# Thermodynamics of RNA Internal Loops with a Guanosine-Guanosine Pair Adjacent to Another Noncanonical Pair†

Mark E. Burkard, Tianbing Xia,‡ and Douglas H. Turner*

*Department of Chemistry, University of Rochester, Rochester, New York 14627-0216*

*Received May 30, 2000; Revised Manuscript Received December 19, 2000*

ABSTRACT: Thermodynamic parameters measured by optical melting are reported for formation of RNA duplexes containing tandem noncanonical pairs with at least one guanosine-guanosine (GG) pair. For selected sequences, imino proton NMR provides evidence that the desired duplex forms and that the structure of a GG pair adjacent to a noncanonical pair depends on context. A GG pair next to a different noncanonical pair is more stable than expected from measurements of adjacent GG pairs. This is likely due to an unfavorable stacking interaction between adjacent GG pairs, where areas of high negative charge probably overlap. The results suggest a model where tandem noncanonical pairs closed by two GC pairs are assigned the following free energy increments at 37 °C: 0.8 kcal/mol for adjacent GG pairs, 1.0 kcal/mol for GG next to UU, and −0.3 kcal/mol for all others. These values are adjusted by 0.65 kcal/mol for each closing AU pair.

Studies of the thermodynamic stability of adjacent, or tandem, noncanonical pairs in RNA have provided insights into the nature of noncanonical pairs and have improved predictions of RNA secondary structure (*1*−*4*). Results for specific noncanonical pairs in the tandem context can often be generalized to other contexts (*4*), but GG pair stabilities are disparate between tandem and single contexts. For example, the tandem GG motif in the 2 × 2 internal loop $^{5'C\mathbf{GG}G3'}_{3'G\mathbf{GG}C5'}$ destabilizes a duplex by 0.8 kcal/mol, whereas the tandem GA motif in $^{5'C\mathbf{GA}G3'}_{3'G\mathbf{AG}C5'}$ stabilizes a duplex by −0.7 kcal/mol (*1*, *4*). When flanked by two adjacent GC pairs, however, a single GG can stabilize a duplex by ∼2 kcal/mol, whereas other single noncanonical pairs, including GA, are either destabilizing or essentially neutral (*5*−*7*). The stability of a single GG pair is rationalized by a three-hydrogen bond structure; in the context $^{5'A\mathbf{G}G3'}_{3'U\mathbf{G}C5'}$, the guanines swap roles in a syn−anti GG pair that is nearly isosteric to Watson−Crick pairs (*8*). The same dynamic pair is likely in other single GG contexts, although conformational exchange impairs identification of the pair (*7*, *8*).

Given the stability and structure of single GG pairs, it is surprising that adjacent GG pairs are destabilizing in 2 × 2 internal loops. In addition to the syn−anti GG pairing in the single context, the tandem GG context may provide room to form one of the other three GG pairings with two hydrogen bonds (*9*, *10*) or a pairing with a single bifurcated hydrogen bond linking O6 with imino and amino protons (*10*). This ability of GG to adopt multiple structures has been demonstrated with the 3 × 2 internal loop of the Rev-responsive element. In a crystal structure of free Rev RNA, the GG adopts a syn−anti pairing with NH1−O6 and NH2−N7 hydrogen bonds (*11*), as with a GG pair in a single mismatch (*8*). When bound to Rev, however, NMR studies indicate that the GG pair makes a symmetric pair with two NH1−O6 hydrogen bonds (*12*, *13*). This range of possibilities suggests that 2 × 2 internal loops may be stabilizing when a GG pair is adjacent to a different noncanonical pair.

Here, the thermodynamic stabilities of duplexes with a GG pair in a 2 × 2 internal loop are measured by monitoring UV[1] absorbance as a function of temperature. Imino proton NMR spectra are also obtained for several duplexes to provide basic structural information. The results show that a 2 × 2 internal loop with a GG pair adjacent to a noncanonical pair other than GG or UU can help stabilize an RNA duplex.

## MATERIALS AND METHODS

*RNA Synthesis*. RNA oligonucleotides were synthesized on an ABI 392 instrument using 2′-*tert*-butyldimethylsilyl-blocked phosphoramidites (*14*). After synthesis, the solid support was incubated in a 3:1 (v/v) ammonia/ethanol mixture at 55 °C overnight (*15*). The liquid phase was collected and evaporated. The residue was incubated in 1 M triethylammonium fluoride in anhydrous pyridine for 48 h at 55 °C. The oligomers were desalted on a Sep-pak C18 cartridge (Waters), purified on 500 $\mu$m silica gel TLC plates (Baker) developed with a 55:35:10 1-propanol/ammonia/water mixture, and desalted a second time. Oligomer purity was checked by reverse-phase HPLC, and all were ∼95% pure.

† This work was supported by NIH Grant GM 22939. M.E.B. is a trainee in the Medical Scientist Training Program funded by NIH Grant T32 GM07356.

* To whom correspondence should be addressed.Telephone: (716) 275-3207. Fax: (716) 473-6889. E-mail: turner@chem.rochester.edu.

