128 scans were averaged. In some spectra, the residual HDO resonance was attenuated by low-power presaturation during the recycle delay, which was set to 2.5 s. Data were apodized with phase-shifted sine-bell squared functions.

Double quantum filtered correlation spectroscopy (DQF-COSY) spectra were recorded over spectral widths of 2000 and/or 4000 Hz at 25 or 30 °C. For each spectrum, 512 FIDs were collected with 2K complex points and 96 scans per transient. The recycle delay was set to 2.5 s. Data were zero-filled to 4K complex points in the t2 dimension and apodized with phase-shifted sine-bell squared functions.

$^{1}H-^{31}P$ heteronuclear correlated (HETCOR) spectra were acquired with the pulse sequence described by Sklenar et al. (*29*) with spectral widths of 1000 or 2000 Hz in the ^{1}H dimension and 2000 or 1200 Hz in the ^{31}P dimension. A total of 64−128 FID's with 200 scans/FID were collected in 1K or 2K complex points. The recycle delay was set to 2.0 or 2.5 s. Data were apodized with phase-shifted sine-bell squared functions.

Generation of Distance Restraints. Interproton distance restraints between nonexchangeable protons were obtained from the intensities of their NOE cross-peaks in 2D NOESY spectra in D_2O at 60, 100, and 150 ms mixing times and are available as Supporting Information. The volumes of the pyrimidine H5−H6 cross-peaks, which correspond to a distance of 2.45 Å, were averaged and used as the internal reference. Distances were usually given a limit of ±20%. Partially overlapped or noisy peaks were given a limit of ±30%. Severely overlapped or very noisy peaks were given a limit of ±40% or an upper limit distance of 6 Å. A lower limit distance of 1.6 Å was given if error limits allowed a smaller distance. If two peaks were almost completely overlapped, the peaks were treated as a single peak. The calculated lower limit was used as the lower limit for each of the distances. Hydrogen-bonded atoms of the Watson−Crick base pairs were given the following restraints in angstroms: for G-C pair, 2.81−3.01 for GO6 to CN4, 2.85−3.05 for GN1 to CN3, and 2.76−2.96 for GN2 to CO2; for A-U pair, 2.72−2.98 for AN1 to UN3 and 2.85−3.05 for AN6 to UO4; for G·U pair, if used, 2.76−3.05 for GO6 to UN3 and 2.76−3.05 for GN1 to UO2 (*30, 31*).

Generation of Torsion Angle Restraints. In principle, all nonphosphate backbone torsion angles can be constrained with semiquantitative estimates of *J* coupling constants obtained from DQF-COSY and ^{1}H-^{31}P HETCOR experiments. The appropriate Karplus equations [$J_{H-C-C-H} = 10.2 \cos^2\theta - 0.8 \cos\theta$ (*32, 33*) and $J_{H-C-O-P} = 15.3 \cos^2\theta - 6.1 \cos\theta + 1.6$ (*34*)] used to convert the *J*-coupling constants to torsion angles have up to four values. By use of multiple couplings that depend on the same torsion angle, it is often possible to determine a torsion angle range consistent with the data. All torsion angles were given broad limits. The determination of the torsion angles is described below.

The phosphate torsion angles, α (O3′−P−O5′−C5′) and ξ (C3′−O3′−P−O5′), cannot be determined directly from *J* coupling data but can be somewhat restrained indirectly from ^{31}P chemical shifts. A trans conformation for either α or ξ around a phosphorus will result in a large downfield shift of its resonance (*35*). Therefore, if a resonance of ^{31}P is within the range expected for the A-form conformation, the trans conformation for α and ξ angles can be excluded. Recently, Rife et al. (*36*) proposed a slightly more sophisticated model to get the restraints for α and ξ angles from ^{31}P chemical shifts. All ^{31}P resonances in this study are within 1 ppm of each other and resonate as expected for the A-form conformation. Thus the α and ξ angles were restrained to 0° ± 150°.

The dihedral angles β (P−O5′−C5′−C4′) were restrained from semiquantitative estimates of $^{3}J_{P-H5'}$ and $^{3}J_{P-H5''}$. These $^{31}P-^{1}H$ coupling constants were estimated from HETCOR experiments and were either absent or very weak, as expected for regular A-form helices with β in the trans (180° ± 30°) conformation (*37*).

The dihedral angles γ (O5′−C5′−C4′−C3′) were restrained from semiquantitative estimates of the $^{3}J_{H4'-H5'}$ and $^{3}J_{H4'-H5''}$ coupling constants obtained from DQF-COSY experiments. In regular A-form helices, γ is in the gauche^{+} conformation and both couplings are very small, whereas in the trans or gauche^{-} conformations, either of the couplings is large. The coupling constants $^{3}J_{H4'-H5'}$ and $^{3}J_{H4'-H5''}$ were often very small or absent and therefore the corresponding torsion angles were restrained to 60° ± 30°. If any of the coupling constants were large, however, γ was not restrained.

The dihedral angles ε (C4′−C3′−O3′−P) were restrained from semiquantitative estimates of the $^{3}J_{H3'-P}$ coupling constants from HETCOR experiments. Strong sequential correlation of (*n*−1)C3′−(*n*)P was always observed. This is consistent with trans and gauche^{-} conformations. The gauche^{-} conformation, however, is usually associated with the C2′-endo sugar pucker as predicted from steric considerations (*38*). Nucleotides with C3′-endo sugar pucker and gauche^{-} conformation have never been observed and thus are considered forbidden (*39*). Therefore, if the sugar pucker was clearly C3′-endo, as indicated by the weak or absent H1′−H2′ coupling constant, the dihedral angle ε was restrained as trans (−130° ± 30°).

The sugar pucker defines the final torsion angle, δ (C5′−C4′−C3′−O3′). It can be restrained by the magnitude of the $^{3}J_{H1'-H2'}$ scalar coupling. For C3′-endo sugar pucker, the coupling constant is small (<3 Hz), resulting in a weak or absent H1′−H2′ cross-peak in the DQF-COSY spectrum. For C2′-endo sugar pucker, the coupling constant is large (>8 Hz), resulting in a large cross-peak. The sugar pucker can also be confirmed by the H3′−H4′ coupling constant, since for C3′-endo sugar pucker, $^{3}J_{H3'-H4'}$ is also large (*40*).

Structure Modeling. The structure modeling was carried out with Biosym's Discovery program package running on a Silicon Graphics ONYX workstation. Calculations used the Amber95 force field (*41*). A cutoff value of 12 Å was used for van der Waals interactions. The cell multiple method (Discover 95.0/3.00 user guide) was used to treat electrostatic interactions. A distance-dependent dielectric constant of 2*r* was used to mimic solvent dielectric effects. Flat-bottomed quadratic penalty terms for distances and torsion angle restraints with force field constants of 25 kcal/(mol·Å^2) and 25 kcal/(mol·rad^2), respectively, were added to the force field.

In all of our structure modeling, the full scaled van der Waals and electrostatic interactions of the AMBER95 force field (*41*) were used along with the NMR restraints. There are discussions in the literature about the treatments of electrostatic interactions (*36, 42, 43*). It is believed that inclusion of the electrostatic interactions in the structural modeling of E73, an RNA oligonucleotide containing the

sarcin/ricin loop (SRL) from rat 28S rRNA, made the final structure more similar to the crystal structure of the same sequence (*36*). In our modeling, inclusion of electrostatic terms was necessary to generate structures that satisfied the NMR restraints and belonged to the A-form family. The effects of the charges on the phosphate group were also investigated and were found to have little effect when the charges were between −0.6 and −1.0. This includes the value of −0.7 found to be optimal by Rife et al. (*36*). No constraint was applied to force hydrogen bonding between G and U in the G·U pairs during the calculation. We attempted to force two hydrogen bonds in the G·U pair for r(GGCGUGCC)$_2$ with a force field constant of 100 kcal/(mol·Å^2) and had essentially the same structure.

The following were used as starting structures: (1) A-form duplex, (2) B-form duplex, (3) duplex with two separated, A-form single strands, and (4) a duplex with an A-form single strand and a B-form single strand. Ten structures were usually created from each starting structure.

Structure modeling was accomplished by restrained molecular dynamics and energy minimization. First, 400 steps of energy minimization were performed to bring the molecule into a reasonable conformation. Second, the dynamics were done by jumping the temperature to 1000 K and then ramping it down gradually to 300 K in 5−50 ps. No substantial difference was found between longer and shorter dynamics runs. After this, restrained energy minimization was performed until either the calculation reached the small derivative of 0.01 Å^{-1} kcal/mol set in the structure modeling protocol or the calculation repetition number reached 40 000. Finally the restraints were turned off and a new round of energy minimization was performed. This last step had a negligible effect on the structures.