‡ Present address: Department of Chemistry, Mail Code 147-75 BR, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125.

[1] Abbreviations: $C_T$, total concentration of all strands of oligomers in solution; $\Delta G°_T$, Gibbs free energy change at temperature $T$ under standard conditions; $\Delta H°$, enthalpy change under standard conditions; NOE, nuclear Overhauser effect; $R$, gas constant, 1.987 cal mol$^{-1}$ K$^{-1}$; $\Delta S°$, entropy change under standard conditions; $T_M$, melting temperature in kelvin; $T_m$, melting temperature in degrees Celsius; TSP, trimethylsilylsodium propionate; UV, ultraviolet.

10.1021/bi0012181 CCC: $20.00 
Published on Web 02/01/2001

Table 1: Occurrences of Noncanonical Pairs Adjacent to a GG Pair in 2 × 2 Internal Loops in Secondary Structures of 75 Group I Introns and 101 Small and 218 Large Subunit rRNAs

| 5′XG3′<br>3′YG5′<br>↓Y X→ | A | C | G | U |
|---|---|---|---|---|
| A | 6 | 8 | 1 | − |
| C | 2 | 6 | − | 1 |
| G | 2 | − | 0 | − |
| U | − | 3 | − | 5 |

*Melting Experiments and Analysis*. For melting experiments, oligomers were dissolved in 1 M NaCl, 0.5 mM $Na_2EDTA$, and 20 mM cacodylic acid, buffered to pH 7. Optical melting curves were obtained with a Gilford 250 spectrophotometer and a Gilford 2527 thermoprogrammer that ramped temperature at a rate of 1 °C/min. Absorbance at 280 nm was monitored because changes in absorbance due to the temperature-dependent hydration of cytosine are minimized at 280 nm (*16*, *17*). Non-self-complementary oligomers were mixed at a 1:1 ratio. Small errors in mixing do not affect results significantly (*18*).

The Meltwin software package (*19*) was used to fit curves of absorbance versus temperature to a two-state model incorporating sloping linear baselines. For each duplex, separate curves were obtained at nine strand concentrations, roughly from 10 $\mu$M to 1 mM. The inverse of the melting temperature, $T_M^{-1}$, for each curve was plotted against the logarithm of the total strand concentration, allowing determination of $\Delta H°$ and $\Delta S°$ by a van't Hoff analysis (*20*):

$$\frac{1}{T_M} = \frac{R}{\Delta H°} \ln(C_T/a) + \frac{\Delta S°}{\Delta H°} \quad (1)$$

where $a = 1$ for self-complementary duplexes and $a = 4$ for non-self-complementary duplexes.

*NMR Spectra*. Imino proton spectra for selected duplexes were obtained with a Varian Inova 500 MHz spectrometer. Samples were lyophilized and redissolved in a 9:1 $H_2O/D_2O$ mixture containing 80 mM NaCl, 0.5 mM $Na_2EDTA$, and 10 mM phosphate, buffered to pH 7. A 1:$\bar{3}$:3:$\bar{1}$ binomial pulse sequence was used to suppress the water signal, and offset delays were set to maximize peak intensity at ~12.5 ppm (*21*). For one-dimensional NOE experiments, resonances were saturated by a continuous wave for 2 s prior to acquisition, and spectra were subtracted from a reference spectrum obtained following off-resonance saturation. Spectra were referenced to TSP at 0 ppm. In samples without internal TSP, spectra were referenced to the residual water signal with a known chemical shift relative to TSP.

*Molecular Modeling*. For modeling a tandem GG structure, the Discover 95 software package was used with the Amber 95 force field (*22*). Desired hydrogen bond restraints were added, and up to 1000 steps of conjugate-gradient energy minimization were performed.

## RESULTS

*Natural Occurrence of GG Pairs in 2 × 2 Internal Loops*. A search of known RNA secondary structures of group I introns (*23*) and ribosomal RNAs (*24*, *25*) reveals 34 such 2 × 2 internal loops containing at least one GG pair. Table 1 shows the distribution of noncanonical pairs next to GG in 2 × 2 internal loops. The most common adjacent noncanonical pair is AC, with 10 occurrences. Adjacent AA, CC, and UU pairs are also relatively abundant, but the motif $^{5'\mathbf{GG}3'}_{3'\mathbf{GG}5'}$ is absent in this database. Because natural sequences contain 2 × 2 internal loops with GG pairs, thermodynamic parameters for such motifs may improve predictions of RNA secondary structure from sequence.

Compared to other 2 × 2 internal loops, those containing GG pairs are rare. For example, the database used for Table 1 contains 99 such 2 × 2 internal loops containing at least one AA pair, and 271 containing two GA pairs. The total of 34 GG pairs found in 2 × 2 internal loops is also considerably fewer than the 196 single GG pairs in this database. In part, the rarity of 2 × 2 internal loops with GG pairs may be attributed to the ability of G to form pairs with C, U, G, and A without large backbone distortions (*8*, *26*−*29*). The ability of G to form multiple helical pairs increases the likelihood of misfolding, which may provide an evolutionary disadvantage for the inclusion of G in unpaired regions of RNA (*30*). Such a disadvantage may result in the scarcity of tandem noncanonical pairs with GG in this database of RNAs with ancient origins.

*Thermodynamics*. The measured thermodynamics for formation of several duplexes containing 2 × 2 internal loops with GG are given in Table 2. Except for GG next to CC, all the possible adjacent noncanonical pairs are represented. Several attempts to prepare duplexes with a GG pair next to a CC pair were unsuccessful; imino proton NMR showed that the expected Watson−Crick stems were not formed by 5′GCAGCACC3′/3′CGUGCUGG5′, 5′CUGGCGAC3′/3′GACGCCUG5′, and 5′CAGCGGUC3′/3′GUCCGCAG5′. It is likely that these sequences form other duplexes, e.g., helixes that are slipped relative to the desired helix.

Table 3 provides the free energy contribution of the measured internal loops, $\Delta G°_{37,loop}$, calculated with

$$\Delta G°_{37,loop} = \Delta G°_{37,duplex} - \Delta G°_{37,ref} + \Delta G°_{37,NN} \quad (2)$$

where $\Delta G°_{37,duplex}$ is the free energy for duplex formation, as given in the first numerical column of Table 2, $\Delta G°_{37,ref}$ is the corresponding free energy for formation of the reference duplex with the same stem sequence but no internal loop, and $\Delta G°_{37,NN}$ is the free energy attributed to the nearest neighbor interrupted by the internal loop, as given by Xia et al. (*31*). The $\Delta H°_{loop}$ and $\Delta S°_{loop}$ values in Table 3 are calculated in a similar manner.

*Adjacent Noncanonical Pairs and Orientation*. When a 2 × 2 internal loop is flanked by GC pairs, the identity of the noncanonical pair next to GG affects loop stability. For a UU pair next to a GG pair, the average $\Delta G°_{37,loop}$ is 1.0 kcal/mol, which is similar to the value of 0.8 kcal/mol reported for adjacent GG pairs (*1*, *4*). For all other noncanonical pairs adjacent to a GG pair, the range in $\Delta G°_{37,loop}$ is from −0.86 to 0.22 kcal/mol with an average of −0.3 kcal/mol. Thus, a 2 × 2 internal loop with a GG pair is destabilizing if it has an adjacent GG or UU pair. Other adjacent noncanonical pairs allow the internal loop to be stabilizing.

Compatibility of adjacent mismatches is an important determinant of stability of 2 × 2 internal loops (*3*). For example, UU and AG pairs can both be stabilizing pairs in a 2 × 2 symmetric internal loop, but when adjacent to each

Table 2: Thermodynamics of Duplexes Containing 2 × 2 Internal Loops with a GG Pair