Thermodynamic Measurements. Oligonucleotides were melted in a buffer system of 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na$_2$EDTA at pH 7.0 on a Gilford 250 spectrometer with a heating rate of 1 °C/min. The extinction coefficient was calculated according to the nearest-neighbor model (*44*, *45*) by assuming mU and iG equal to U and G, respectively. Thermodynamic parameters for duplex formation were calculated by two methods: (1) by fitting the shape of each melting curve to a two-state model (*46*) and (2) by plotting the reciprocal of the melting temperature, T_M, versus ln C_T, where T_M and C_T are melting temperature in kelvins and total strand concentration, respectively (*47*).

Contour Map of Potentials and Electrostatic Interaction Calculations. Electrostatic interaction calculations were based on Coulomb's law with a dielectric constant of 1. The partial charges used in all the calculations were taken from the Amber95 force field (*41*) and only the atoms on the bases and the C1′ atoms were taken into consideration.

RESULTS

Assignment of Exchangeable Proton Resonances. Figure 1 shows imino proton spectra for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$. Assignments were obtained from a series of 1D and 2D experiments in H_2O by standard procedures (*40*) (Supporting Information). As is commonly observed, the imino protons of the G·U pair are shifted upfield, with GH1 around 11 ppm, further upfield shifted than UH3 (*48*).

For r(GGCGUGCC)$_2$, no NOEs are observed between the resonances of the G·U pair and its adjacent G-C pair. This is expected because a G·U pair does not typically stack to the 5′ side of the G (*49*); thus the distance from G4H1 to G6H1 is expected to be about 5 Å. Therefore, the imino proton assignments of the Watson−Crick pairs were obtained from a 2D NOESY experiment in H_2O, in which the read pulse was replaced by a symmetrically shifted soft pulse (*28*). Following the connectivity for a G-C pair, i.e., from GH1 to H5 of the complementary C (*40*), the resonances at 13.40 and 12.98 ppm were assigned to G2H1 and G6H1, respectively (Figure 2). The chemical shifts assigned to G1, G2, and G6 for r(GGCGUGCC)$_2$ are similar to those assigned for the corresponding nucleotides in r(GGC*AG*GCC)$_2$ (*50*), as expected.

There are weaker NOEs between G4H1 and U5H3 for r(GGCGUGCC)$_2$ compared with the NOEs between G5H1 and U4H3 for r(GAGUGCUC)$_2$ (Supporting Information). As shown in Figure 2, G4H1 and U5H3 of r(GGCGUGCC)$_2$ have much larger cross-peaks with water (5.08 ppm) than other imino resonances in this duplex. In contrast, for r(GAGUGCUC)$_2$, all the imino protons have weak cross-peaks with water (data not shown). Surprisingly, the NOESY spectrum of r(GAGUGCUC)$_2$ in H_2O (data not shown) contains strong cross-peaks of amino and imino protons of G3. These cross-peaks can be seen even at 25 °C. These cross-peaks suggest strong hydrogen bonds and correlate with the favorable stability of 5′-GUGC-3′ in this duplex.

Assignment of Nonexchangeable Protons and Phosphorus. Chemical shifts of all assigned proton and phosphorus resonances are given in Supporting Information. The base (H8/H6/H2/H5) and sugar H1′ protons were assigned on the basis of standard connectivities (*40*) from NOESY spectra shown in Figure 3 for r(GAGUGCUC)$_2$ and in Figure 4 for r(GGCGUGCC)$_2$. The assignments were confirmed by the H5−H6 cross-peaks, which were confirmed by the splitting of these cross-peaks in NOESY and DQF-COSY spectra. The weak cross-peaks between (*n*)H5 and (*n*−1)H8/H6 also confirmed the assignments. The resonance in Figure 3 at 7.57 ppm was assigned to A2H2 from cross-peaks to G3H1′ and C8H1′ on the opposite strand. Overall, the cross-peak patterns and intensities in the base to H5/H1′ region for both duplexes verify that the structures are generally A-form and all nucleotides have anti glycosidic torsion angles (*51*).

The sugar protons (H2′, H3′, H4′, and some of H5′ and H5″) were assigned from a combination of NOESY and COSY spectra (*40*, *52*). Almost all H1′−H2′ cross-peaks in this study are missing in the DQF-COSY spectra, suggesting that most sugars are in C3′-endo conformations. One exception is the 3′ terminal nucleotide, indicating its sugar has a fairly high fraction of C2′-endo conformation. The usually large cross-peaks of H3′-H4′ also reveal the predominant C3′-endo sugar pucker for all nucleotides.

Three-Dimensional Models. The restraints used in the structural modeling for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$ are supplied as Supporting Information and are summarized in Table 2. Ensembles of converged structures are shown in Figures 5 and 6. The average RMSDs of all the structures with the average structure are 0.20 and 0.41 Å for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$, respectively. Individual average RMSDs for each base pair are also shown in Table 2. The all-atom pairwise RMSD between the two G·

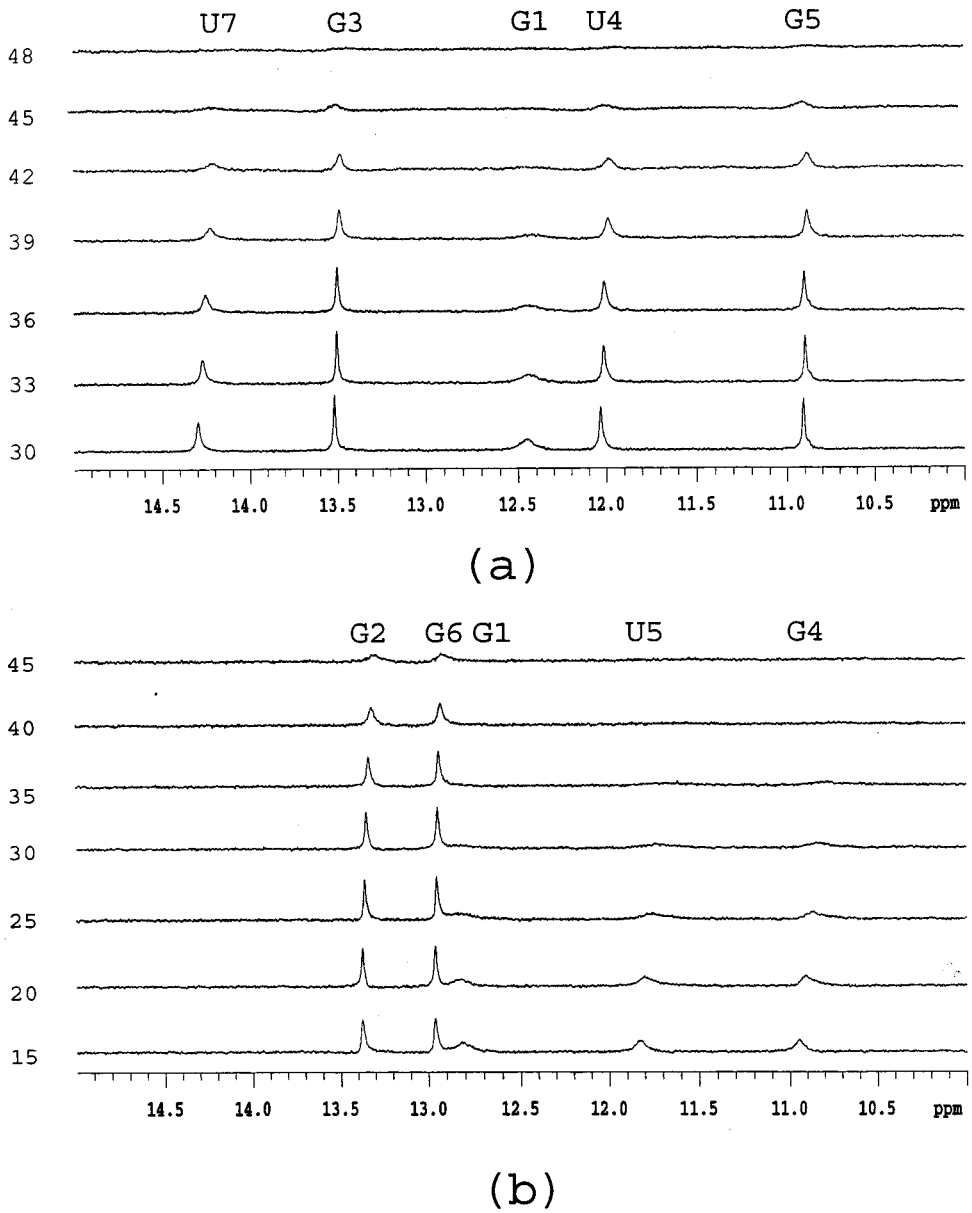

FIGURE 1: Imino proton spectra as a function of temperature for (a) r(GAGUGCUC)$_2$ and (b) r(GGCGUGCC)$_2$. Assignments are shown on top of the plot. In each plot, the temperatures (°C) are designated to the left. For clarity, the temperature step in panel a is smaller than in panel b.