| | $T_M$ dependence on concentration[a] | | | | curve fit[b] | | | |
|---|---|---|---|---|---|---|---|---|
| duplex | $\Delta G°_{37}$ (kcal/mol) | $\Delta H°$ (kcal/mol) | $\Delta S°$ (eu) | $T_m$ (°C) at 0.1 mM | $\Delta G°_{37}$ (kcal/mol) | $\Delta H°$ (kcal/mol) | $\Delta S°$ (eu) | $T_m$ (°C) at 0.1 mM |
| 5′GAGGCGAG3′<br>3′CUCGUCUC5′ | −5.67 ± 0.03 | −41.7 ± 0.9 | −116.2 ± 2.9 | 30.7 | −5.60 ± 0.08 | −43.5 ± 2.8 | −122.1 ± 9.2 | 30.5 |
| 5′GAGCGGAG3′<br>3′CUCUGCUC5′ | −5.72 ± 0.05 | −44.4 ± 1.6 | −124.8 ± 5.2 | 31.5 | −5.68 ± 0.09 | −46.2 ± 1.9 | −130.6 ± 6.1 | 31.4 |
| 5′GAGCGGAG3′<br>3′CUCAGCUC5′ | −5.67 ± 0.04 | −46.7 ± 1.4 | −132.2 ± 4.6 | 31.4 | −5.87 ± 0.35 | −43.7 ± 6.2 | −122.0 ± 21.1 | 32.4 |
| 5′GUGGCGUG3′<br>3′CACGUCAC5′ | −5.27 ± 0.05 | −53.5 ± 1.4 | −155.4 ± 4.6 | 29.9 | −5.45 ± 0.07 | −45.8 ± 4.9 | −130.1 ± 15.8 | 29.9 |
| 5′CUGGAGUC3′<br>3′GACGGCAG5′ | −5.09 ± 0.04 | −55.4 ± 1.9 | −162.5 ± 6.3 | 28.4 | −5.09 ± 0.13 | −48.8 ± 1.4 | −141.0 ± 4.6 | 28.1 |
| 5′GAGGCGAG3′<br>3′CUCGACUC5′ | −5.55 ± 0.05 | −46.2 ± 1.7 | −131.0 ± 5.5 | 30.5 | −5.73 ± 0.22 | −41.3 ± 3.0 | −114.6 ± 10.4 | 31.1 |
| 5′CUGAGGUC3′<br>3′GACGGCAG5′ | −5.15 ± 0.04 | −52.5 ± 2.1 | −152.8 ± 6.9 | 29.0 | −5.25 ± 0.14 | −47.9 ± 2.3 | −137.5 ± 7.7 | 28.9 |
| 5′GUGGAGUG3′<br>3′CACGACAC5′ | −4.66 ± 0.06 | −33.6 ± 1.3 | −93.4 ± 4.5 | 20.7 | −4.68 ± 0.10 | −34.5 ± 4.7 | −96.2 ± 15.6 | 21.3 |
| 5′GUGAGGUG3′<br>3′CACAGCAC5′ | −4.51 ± 0.13 | −37.4 ± 2.6 | −106.1 ± 8.6 | 21.1 | −4.63 ± 0.24 | −36.6 ± 3.9 | −103.1 ± 13.3 | 21.7 |
| 5′GAGGUGAG3′<br>3′CUCGCCUC5′ | −5.03 ± 0.13 | −40.5 ± 2.3 | −114.4 ± 7.9 | 25.9 | −5.16 ± 0.18 | −38.5 ± 2.6 | −107.5 ± 9.0 | 26.3 |
| 5′GAGUGGAG3′<br>3′CUCCGCUC5′ | −5.02 ± 0.02 | −44.9 ± 0.7 | −128.7 ± 2.5 | 26.9 | −4.98 ± 0.11 | −45.7 ± 3.9 | −131.1 ± 12.9 | 26.8 |
| 5′GAGGUGAG3′<br>3′CUCGUCUC5′ | −4.56 ± 0.16 | −44.6 ± 2.8 | −129.1 ± 9.4 | 23.9 | −4.64 ± 0.22 | −44.6 ± 4.4 | −128.8 ± 14.8 | 24.4 |
| (5′UGCGGGCA3′)$_2$[c] | −5.04 ± 0.32 | −63.7 ± 3.6 | −189.2 ± 11.5 | 33.9 | −5.37 ± 0.12 | −52.3 ± 2.2 | −151.4 ± 7.3 | 35.0 |
| 5′GAGUGGAG3′<br>3′CUCUGCUC5′ | −3.87 ± 0.19 | −57.8 ± 3.5 | −173.9 ± 11.8 | 23.4 | −4.05 ± 0.20 | −53.8 ± 5.9 | −160.5 ± 19.5 | 23.3 |
| (5′GCAGGUGC3′)$_2$ | −4.37 ± 0.04 | −53.9 ± 1.8 | −159.5 ± 5.8 | 29.6 | −4.38 ± 0.17 | −57.9 ± 8.7 | −172.5 ± 28.3 | 30.2 |
| (5′GCUGGAGC3′)$_2$ | −4.27 ± 0.10 | −48.7 ± 2.5 | −143.3 ± 8.4 | 28.3 | −4.18 ± 0.16 | −53.2 ± 6.8 | −158.0 ± 22.3 | 28.5 |

| | $T_M$ dependence on concentration[a] | | | | curve fit[b] | | | |
|---|---|---|---|---|---|---|---|---|
| reference duplex | $\Delta G°_{37}$ (kcal/mol) | $\Delta H°$ (kcal/mol) | $\Delta S°$ (eu) | $T_m$ (°C) at 0.1 mM | $\Delta G°_{37}$ (kcal/mol) | $\Delta H°$ (kcal/mol) | $\Delta S°$ (eu) | $T_m$ (°C) at 0.1 mM |
| 5′GAGGAG3′[d]<br>3′CUCCUC5′ | −8.50 ± 0.05 | −55.7 ± 1.7 | −152.2 ± 5.2 | 48.4 | −8.66 ± 0.20 | −58.8 ± 4.9 | −161.8 ± 15.2 | 48.6 |
| 5′GUGGUG3′[e]<br>3′CACCAC5′ | −7.67 ± 0.10 | −48.8 ± 3.3 | −132.7 ± 10.5 | 47.4 | −7.78 ± 0.20 | −55.7 ± 3.2 | −154.6 ± 10.2 | 46.6 |
| 5′CUGGUC3′<br>3′GACCAG5′ | −8.02 ± 0.04 | −55.1 ± 1.5 | −151.7 ± 4.8 | 45.6 | −8.13 ± 0.14 | −57.7 ± 3.2 | −159.7 ± 9.8 | 45.9 |
| (5′GCAUGC3′)$_2$[f] | −7.38 | −62.3 | −177.2 | 45.7 | −7.41 | −59.6 | −168.4 | 46.3 |
| (5′GCUAGC3′)$_2$[g] | −7.92 | −59.1 | −165.1 | 49.3 | −8.07 | −63.0 | −177.1 | 49.3 |

[a] Parameters determined from the dependence of $T_M$ on strand concentration. [b] Average parameters, as determined by fitting individual melting curves. [c] From ref *1*. [d] From ref *3*. [e] From ref *31*. [f] From ref *41*. [g] From ref *42*.

other, they are destabilizing. According to the hypothesis proposed by Xia et al. (*3*), the pair formed by two pyrimidines is incompatible with a larger purine-purine pair because these tend to distort the backbones in different directions, and both cannot be accommodated as neighbors. The work presented here suggests that the small, stable UU pair is also incompatible with the GG pair, but that the GA pair is compatible with the GG pair. The compatible fit hypothesis may also explain why a CU pair adjacent to a GG pair is more stable than a UU or UC pair next to a GG pair (Table 3). A CU pair can form a water-inserted pair that is sterically compatible with purine-purine pairs, and which can covary with GA (*10*). This could rationalize the results in Table 3 if the formation and/or energetics of this structure are dependent on the flanking pairs so that it is less stable for the sequences with UC pairs. Interestingly, the destabilizing motifs in Table 3 contain either two GG mismatches or the sequence 5′GUG3′. The sequence 5′GUG3′ may also have unusual stacking in the single strand (*32*).

*Imino Proton Spectra*. Imino proton spectra for five duplexes are shown in Figure 1. All the duplexes were designed to have two AU pairs, and all spectra but one exhibit two resonances above 14 ppm where the imino proton of U in an AU pair is expected to resonate. The exception is (5′GCAGGUGC3′)$_2$ where the two AU pairs are equivalent and next to GG mismatches. For (5′GCAGGUGC3′)$_2$, there is a resonance at 13.4 ppm, which likely corresponds to the AU pairs in this unusual environment. A similar upfield shift is seen with an AU pair next to a single GG pair (*8*). The spectra also contain the expected number of resonances for GC pairs, although $^{5'GUGGAGUG3'}_{3'CACGACAC5'}$ has additional small resonances (panel E). One-dimensional NOE difference spectra for $^{5'CUGAGGUC3'}_{3'GACGGCAG5'}$ support the assignment of resonances above 12 ppm to protons in Watson−Crick pairs, so resonances below 12 ppm arise from protons in the 2 × 2 internal loop (Supporting Information).

On the basis of the NMR spectra, structures of GG pairs may depend on loop sequence. The GG imino resonances

Table 3: Thermodynamics of Loop Formation for 2 × 2 Internal Loops Containing GG[a]

| internal loop | $\Delta G^\circ_{37,loop}$ | predicted[b] | $\Delta H^\circ_{loop}$ | $\Delta S^\circ_{loop}$ |
|---|---|---|---|---|
| 5′GGCG3′<br>3′CGUC5′ | −0.86 ± 0.13 | −0.3 (1.1) | −18.1 ± 3.8 | −55.4 ± 12.1 |
| 5′GCGG3′<br>3′CUGC5′ | −0.48 ± 0.10 | −0.3 (1.0) | −2.1 ± 2.6 | −5.3 ± 8.3 |
| 5′GCGG3′<br>3′CAGC5′ | −0.43 ± 0.09 | −0.3 (0.9) | −4.4 ± 2.5 | −12.7 ± 7.9 |
| 5′GGCG3′<br>3′CGUC5′ | −0.43 ± 0.09 | −0.3 (1.1) | 0.6 ± 2.3 | 3.3 ± 7.1 |
| 5′GGAG3′<br>3′CGGC5′ | −0.33 ± 0.09 | −0.3 (0.1) | −13.7 ± 2.7 | −43.5 ± 8.8 |
| 5′GGCG3′<br>3′CGAC5′ | −0.31 ± 0.10 | −0.3 (1.4) | −3.9 ± 2.7 | −11.5 ± 8.5 |
| 5′GAGG3′<br>3′CGGC5′ | −0.29 ± 0.09 | −0.3 (−0.3) | −10.8 ± 4.6 | −33.8 ± 9.2 |
| 5′GGAG3′<br>3′CGAC5′ | −0.25 ± 0.14 | −0.3 (1.1) | 1.8 ± 3.8 | 6.6 ± 12.0 |
| 5′GAGG3′<br>3′CAGC5′ | −0.10 ± 0.18 | −0.3 (1.2) | −2.0 ± 4.4 | −6.1 ± 14.1 |
| 5′GGUG3′<br>3′CGCC5′ | 0.21 ± 0.16 | −0.3 (1.1) | 1.8 ± 3.1 | 5.1 ± 10.2 |
| 5′GUGG3′<br>3′CCGC5′ | 0.22 ± 0.09 | −0.3 (0.9) | −2.6 ± 2.2 | −9.2 ± 6.9 |
| 5′GGUG3′<br>3′CGUC5′ | 0.68 ± 0.18 | 1.0 (1.2) | −2.3 ± 3.5 | −9.6 ± 11.4 |
| (5′CGGG3′)$_2$[c] | 0.8 | 0.8 (0.8) | | |
| 5′GUGG3′<br>3′CUGC5′ | 1.37 ± 0.21 | 1.0 (1.2) | −15.5 ± 4.1 | −54.4 ± 13.4 |
| (5′AGGU3′)$_2$ | 1.91 | 2.1 (1.4) | −1.0 | −9.0 |
| (5′UGGA3′)$_2$ | 2.32 | 2.1 (1.5) | 2.7 | 1.3 |

[a] Errors are calculated as the square root of the sum of squared errors for the duplex containing the internal loop, for the reference duplex, and for the nearest neighbor interrupted by the loop as given in ref *31*. [b] Predicted according to the approximations described in the Discussion or, in parentheses, the method of ref *4*. [c] From ref *1*, using the −2.36 kcal/mol nearest neighbor parameter for $^{5'\mathbf{CG}3'}_{3'\mathbf{GC}5'}$ (*31*).