U pairs of r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$ is 0.93 Å. The value is 0.68 Å if only the bases are compared. The helical parameters for the two duplexes, as well as standard A-form and B-form values, are reported in Supporting Information. All these parameters are close to A-form. The most definitive criterion to distinguish A-form from B-form helices is the displacement from the helical axis (≈−4 Å for A-form but ≈0.8 Å for B-form) (*38*, *42*). Examination of this helical parameter clearly reveals these structures are A-form. The predominance of C3′ endo sugar puckers is also consistent with A-form structure. Evidently, the insertion of the G·U pairs has little effect on the global geometry of the duplexes. The effects are localized in the G·U pair regions. Thus the structures are consistent with the nearest neighbor model commonly used for predicting RNA stability (*53*, *54*).

Despite global similarities to A-form, examination of the helical parameters reveals some special structural features. For example, in r(GAGUGCUC)$_2$, the helical twist angles between adjacent G-C and G·U pairs are small, and the helical twist angle between the two G·U pairs is large. In r(GGCGUGCC)$_2$, however, the helical twist angles between adjacent G-C and G·U pairs are large, and the twist angle between the two G·U pairs is small. This is consistent with previous observations on G·U and G·T pairs in RNA and A-form DNA, respectively (*55*−*61*). This is because the G·U or G·T pair tends to overlap extensively with the pair 3′ to its G, whereas overlap is minimum with the pair 5′ to its G (*62*). Thus the Gs of the G·U pairs in r(GAGUGCUC)$_2$ overlap extensively with the adjacent C in the same strand and also cross-strand overlap with each other (see Figures 7b,c), whereas in r(GGCGUGCC)$_2$ the ring of the G in a G·U pair overlaps almost completely with the adjacent U in the same strand (see Figure 7f). This is a common feature for tandem G·U pairs in RNA and tandem G·T pairs in A-form DNA, in both solution and crystal structures (31, *55*−*59*, *61*). A common consequence of destacking between the two base pairs in a step (Figure 7c,e) is that the interstrand P−P distance is usually shorter than standard A-form.

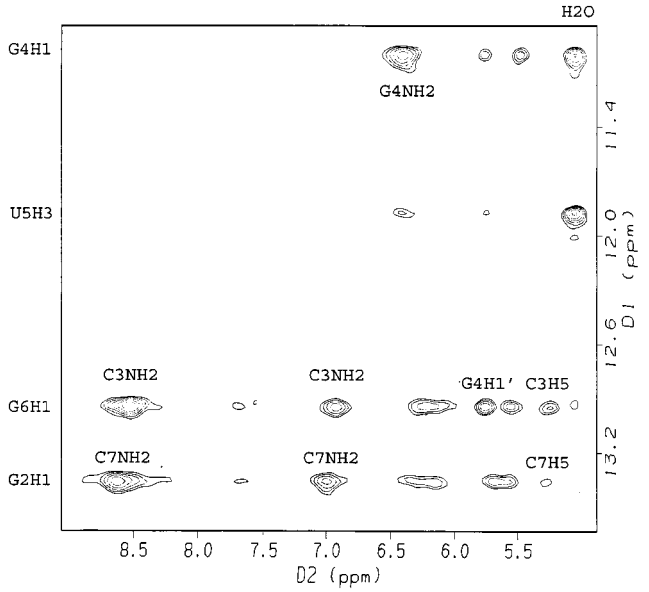

FIGURE 2: Portion of 150 ms 2D NOESY spectrum in H_2O for r(GGCGUGCC)$_2$ at 0 °C showing the imino to amino proton region. D1 dimensional assignments are designated to the left. Some important assignments are labeled. Note the large cross-peaks to H_2O in the D2 dimension from G4H1 and U5H3.

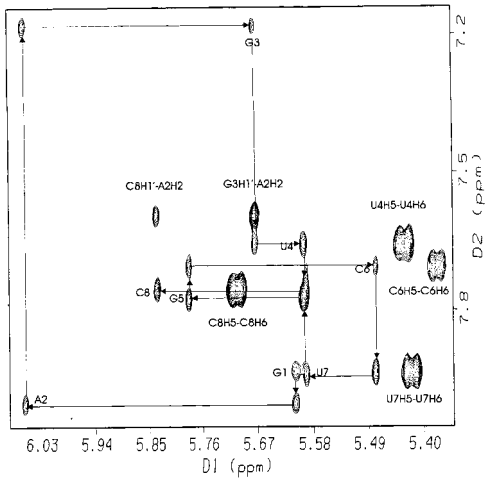

FIGURE 3: NOESY spectrum (350 ms) of r(GAGUGCUC)$_2$ at 32 °C in the H8/H6/H2 to H5/H1′ regions. Intranucleotide base to sugar H1′ cross-peaks are labeled with the base number.

The striking difference between the two NMR structures is the hydrogen bonding patterns of the G·U pairs (Figures 8b,c). In r(GAGUGCUC)$_2$, two hydrogen bonds are observed as expected for the G·U pair. In r(GGCGUGCC)$_2$, however, only one three-centered hydrogen bond is seen. The distance between G4O6 and U5H3 in r(GGCGUGCC)$_2$ is about 3.4 ± 0.1 Å, too far to form a hydrogen bond. A three-centered hydrogen bond has been shown to be weaker than a two-centered hydrogen bond (*63*). Fewer hydrogen bonds in the 5′-CGUG-3′ motif can rationalize its decreased stability relative to 5′-GGUC-3′, 5′-GUGC-3′, and 5′-CUGG-3′.

Shown in Figure 9 is the juxtaposition of a G-C pair and the G·U pairs in r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$, in which the Gs are superimposed. The distances of C1′ and N1 of U relative to C1′ and N1 of C are 1.36 and 1.34 Å, respectively, for r(GAGUGCUC)$_2$. The distances are 1.09 and 0.55 Å, respectively, for r(GGCGUGCC)$_2$, however. Clearly, the G·U pair in r(GAGUGCUC)$_2$ deviates more from

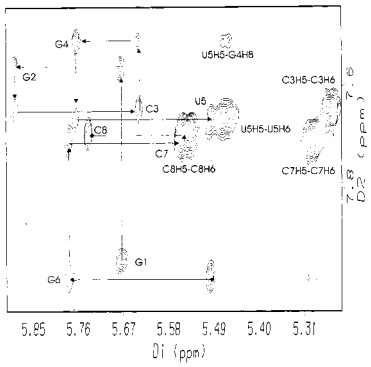

FIGURE 4: NOESY spectrum (150 ms) of r(GGCGUGCC)$_2$ at 28 °C in the H8/H6/H2 to H5/H1′ regions. Intranucleotide cross-peaks are labeled with the base number.

Table 2: Number of Restraints and Average RMSD[a] (Å) between the Converged and the Average Structures for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$

no. of restraints in structural modeling		
	r(GAGUGCUC)$_2$	r(GGCGUGCC)$_2$
distance	182	166
dihedral	100	100
hydrogen bonding	16	18

average RMSD for individual base pairs and the whole molecules			
r(GAGUGCUC)$_2$		r(GGCGUGCC)$_2$	
G1-C8′ [b]	0.30	G1-C8′	0.60
A2-U7′	0.12	G2-C7′	0.48
G3-C6′	0.09	C3-G6′	0.25
U4·G5′	0.16	G4·U5′	0.22
G5·U4′	0.11	U5·G4′	0.24
C6-G3′	0.08	G6-C3′	0.22
U7-A2′	0.12	C7-G2′	0.34
C8-G1′	0.35	C8-G1′	0.55
whole molecule	0.20	whole molecule	0.41

[a] Average RMSD is given in angstroms. [b] Prime designates the opposite strand.

Watson−Crick geometry than the G·U pair in r(GGCGUGCC)$_2$.