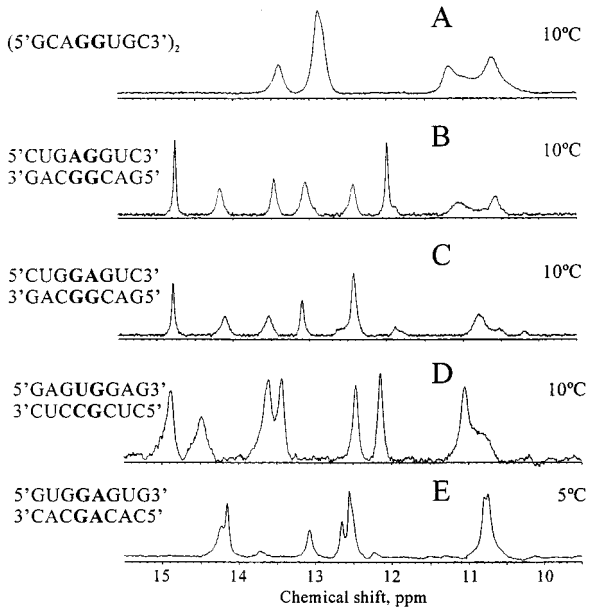


FIGURE 1: Imino proton spectra for selected duplexes with 2 × 2 internal loops containing a GG pair.

and possibly a resonance from the adjacent noncanonical pair appear between 10 and 11.5 ppm in Figure 1. For some sequences, the resonances from the GG pair are nearly equal in size, but others have unequal resonances, suggesting different structures. In particular, the NMR spectra also suggest that a GG pair in a 2 × 2 internal loop sometimes has a structure different from that of a single GG pair. For example, the noncanonical guanines in $^{5'\mathrm{GUG}\mathbf{GA}\mathrm{GUG}3'}_{3'\mathrm{CAC}\mathbf{GA}\mathrm{CAC}5'}$ have equally sharp resonances, both near 10.8 ppm. In contrast, the G imino protons in a single GG pair in the $^{5'\mathrm{A}\mathbf{G}\mathrm{G}3'}_{3'\mathrm{U}\mathbf{G}\mathrm{C}5'}$ context have unequal resonances, indicating exchange at different rates with solvent (*8*).

From the imino proton spectra for $^{5'\mathrm{CUG}\mathbf{AG}\mathrm{GUC}3'}_{3'\mathrm{GAC}\mathbf{GG}\mathrm{CAG}5'}$ and $^{5'\mathrm{CUG}\mathbf{GA}\mathrm{GUC}3'}_{3'\mathrm{GAC}\mathbf{GG}\mathrm{CAG}5'}$ shown in panels B and C of Figure 1, it is difficult to determine if the GA pair is in the imino hydrogen-bonded conformation, the sheared conformation, or a different conformation. An imino GA conformation typically has a strong G imino resonance near 12.5 ppm, with a strong NOE to the AH2 near 7.8 ppm (*33*). From the one-dimensional NOE spectrum for $^{5'\mathrm{CUG}\mathbf{AG}\mathrm{GUC}3'}_{3'\mathrm{GAC}\mathbf{GG}\mathrm{CAG}5'}$, the resonances between 12 and 13 ppm are more likely G imino protons in GC pairs, since NOEs to the 7.8−8.5 ppm region are to broad resonances, suggestive of C amino protons (Supporting Information). Thus, the GA pairs do not exhibit expected NMR properties of an imino hydrogen-bonded conformation. Sheared GA conformations are possible as these usually exhibit a resonance near 10 ppm (*34*−*36*). This would be surprising, however, for $^{5'\mathrm{CUG}\mathbf{AG}\mathrm{GUC}3'}_{3'\mathrm{GAC}\mathbf{GG}\mathrm{CAG}5'}$ since sheared GA conformations are thought to be prohibited when a Watson−Crick pair is 5′ to the A (*37*).