Temperature Dependence of Imino Proton Resonances. The different hydrogen-bonding patterns for G·U pairs in the models for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$ suggested there could also be different patterns for exchange of imino protons with water. To test this, imino proton spectra were measured as a function of temperature (Figure 1). For r(GAGUGCUC)$_2$, the imino proton resonances of the G·U pairs disappear at almost the same temperature as those of Watson−Crick pairs (Figure 1a). For r(GGCGUGCC)$_2$, however, the imino proton resonances of the G·U pairs disappear at 10−15 °C lower temperature than those of Watson−Crick pairs (Figure 1b). This is consistent with less favorable hydrogen bonding in the G·U pairs of r(GGCGUGCC)$_2$.

NMR structural models have been reported previously for two duplexes with the 5′-GU-3′ motif. One, r(GGAGUUCC)$_2$, has a relatively unstable 5′-GU-3′ motif, while the other, r(GAGGUCUC)$_2$, has a relatively stable 5′-GU-3′ motif (Table 1). Both NMR models had G·U pairs with two hydrogen bonds (*57*, *58*). For r(GGAGUUCC)$_2$, He et al. (*20*) reported that the imino protons for the G·U pairs disappear between 15 and 30 °C, whereas those for Watson−Crick pairs are still present at 30 °C. Measurements of imino

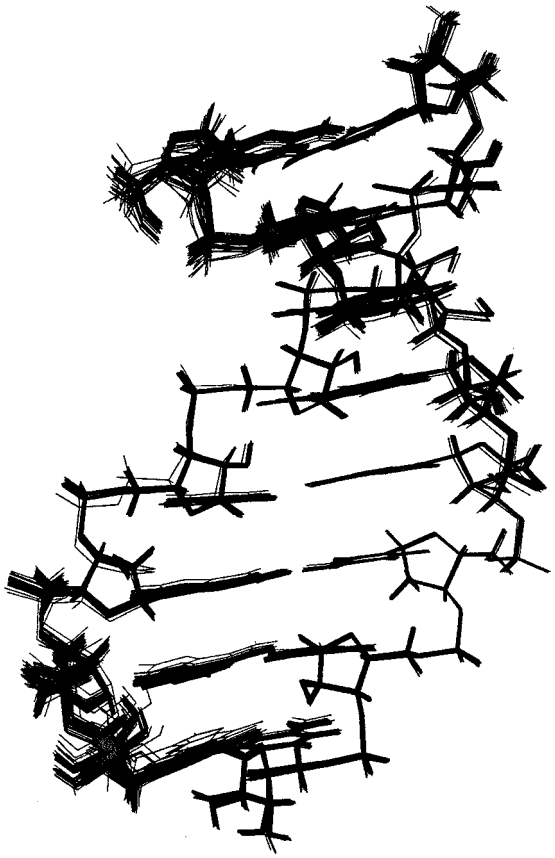

FIGURE 5: Converged structures of r(GAGUGCUC)$_2$.

FIGURE 6: Converged structures of r(GGCGUGCC)$_2$.

proton spectra of r(GAGGUCUC)$_2$ as a function of temperature, however, show that the resonances for imino protons in G·U and Watson−Crick pairs disappear at almost the same temperature (Table 3).

The protocol used previously (*58*) for structure determination of r(GGAGUUCC)$_2$ and r(GAGGUCUC)$_2$ differed from the protocol used here for r(GGCGUGCC)$_2$ and r(GAGUGCUC)$_2$. For example, the previous protocol used the AMBER86 force field (*64*), and the force constants of the van der Waals and electrostatic interactions were scaled during molecular dynamics. To see if differences in the protocols for generating structural models lead to the difference in G·U pair geometry, we used the protocol described here for r(GGCGUGCC)$_2$ with the published NMR restraints to model the structures for r(GGAGUUCC)$_2$ and r(GAGGUCUC)$_2$. In the new model for r(GGAGUUCC)$_2$, there is only one hydrogen bond in each G·U pair, and the violations of NMR restraints are similar to those observed with the published model. Evidently, the NMR data are consistent with either a one- or two-hydrogen-bond model, raising the possibility of an equilibrium between the two structures. In the new model for r(GAGGUCUC)$_2$, however, there are still two hydrogen bonds in each G·U pair.

For comparison, imino proton spectra as a function of temperature were also measured for r(GGCUGGCC)$_2$ and r(GGUUGACC)$_2$. As shown in Table 3, the resonances from the imino protons of G·U and Watson−Crick pairs disappear at essentially the same temperature for both duplexes. This is consistent with the favorable free energy increments of −4.2 and −2.2 kcal/mol, respectively, determined for the tandem G·U motifs in these duplexes (Table 1).

Thermodynamics of G·mU (N3-Methyluridine) and (Isoguanosine) iG·U Pairs. The above results suggest that the sequence dependence of symmetric tandem G·U pair stability is correlated with a difference in the number of hydrogen bonds in the G·U pair, particularly whether there is a hydrogen bond between UH3 and GO6. Formation of this hydrogen bond can be prevented by substituting mU for U or iG for G (see Figure 8d,e). The thermodynamics of formation for RNA duplexes with mU or iG substitutions, as well as two Watson−Crick duplexes, are listed in Table 4. The stabilities for duplexes with and without the substitutions are compared in Table 5. The change in stability for each substitution, $\Delta\Delta G^\circ_{37}$, along with the presumed number of hydrogen bonds in each G·U pair in the parent duplex are also listed in Table 5. When mU and iG substitutions are made in motifs thought to have one hydrogen bond per G·U pair, the change in stability is less than when substitutions are made in motifs having two hydrogen bonds per G·U pair.

Electrostatic Calculations. On the basis of previous NMR and thermodynamic studies (*57*, *58*, *65*), it was proposed that electrostatics play an important role in RNA structure and stability. Figure 10 shows contour plots of electrostatic potentials for the base pair steps between adjacent G-C and G·U pairs for the duplexes (GAGUGCUC)$_2$ and (GGCGUGCC)$_2$. The numbers beneath each plot are the electrostatic energies between the base pairs as calculated with Coulomb's law. The geometries in the left two plots were taken from the NMR-models reported here, whereas the right two plots were constructed by flipping the adjacent G-C pairs in the NMR models. By comparing the overlap patterns and

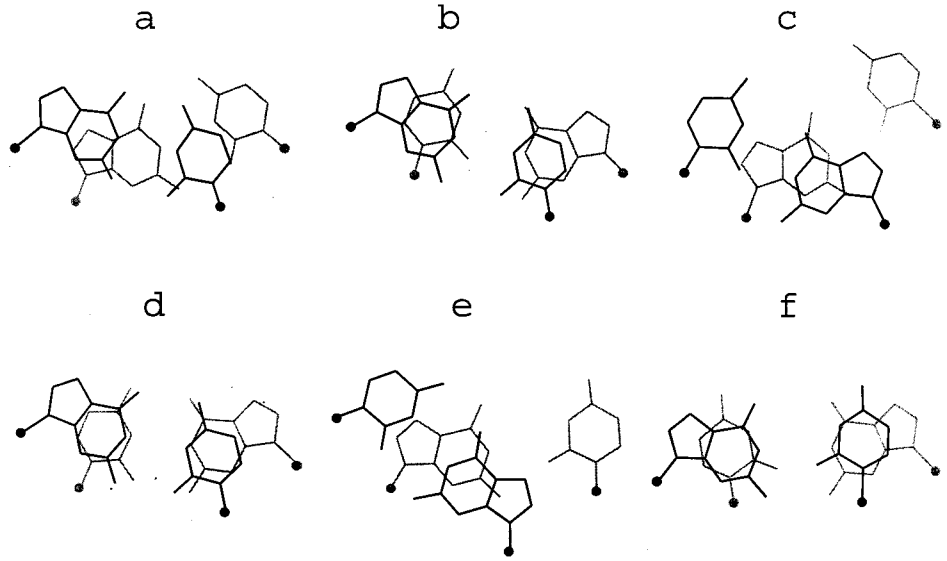

FIGURE 7: Stacking patterns of (a−c) r(5′GAGUGCUC3′)$_2$ and (d−f) r(5′GGCGUGCC3′)$_2$ showing some of the internal base steps: (a) 5′-AG-3′, (b) 5′-GU-3′, (c) 5′-UG-3′, (d) 5′-GC-3′, (e) 5′-CG-3′, (f) 5′-GU-3′, where underlined bases are in G·U pairs. The left base pairs in each step are shown in black. The solid dots designate the C1′ atoms.

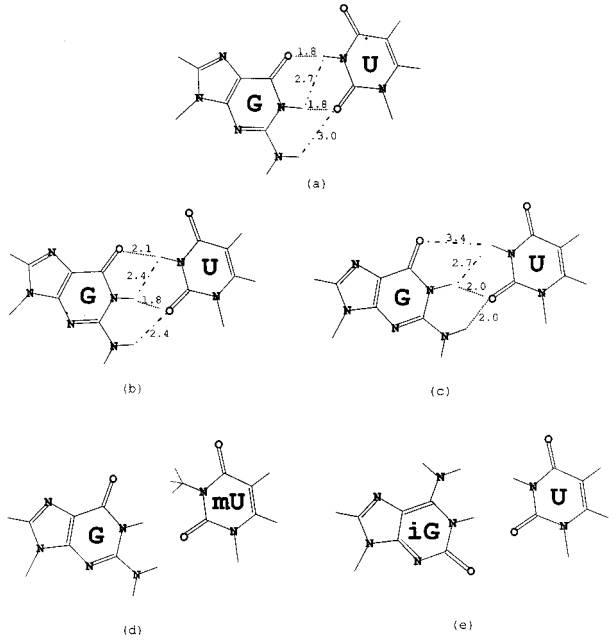

FIGURE 8: G·U pair geometry in (a) crystal structure of r(GUAUGUA)dC (*59*); (b) NMR model of r(GAGUGCUC)$_2$; (c) NMR model of r(GGCGUGCC)$_2$, in which only one hydrogen bond is seen; (d) G·mU pair; (e) iG·U pair. In the NMR models, hydrogen bonds are shown by dotted lines. Some other distances are shown in angstroms by broken lines.

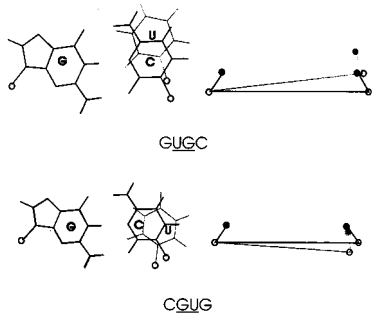

FIGURE 9: Juxtaposition of G·U and G-C pairs designated by gray and black structures, respectively. Gs in each pair are superimposed. The geometry of each G·U pair is taken from the NMR model of the motif listed below the pair. The geometry of the G-C pair is taken from G2-C7 of the NMR model for r(GGCGUGCC)$_2$. ●, N9/N1; ○, C1′.

Table 3: Temperatures at Which Imino Proton Resonances Disappear in Different Duplexes in NMR Buffer[a]

sequence	GU	Watson−Crick	T_m[d]
Duplexes with Stable G·U Motifs			
GAGUGCUC	≈50	≈50	50
GGCUGGCC	≈66	≈68	65
GGUUGACC	≈51	≈51	51
GGAUGUCC[b]	>50	>50	52
GAGGUCUC	≈45	≈43	47
Duplexes with Unstable G·U Motifs[c]			
GGCGUGCC	≈35	≈50	56
GGAGUUCC[b]	15−30	>30	39

[a] Sample concentrations range from 0.5 to 3.5 mM and the NMR buffer contains 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM EDTA at pH 7.0. Temperatures are given in degrees Celsius. [b] Reference 20. [c] It has previously been reported (*20*) that imino resonances for G·U pairs also disappear at lower temperatures than those for Watson−Crick pairs for the duplexes r(GUCGUGAC)$_2$ and r(CCUGUAGG)$_2$. [d] Melting temperature calculated at 10^{-4} M strand concentration in 1 M NaCl from thermodynamic parameters determined by optical melting.

calculated values in the same row, one finds that flipping the adjacent base pair causes little electrostatic change for the 5′-GUGC-3′ geometry, whereas a large effect is seen for the 5′-CGUG-3′ geometry. This calculation suggests that 5′-GUGC-3′ and 5′-CUGG-3′ will assume similar geometries with similar energetics, whereas 5′-CGUG-3′ and 5′-GGUC-3′ will not. Similar calculations on the NMR-derived structural models for 5′-GGUC-3′, from the literature (*57*) or from the protocol described here, give more favorable electrostatic interactions by 2.8 and 5.1 kcal/mol, respectively, for the 5′-GG-3′ steps relative to the 10.8 kcal/mol calculated for the model with one hydrogen bond per G·U pair in Figure 10. The sum of the hydrogen-bonding interactions and electrostatic components of the stacking interactions in 5′-GGUC-3′ also appears to be more favorable for the NMR models with two hydrogen bonds per G·U pair

Table 4: Thermodynamics of Formation of Self-Complementary RNA Duplexes

sequence (5′→3′)	$1/T_M$ versus ln(C_T) parameters $-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°_{37}$ (kcal/mol)	$-\Delta S°_{37}$ [(cal/(K·mol)]	T_M[a] (°C)	melt curve-fitting parameters $-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°_{37}$ (kcal/mol)	$-\Delta S°_{37}$ [(cal/(K·mol)]	T_M[a] (°C)
Duplexes with A-mU/G·mU Pairs								
GAG**mU**ACUC	3.1	52.7	159.9	22.6	3.6	47.4	141.4	23.8
GGCA**mU**GCC	6.6	54.3	153.7	42.5	6.8	57.7	164.1	42.9
GAGG**mU**CUC	3.7	51.5	154.2	25.4	3.8	49.8	148.1	25.8
GAG**mUG**CUC	2.8	44.6	134.8	17.9	2.7	50.9	155.5	19.6
GGAG**mU**UCC	3.6	59.3	179.6	26.7	3.9	56.1	168.3	27.4
GGA**mUG**UCC	2.5	63.5	196.6	22.1	3.1	54.2	164.8	22.9
GGCG**mU**GCC	7.1	59.0	167.2	44.7	7.3	60.6	172.0	45.2
GGC**mUG**GCC	6.6	59.6	171.2	41.6	6.5	51.7	145.9	42.2
Duplexes with iG·U Pairs								
GAGU**iG**CUC	5.3	50.0	144.4	34.4	5.2	58.2	170.8	34.7
GAG**iGU**CUC	5.2	51.6	149.8	33.8	5.1	57.1	167.8	33.9
GGAU**iG**UCC	5.2	53.7	156.2	34.5	5.3	61.1	180.2	34.9
GGA**iGU**UCC	4.8	68.8	206.5	33.0	5.0	61.7	182.7	33.6
GGCU**iG**GCC	8.1	71.0	202.5	48.2	8.0	66.3	187.8	48.5
GGC**iGU**GCC	7.9	60.4	169.2	48.9	8.0	64.3	181.3	48.8
Duplexes with only Watson−Crick Pairs								
GAGUACUC	9.7	70.0	194.3	55.9	10.2	79.5	223.3	55.9
GGCAUGCC	13.3	82.2	222.0	68.8	14.0	90.2	245.7	68.6

[a] Calculated for an oligomer strand concentration of 10^{-4} M.

Table 5: Comparison of Stability[a] of Self-Complementary Duplexes with and without *N*3-methyluridine or Isoguanosine Substitutions[b]

	G·U pairs sequence (5′ → 3′)	no. of HB[c]	$\Delta G°$	T_m[d] (°C)	G·mU/A-mU pairs sequence (5′→3′)	$\Delta G°$	T_m[d] (°C)	$\Delta\Delta G°$[e]	iG·U pairs sequence (5′→3′)	$\Delta G°$	T_m[d] (°C)	$\Delta\Delta G°$[e]
set I	GAG**UG**CUC	2	−9.4	51.6	GAG**mU**GCUC	−2.8	17.9	6.6	GAGU**iG**CUC	−5.3	34.4	4.1
	GAG**GU**CUC	2	−8.8	49.1	GAG**GmU**CUC	−3.7	25.4	5.1	GAG**iGU**CUC	−5.2	33.8	3.6
set II	GGC**UG**GCC	2	−13.1	65.9	GGC**mUG**GCC	−6.6	42.3	6.5	GGCU**iG**GCC	−8.1	48.2	5.0
	GGC**GU**GCC	1	−9.7	55.0	GGC**GmU**GCC	−7.1	44.7	2.6	GGC**iGU**GCC	−7.9	48.9	1.8
set III	GGA**UG**UCC	2	−8.4	49.0	GGA**mUG**UCC	−2.5	22.1	5.9	GGAU**iG**UCC	−5.2	34.5	3.2
	GGA**GU**UCC	1	−6.4	40.2	GGA**GmU**UCC	−3.6	26.7	2.8	GGA**iGU**UCC	−4.8	33.0	1.6
control	GGCAUGCC	2	−13.3	68.8	GGCA**mU**GCC	−6.6	42.5	6.7				
	GAGUACUC	2	−9.7	55.9	GAG**mU**ACUC	−3.1	22.6	6.6				

[a] $\Delta G°$ and $\Delta\Delta G°$ in kilocalories per mole at 37 °C. [b] The G·U/mU, A-U/mU and G·U/iG·U pairs are underlined, and the positions of substitutions are shown in bold. In each set of the duplexes, the only difference is the sequence of the G·U pairs. Values are from $1/T_M$ versus ln (C_T) plots. For sequences not listed in Table 4, reference are given in Table 1. [c] Number of hydrogen bonds proposed in a G·U or A-U pair. [d] Calculated for an oligomer strand concentration of 10^{-4} M. [e] $\Delta\Delta G°_{37}$ (kcal/mol) is the free energy difference between a substituted and regular sequence.

than for the model with one hydrogen bond per G·U pair. This may explain why 5′-GGUC-3′ has a structure that differs from 5′-CGUG-3′.