## DISCUSSION

Previously, it has been shown that a single GG pair is as much as 3 kcal/mol more stable than other noncanonical pairs, including the GA pair (*5*−*7*). Here, we find that in 2 × 2 internal loops, the GG pair can stabilize the loop by more than 1 kcal/mol relative to other noncanonical pairs. For examples, $^{5'\mathrm{G}\mathbf{GA}\mathrm{G}3'}_{3'\mathrm{C}\mathbf{GA}\mathrm{C}5'}$ has a $\Delta G^\circ_{37,loop}$ of −0.3 kcal/mol, compared to a value of 1.4 kcal/mol for $^{5'\mathrm{G}\mathbf{AA}\mathrm{G}3'}_{3'\mathrm{C}\mathbf{AA}\mathrm{C}5'}$, and

$^{5'\mathbf{GG}C3'}_{3'\mathbf{GA}C5'}$ has a $\Delta G°_{37,loop}$ of −0.3 kcal/mol, compared with a value of 1.1 kcal/mol for $^{5'\mathbf{GA}C3'}_{3'\mathbf{CA}C5'}$ (Table 3 and refs *3* and *4*). These results suggest that the GG pair can make favorable pairings that are compatible with multiple adjacent noncanonical pairs. The favorable pairing of a GG pair can have a large effect on RNA folding since the equilibrium constant for folding becomes 10-fold more favorable for every 1.4 kcal/mol in free energy at 37 °C.

*Two Adjacent GG Pairs Are Less Stable Than GG Pairs Next to Other Noncanonical Pairs Except UU Pairs*. The compatibilities of noncanonical pair shape and stacking interactions with the closing base pairs are important determinants of stability and structure for 2 × 2 internal loops (*3*, *19*, *33*, *35*). The current work shows that stacking interactions between the noncanonical pairs in the 2 × 2 internal loop are also crucial, since internal loops with two adjacent GG pairs are less favorable than internal loops of GG pairs next to most other noncanonical pairs, and also less favorable than single GG pairs. For example, at 37 °C the loop free energy increment for $^{5'C\mathbf{GG}G3'}_{3'G\mathbf{GG}C5'}$ is 0.8 kcal/mol (*1*, *4*), but for $^{5'C\mathbf{GA}G3'}_{3'G\mathbf{GG}C5'}$ and $^{5'C\mathbf{AG}G3'}_{3'G\mathbf{GG}C5'}$, it is −0.3 kcal/mol (Table 3). Single GG pairs in the $^{5'C\mathbf{G}G3'}_{3'G\mathbf{G}C5'}$ context are stabilizing by 1.4 kcal/mol (*7*), so the destabilization of tandem GG pairs is not attributable to stacking with neighboring Watson−Crick pairs. The additional requirement for adjacent GG pairs is stacking of the two GG pairs.

The poor stacking stability of adjacent GG pairs is most likely attributable to unfavorable electrostatic interactions. (a) Among the four natural bases in RNA, G has the largest dipolar charge separation (*38*). (b) The van der Waals interactions are expected to be approximately similar for G and A, but a GA pair adjacent to a GG pair is more favorable than two adjacent GG pairs. Thus, the electrostatic portion of stacking interactions is likely the most sensitive to base composition and geometry.

To identify possible unfavorable electrostatic interactions between adjacent GG pairs, the structure was modeled by assuming the N1-carbonyl, N7-amino pairing observed with a single GG pair (*8*). Of the four possible GG geometries with two hydrogen bonds, this pairing distorts the backbone the least. Figure 2 shows two models of adjacent GG pairs, obtained by energy minimization with the Amber 95 force field (*22*). The GG pairs were restrained to have H1−O6 and NH2−N7 hydrogen bonds, as observed with a single GG pair in a duplex (*8*). Regions with strong negative electric fields, shown as shaded regions, are more localized than positively charged regions, and overlaps indicate unfavorable electrostatic interactions. When the 3′ G is syn (Figure 2A), adjacent guanines are stacked within a strand. In this conformation, there are small unfavorable electrostatic overlaps and the negatively charged regions are localized to the major groove and pairing faces. When the 5′ G is syn (Figure 2B), the 3′ guanines are stacked cross-strand, rather than within a strand, but there are highly unfavorable overlaps of negatively charged regions. In both panels, the four negatively charged O6 groups are closely juxtaposed in the major grove, so mutual repulsions of these groups may destabilize adjacent GG pairs. Because of the especially large charge separations in guanine, other GG pair conformations are also likely to produce unfavorable overlaps, and these in addition will require significant backbone distortion (*8*).

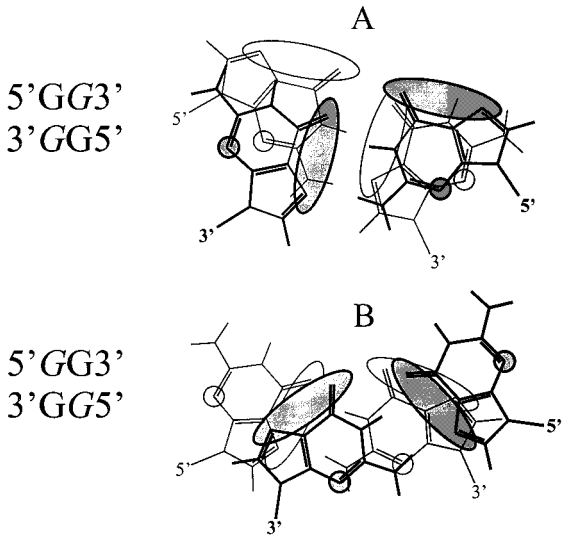


FIGURE 2: Models of adjacent GG pairs, using the syn−anti pairing observed for a single GG pair (*8*). The shaded areas represent regions with strong negative electric potential. The darker bonds and shading indicate the bases closest to the reader. The guanosine in the syn glycosidic conformation is designated with italics. (A) The 3′ guanosines are in the syn glycosidic conformation. (B) The 5′ guanosines are syn.