DISCUSSION

Much is not understood about the relationships between sequence, stability, structure, and function of RNA. Studies of small motifs in oligoribonucleotides are a first step toward understanding these relationships. It was reported previously that the thermodynamics of duplexes with symmetric 5′-GU-3′ motifs do not fit into the nearest-neighbor model (*20*). Nevertheless, if these duplexes are divided into two groups, one with 5′-GGUC-3′, and the other with 5′-CGUG-3′, 5′-UGUA-3′ and 5′-AGUU-3′, either group can fit into the nearest-neighbor model. This suggests that the 5′-GU-3′ motif may have different structures in different contexts. Nearest-neighbor analysis has shown that 5′-CGUG-3′ is about 3 kcal/mol less stable on average than the other symmetric tandem G·U pair motifs with G-C as the adjacent pairs: 5′-GGUC-3′, 5′-GUGC-3′, and 5′-CUGG-3′ (Table 1). We also noted that, in databases of known secondary structures (*1*, *2*), 5′-CUGG-3′ and 5′-GUGC-3′ occur 33 and 24 times, respectively, while 5′-CGUG-3′ and 5′-GGUC-3′ occur only three and two times, respectively.

There is evidence suggesting that the stability differences among duplexes containing symmetric tandem G·U pairs arise from local effects. For example, thermodynamic studies on these duplexes showed that the free energy increments for tandem G·U pair motifs are mostly independent of the base pair compositions for nonadjacent base pairs (Table 1). It is also found that introduction of symmetric tandem G·U pairs has little effect on the overall backbone geometry of the helices (*57*−*59*). Finally, imino protons of the relatively less stable tandem G·U motifs exchange with water at lower temperatures than the surrounding Watson−Crick base pairs, whereas those of the more stable tandem G·U motifs exchange with water at the same temperature as the surrounding Watson−Crick base pairs (Table 3).

NMR structural models are reported here for r(GAGUGCUC)$_2$ and r(GGCGUGCC)$_2$. These models can explain the

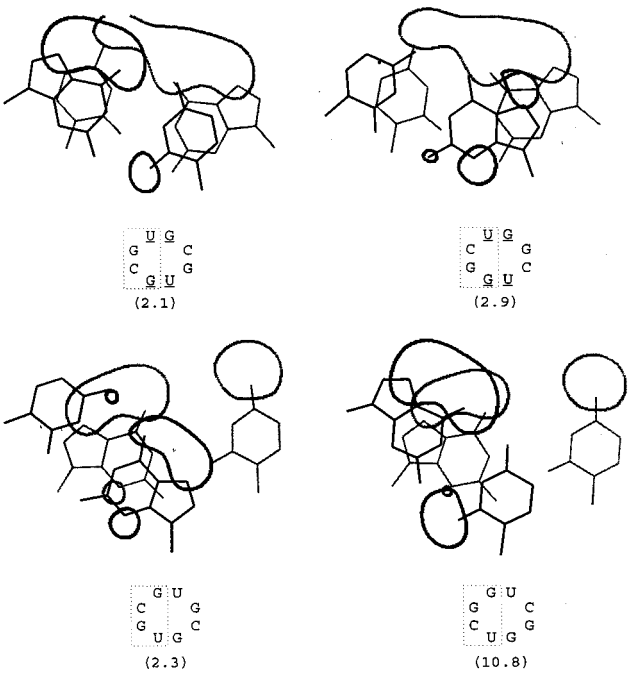

FIGURE 10: Stacking interactions for steps of 5′-GU-3′/3′-CG-5′ (upper left) and 5′-CG-3′/3′-GU-5′ (lower left) in NMR models and for steps of 5′-CU-3′/3′-GG-5′ (upper right) and 5′-GG-3′/3′-CU-5′ (lower right), modeled by flipping the G-C pairs of the NMR models. Contours encircle regions of highly negative potential and are drawn in the best plane through the middle of the stack of the two base pairs shown. The steps are boxed in the corresponding motifs shown beneath each plot, along with the electrostatic interaction energies (kilocalories per mole) calculated according to Coulomb's law, with partial charges from Cornell et al. (*41*). The larger the interaction energy, the less stable is the motif. The left base pairs in each motif are shown in black.

unusual sequence dependence of stability for symmetric tandem G·U motifs, and the temperature dependence of the imino proton resonances (Tables 1 and 3). The key suggestion is that there are different numbers of hydrogen bonds in the G·U pairs in different contexts. For r(GAGUGCUC)$_2$, two hydrogen bonds are seen in each G·U pair, i.e., GH1 to UO2 and GO6 to UH3 (Figure 8b). In contrast, only one three-centered hydrogen bond is seen for each G·U pair in r(GGCGUGCC)$_2$, i.e., GH1 and GH2 to UO2 (Figure 8c). Such a three-centered hydrogen bond has been shown to be weaker than a two-centered hydrogen bond (*63*). Although G·U pairs with different one-hydrogen-bond geometries are seen in other contexts (*31*, *42*, *66*), the one-hydrogen-bond scheme in an RNA helix has not been suggested previously.

Effects of N3-Methyluridine and Isoguanosine Substitution. The NMR models suggest that the stability differences between G·U motifs (Table 1) correlate with the number of hydrogen bonds in the G·U pair. In particular, UH3 forms a hydrogen bond with GO6 in more stable tandem G·U motifs, but does not in less stable motifs (Figure 8b,c). Replacing uridine (U) with *N*3-methyluridine (mU) guarantees that there is no imino proton to form a hydrogen bond with GO6 (Figure 8d). Thus the NMR models predict that the mU substitution should have a larger effect on duplex stability for the more stable tandem G·U motifs than for the less stable motifs. As shown in Table 5, the data fit this expectation. Motifs containing G·U pairs with two hydrogen bonds lost 6.0 ± 0.7 kcal/mol on average, while those with only one hydrogen bond lost 2.7 ± 0.1 kcal/mol on average. As a control, A-U pairs, each of which have two hydrogen bonds, lost 6.7 ± 0.5 kcal/mol, which is close to the loss for G·U pairs with two hydrogen bonds. The loss of stability for G·U pairs with single hydrogen bonds may be due to steric hindrance by the bulky methyl group, changes in solvation, or other factors.

Table 6: Comparison of Free Energy Increments for Symmetric Tandem G·U and A-U Motifs[a]

	central pairs					
adjacent base pairs	5′UG3′ [b] 3′GU5′	5′UA3′ [c] 3′AU5′	ΔΔ*G*[d]	5′GU3′ [b] 3′UG5′	5′AU3′ [c] 3′UA5′	ΔΔ*G*[d]
5′G 3′C	−4.9	−5.8	0.9	−4.1	−5.8	1.7
5′C 3′G	−4.2	−5.5	1.3	−1.1	−5.3	4.2
5′U 3′A	−2.6	−3.2	0.6	−0.3	−3.8	3.5
5′A 3′U	−1.9	−3.5	1.6	0.2	−3.0	3.2

[a] Kilocalories per mole at 37 °C. [b] Averaged if applicable. [c] Calculated from the nearest-neighbor parameters, e.g., Δ*G*°(GUAC) = Δ*G*°(GU) + Δ*G*°(UA) + Δ*G*°(AC). [d] ΔΔ*G* is the difference between G·U motifs and the corresponding A-U motifs.

The hydrogen bond between GO6 and UH3 of a wobble G·U pair also cannot form in an iG·U pair (Figure 8e). Thus the NMR models predict that the iG substitution should have a larger effect on duplex stability for the more stable tandem G·U motifs than for the less stable motifs. As shown in Table 5, the data also fit this expectation. Motifs containing G·U pairs with two hydrogen bonds lost 4.0 ± 0.8 kcal/mol on average, while those with only one hydrogen bond lost 1.7 ± 0.1 kcal/mol on average. The loss of stability for G·U pairs with a single hydrogen bond may be due to steric hindrance between iGNH2 and UH3, changes in solvation or other factors.

Nearest-Neighbor Analysis Is Also Consistent with the NMR Models. A wobble G·U pair has two hydrogen bonds, i.e., GH1 to UO2 and GO6 to UH3, and thus is expected to have similar stability as an A-U pair. Table 6 shows a comparison of the free energy increments of symmetric tandem G·U motifs and their A-U counterparts. The 5′-GGUC-3′ motif and all the motifs with 5′-UG-3′ have stabilities similar to those of their A-U counterparts, with the G·U motifs less stable by 1.2 kcal/mol on average. These motifs are all thought to have G·U pairs with wobble structures (Figure 8a,b). Motifs 5′-CGUG-3′, 5′-UGUA-3′, and 5′-AGUU-3′, however, are on average 3.6 kcal/mol less stable than their A-U counterparts. Since a hydrogen bond in a helix typically contributes about 1 kcal/mol to stability (*67*), the further decrease in stability of about 2.4 kcal/mol relative to A-U pairs is consistent with each of the two G·U pairs having only one hydrogen bond as suggested by the NMR and molecular modeling studies and by the nucleotide analogue replacement experiments.

Imino Proton NMR. The resonance of a hydrogen-bonded proton is usually shifted downfield relative to a non-hydrogen-bonded proton. The chemical shift of the hydrogen-bonded U4H3 in (GAGUGCUC)$_2$, however, is only about 0.3 ppm downfield of the non-hydrogen bonded U5H3 in (GGCGUGCC)$_2$, which resonates at 11.8 ppm. The small difference, however, is deceptive because the magnetic

environments of the G·U pairs in the two duplexes are different, and the chemical shift of UH3 in a G·U pair is quite context-dependent. For example, neither the U9H3 proton in a G·U pair of a UUCG tetraloop (*42*, *68*) nor the U74H3 proton in a G·U pair of *Escherichia coli* 5S rRNA (*31*) is hydrogen bonded, but they resonate at 11.9 and 10.8 ppm, respectively. Thus there is precedent for a non-hydrogen-bonded UH3 resonating at 11.9 ppm, as observed for U5H3 in $(GGCGUGCC)_2$.

The temperature dependence of the imino resonances of $r(GGCGUGCC)_2$ is somewhat unusual. The line width of an imino proton resonance can broaden if it is in conformational exchange or in rapid exchange with bulk water. Under extreme cases, the resonance disappears. It is a reasonable assumption that exchange with water effectively takes place only when the nucleotide is in an open state (*69*, *70*). Thus line broadening is common for terminal nucleotides, presumably because of terminal fraying. The stability of the duplex, however, is not usually affected since exchange can occur even if the base pair is only open a small fraction of time. It is also common that internal imino protons have different lifetimes for the open state (*71*). The difference is, however, not as drastic as observed here for $r(GGCGUGCC)_2$, in which the imino protons from the G·U pairs disappear more than 15 °C lower than those from Watson−Crick base pairs. Moreover, those protons with short lifetimes usually are close to termini or some structural disruptions, with the shortest lifetimes observed for protons closest to the structural disruptions. For $r(GGCGUGCC)_2$, the imino resonances of base pairs adjacent to G·U are strong, meaning there is no terminal or structural disruption near the G·U pairs. There are rare examples in which an imino resonance disappears 15 °C lower than the rest of the imino resonances (*72*), but the structural reasons and mechanisms have not been elucidated. What is interesting about the G·U imino proton resonances reported here is that there is a strong correlation between the dependence of the G·U imino proton resonances on temperature and their corresponding thermodynamic stabilities. That is, whenever the G·U pair is not stable, the imino proton resonances disappear more than 10 °C lower than those from Watson−Crick base pairs.

For $r(GGCGUGCC)_2$, the resonances of the U5H3 and G4H1 imino protons have almost the same line widths and disappear at almost the same temperature even though U5H3 is not hydrogen-bonded but G4H1 is. This is presumably because U5H3 is located in the narrow and deep RNA major groove, which prevents rapid exchange with water unless the G·U pairing is disrupted, at least a small fraction of the time. This can also explain why the line widths for all the imino protons are similar below a certain temperature. In fact, the exchange cross-peak of G4H1 with water is slightly weaker than that of U5H3 (Figure 2). The NOEs between G4H1 and U5H3 for $r(GGCGUGCC)_2$ are weaker than those between G5H1 and U4H3 for $r(GAGUGCUC)_2$ (see Supporting Information). This is consistent with greater exchange with water of the G·U pairs in $r(GGCGUGCC)_2$ and with the longer distance from G4H1 to U5H3 (2.7 ± 0.1 Å) in $r(GGCGUGCC)_2$ than from G5H1 to U4H3 (2.4 ± 0.03 Å) in $r(GAGUGCUC)_2$ (Figure 8b,c).

There are not many nonexchangeable protons that can be used to establish definitively the structure and therefore the hydrogen-bonding patterns of the G·U pairs. Moreover, because of their exchangeable nature, the imino and amino protons fail to provide accurate distance restraints. Nevertheless, structural modeling with molecular dynamics inevitably led to a three-centered hydrogen bond for the G·U pairs in $r(GGCGUGCC)_2$. Without employing molecular dynamics, however, 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. Thus we do not exclude the possibility of having a two-hydrogen-bonded structure for the G·U pairs in $r(GGCGUGCC)_2$. While both one and two hydrogen bond models are consistent with the NMR data, the energy calculations strongly favor the one hydrogen bond model. Moreover, this model is consistent with the thermodynamics measured for duplexes with G·U, G·mU, and iG·U pairs.

Comparison with Previous Results. A crystal structure containing 5′-GUGC-3′ has been reported at 2.5 Å resolution (*73*). Comparison of this structure with the NMR model reported here for 5′-GUGC-3′ reveals similar geometry. For example, each G·U pair in both structures has two hydrogen bonds that are somewhat bifurcated, which means that the displacement of the G in the G·U pair toward the minor groove leaves UO2 between GN2 and GH1 (Figure 8b).

Recently, Zhang et al. (*74*) reported NMR studies of ^{15}N chemical shifts for $r(GGCGUGCC)_2$ and $r(GGCUGGCC)_2$ with atom-specific isotopic labeling. The only difference between the two duplexes is the sequence of the tandem G·U pairs. The thermodynamic stabilities of the two duplexes and the corresponding motifs, 5′-CGUG-3′ and 5′-CUGG-3′, however, are very different (Tables 1 and 5). Zhang et al. (*74*) found an interesting difference between the two duplexes in the temperature dependence of the chemical shifts of N2 of the G in the G·U pairs. For $r(GGCGUGCC)_2$, a transition was seen with increasing temperature. In contrast, the chemical shift as a function of temperature was just a straight line with negative slope for $r(GGCUGGCC)_2$. It is possible that such a difference is caused by the different stacking of the G·U pairs in the two duplexes. The structural model obtained here, however, offers another explanation for this difference. In the model of $r(GGCGUGCC)_2$, G4NH2 is involved in a weak three-centered hydrogen bond (Figure 8c). Therefore, when the temperature rises and the duplex dissociates, this hydrogen bond breaks, giving rise to a transition in the chemical shift versus temperature plot. If each G·U pair in $r(GGCUGGCC)_2$ has two hydrogen bonds as in the model for $(GAGUGCUC)_2$, then the distance between U4O2 and G5NH2 is too far (2.4 Å) to form a hydrogen bond (Figure 8b). Thus a similar transition is not expected.

Complexity of G·U Pairs. G·U pairs comprise about 50% of the non-Watson−Crick base pairs and mismatches in a large set of known RNA secondary structures (*1*−*4*). G·U pairs can form either one or two hydrogen bonds. One structure of a G·U pair with one hydrogen bond is shown in Figure 8c. Another structure is seen in the models of the UUCG tetraloop (*42*) and of *E. coli* 5S rRNA (*31*). Evidently, the type of pairing is dependent on secondary and tertiary context. The complexity of G·U pairing will certainly affect studies of RNA structure and function.

Why Do G·U Pairs Adopt Different Geometries? The NMR and molecular modeling studies reported here suggest

that the G·U pairs in the 5′-CGUG-3′ motif have one hydrogen bond, whereas the G·U pairs in the 5′-GGUC-3′ motif have two hydrogen bonds. There are at least two factors that can favor the single hydrogen bond structure. As shown in Figure 9, less distorted juxtaposition of the G·U pair relative to A-form is required to accommodate a G·U pair with one hydrogen bond. Second, the electrostatic interactions between bases in the 5′-CGUG-3′ motif with one hydrogen bond may not be as unfavorable as those in the same motif with two hydrogen bonds for each G·U pair. Evidently, the favorable geometry for a G·U pair depends on a combination of electrostatic interactions with adjacent base pairs, hydrogen bonds within the G·U pair, and backbone distortions.

SUPPORTING INFORMATION AVAILABLE

Two tables of chemical shift assignments, four tables of the restraints used in the structural modeling, four tables of the helical parameters, two figures of 1D imino proton NOEs, and two figures of NOESY spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3502−3507.
2. Gutell, R. R., Gray, M. W., and Schnare, M. N., (1993) *Nucleic Acids Res. 21*, 3055−3074.
3. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
4. Gautheret, D., Konings, D., and Gutell, R. R. (1995) *RNA 1*, 807−814.
5. Crick, F. H. C. (1966) *J. Mol. Biol. 19*, 548−555.
6. Dirheimer, G., Keith, G., Dumas, P., and Westhof, E. (1995) Primary, Secondary, and Tertiary Structures of tRNAs, in *tRNA: Structure, Biosynthesis, and Function* (Soll, D., RajBhandary, U. L., Eds.) American Society for Microbiology Press: Washington, D.C.
7. Sprinzl, M., Horn, C., Brown, M., Loudovitch, A. and Steinberg, S. (1998) *Nucleic Acids Res. 26*, 148−153.
8. Pallanck, L., Pak, M. and Schulman, L. (1995) tRNA Discrimination in Aminoacylation, in *tRNA: Structure, Biosynthesis, and Function* (Soll, D. and RajBhandary, U. L., Eds.) American Society for Microbiology Press: Washington, D.C.
9. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) *Science 246*, 1135−1142.
10. Rould, M. A., Perona, J. J., and Steitz, T. A. (1991) *Nature 352*, 213−218.
11. McClain, W. H., and Foss, K. (1988) *Science 240*, 793−796.
12. Musier-Forsyth, K., Usman, N., Scaringe, S., Doudna, J., Green, R., and Schimmel, P. (1991) *Science 253*, 784−786.
13. McClain, W. H., Chen, Y.-M., Foss, K., and Schneider, J. (1988) *Science*, *242*, 1681−1684.
14. Gabriel, K., Schneider, J., and McClain, W. H. (1996) *Science 271*, 195−197.
15. Francklyn, C., and Schimmel, P. (1989) *Nature 337*, 478−481.
16. Strobel, S. A., and Cech, T. R. (1995) *Science 267*, 675−679.
17. Doudna, J. A., Cormark, B. P., and Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7402−7406.
18. Simpson, L., and Thiemann, O. H. (1995) *Cell 81*, 837−840.
19. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
20. He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., and Turner, D. H. (1991) *Biochemistry 30*, 11124−11132.
21. Strobel, S. A., Cech, T. R., Usman, N. and Beigelman, L. (1994) *Biochemistry 33*, 13824−13835.
22. Seela, F., and Wei, C. F. (1997) *Helv. Chim. Acta 80*, 73−85.
23. Zemlicka, J., and Sorm, F. (1965) *Collect. Czech. Chem. Commun. 30*, 1880−1889.
24. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S. and Usman, N. (1995) *Nucleic Acids Res. 23*, 2677−2684.
25. Usman, N., Ogilvie, K. K., Jingo, M.-Y., and Cedergren, R. J. (1987) *J. Am. Chem. Soc. 109*, 7845−7854.
26. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
27. Hore, P. J. (1983) *J. Magn. Reson. 55*, 283−300.
28. Smallcombe, S. H. (1993) *J. Am. Chem. Soc. 115*, 4776−4785.
29. Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., and Bax, A. (1986) *FEBS Lett. 208*, 94−98.
30. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−980.
31. Dallas, A., and Moore, P. B. (1997) *Structure 5*, 1639−1653.
32. Davies, D. B. (1978) *Prog. NMR Spectrosc. 12*, 135−225.
33. Hosur, R. V., Govil, G., and Miles, H. J. (1988) *Magn. Reson. Chem. 26*, 927−944.
34. Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C., and Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387−1405.
35. Gorenstein, D. G. (1984) *Phorphorus-31 NMR: Principles and Applications*, Academic Press, New York.
36. Rife, J. P., Stallings, S. C., Correll, C. C., Dallas, A., Steitz, T. A., and Moore, P. B. (1999) *Biophysical J. Part 1, 76*, 65−75.
37. Altona, C. (1982) *Rec. Trav. Chem. Pays-Bas 101*, 413−433.
38. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
39. Wijmenga, S. S., Mooren, M. M. W., and Hilbers, C. W. (1993) in *NMR of Macromolecules: A Practical Approach* (Roberts, G. C. K., Ed.) Oxford University Press, Inc., New York.
40. Varani, G., and Tinoco, I., Jr. (1991) *Q. Rev. Biophys. 24*, 479−532.
41. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
42. Allain, F. H.-T., and Varani, G. (1995) *J. Mol. Biol. 250*, 333−353.
43. Moore, P. B. (1999) in *Prebiotic Chemistry, Molecular Fossils, Nucleosides, and RNA* (Soll, D., Nishimura, S., and Moore, P. B., Eds.), Vol. 6 of Comprehensive Natural Products Chemistry (Barton, D., Nakanishi, K., and Meth-Cohn, O., Eds.) Elsevier: Oxford.
44. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. I, p 597, CRC Press, Cleveland, OH.
45. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., Vol. I, p 579, CRC Press, Cleveland, OH.
46. Petersheim, M., and Turner, D. H. (1983) *Biochemistry 22*, 256−263.
47. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843−853.
48. Geerdes, H. A. M., and Hilbers, C. W. (1979) *FEBS Lett. 107*, 125−128.
49. Van Knippenberg, P. H., Formenoy, L. J., and Heus, H. A. (1990) *Biochim. Biophys. Acta 1050*, 14−17.
50. Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) *Biochemistry 36*, 4449−4460.
51. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.
52. Varani, G., Aboul-ela, F., and Allain, F. H.-T. (1996) *Prog. NMR Spectrosc. 29*, 51−127.
53. Tinoco, I., Jr., Uhlenbeck, O. C., and Levine, M. D. (1971) *Nature 230*, 362−367.
54. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
55. Rabinovich, D., Haran, T., Eisenstein, M., and Shakked, Z. (1988) *J. Mol. Biol. 200*, 151−161.

56. Kneale, G., Brown, T., and Kennard, O. (1985) *J. Mol. Biol. 186*, 805−814.
57. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
58. McDowell, J. A., He, L., Chen, X., and Turner, D. H. (1997) *Biochemistry 36*, 8030−8038.
59. Biswas, R., Wahl, M. C., Ban, C., and Sundaralingam, M. (1997) *J. Mol. Biol. 267*, 1149−1156.
60. White, S. A., Nilges, M., Huang, A., Brunger, A. T., and Moore, P. B. (1992) *Biochemistry 31*, 1610−1621.
61. Kieft, J. S., and Tinoco, I., Jr. (1997) *Structure 5*, 713−721.
62. Mizuno, H., and Sundaralingam, M. (1978) *Nucleic Acids Res. 5*, 4451−4461.
63. Yang, J., and Gellman, S. H. (1998) *J. Am. Chem. Soc. 120*, 9090−9091.
64. Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. T. (1986) *J. Comput. Chem. 7*, 230−252.
65. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.
66. Leontis, N. B., and Westhof, E. (1998) *Q. Rev. Biophys. 31*, 399−455.
67. Turner, D. H., Sugimoto, N., Kierzek, R., and Deiker, S. D. (1987) *J. Am. Chem. Soc. 109*, 3783−3785.
68. Allain, F. H.-T., and Varani, G. (1995) *Nucleic Acids Res. 23*, 341−350.
69. Pardi, A., and Tinoco, I., Jr. (1981) *Biochemistry 21*, 4686−4693.
70. Gueron, M., Kochoyan, M., and Leroy, J. L. (1987) *Nature 328*, 89−92.
71. Pardi, A., Morden, K. M., Patel, D. J., and Tinoco, I., Jr. (1982) *Biochemistry 21*, 6567−6574.
72. Cheung, S., Arndt, K., and Lu, P. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3665−3669.
73. Betzel, Ch., Lorenz, S., Furste, J. P., Bald, R., Zhang, M., Schneider, Th. R., Wilson, K. S., and Erdmann, V. A. (1994) *FEBS Lett. 351*, 159−164.
74. Zhang, X., Gaffney, B. L., and Jones, R. A. (1998) *J. Am. Chem. Soc. 120*, 615−618.
75. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
76. Dickerson, R. E. (1989) *Nucleic Acids Res. 17*, 1797−1803.
77. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) *Biochemistry 25*, 5755−5759.

BI992938E