Thus, the structural model of adjacent GG pairs is consistent with unfavorable electrostatic interactions of the overlapped GG pairs. Similarly, the negatively charged carbonyls on uracil may make stacking unfavorable between GG and UU pairs.

*Structures of GG Pairs in 2 × 2 Internal Loops*. The imino proton spectra of Figure 1 show resonances between 10.5 and 11.5 ppm that correspond to the internal loop bases. Thus, imino protons of the internal loops are somewhat protected from exchange with bulk solvent. The favorable $\Delta G°_{37,loop}$ for many of these duplexes suggests favorable hydrogen bonds (*9*, *10*), although other interactions may protect these protons from exchange with water. No conclusions can be made about the exact pairing geometries, however, since chemical shifts between 10 and 11.5 ppm have been observed for GG pairs in several different geometries (*8*, *12*, *13*, *39*).

For internal loops with a CU pair next to a GG pair, a GC or GU pair may form rather than GG and CU pairs. If so, then there are bulged bases in $^{5'GAG\mathbf{UG}GAG3'}_{3'CUC\mathbf{CG}CUC5'}$ since the spectrum in Figure 1D indicates that the Watson−Crick stems are intact.

*Approximations for Stability Increments of Internal Loops with GG Pairs*. Previously, stabilities of 2 × 2 internal loops with a single GG pair were approximated from stabilities of tandem GG pairs (*3*, *4*). Values for $\Delta G°_{37,loop}$ from this model are listed in parentheses in Table 3, and differ significantly from measured values. On the basis of the measured values, when the 2 × 2 internal loop is flanked by two GC pairs, it is reasonable to use a free energy increment of 1.0 kcal/mol at 37 °C for $^{5'\mathbf{GU}3'}_{3'\mathbf{GU}5'}$ and $^{5'\mathbf{UG}3'}_{3'\mathbf{UG}5'}$, and an increment of 0.8 kcal/mol for $^{5'\mathbf{GG}3'}_{3'\mathbf{GG}5'}$. For all other 2 × 2 internal loops flanked by two GC pairs, the range of free energy increments is from −0.86 to 0.22 kcal/mol, with no clear pattern. Thus, these loops are reasonably approximated by an average $\Delta\Delta G°_{37}$ of −0.3 kcal/mol. On the basis of results with other internal loops, these values are adjusted by 0.45 (hydrogen bond loss)

+ 0.20 (stacking correction) = 0.65 kcal/mol when a closing GC pair is replaced with a closing AU pair (*4*, *40*, *41*). Predicted loop free energy increments from these approximations are listed in Table 3.

As GG pairs can form stabilizing 1 × 1 and 2 × 2 internal loops, they may also stabilize larger internal loops. A first approximation for the thermodynamic bonus provided by a GG pair at the end of an internal loop can be provided by comparing the stabilities of loops in Table 3 with previously measured stabilities of 2 × 2 internal loops without the potentially stabilizing GU, GA, UU, or GG pairs. The average free energy increment for 16 such 2 × 2 loops closed by two GC pairs is 1.3 kcal/mol (*1*−*4*). This is 1.6 kcal/mol less favorable than the average free energy increment of −0.3 kcal/mol calculated above for 11 such 2 × 2 internal loops with a single GG pair and no adjacent UU pair. This suggests that the bonus for GG pairs terminating internal loops may be as favorable as −1.6 kcal/mol when the loops have equal numbers of nucleotides on each side. This is more favorable than the bonus of −1.1 kcal/mol currently used for GA pairs (*4*). Studies of 2 × 3 internal loops suggest these bonuses are less favorable when the numbers of nucleotides on each side of the loop differ (*40*).

## ACKNOWLEDGMENT

We thank an anonymous reviewer for bringing to our attention water-inserted CU pairs.

## SUPPORTING INFORMATION AVAILABLE

A set of one-dimensional NOE difference spectra for 5′CUG**AG**GUC3′ 3′GAC**GG**CAG5′. This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES

1. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) *Biochemistry 30*, 8242−8251.
2. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
3. Xia, T., McDowell, J. A., and Turner, D. H. (1997) *Biochemistry 36*, 12486−12497.
4. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
5. Zhu, J., and Wartell, R. M. (1997) *Biochemistry 36*, 15326−15335.
6. Bevilacqua, J. M., and Bevilacqua, P. C. (1998) *Biochemistry 37*, 15877−15884.
7. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) *Biochemistry 38*, 14214−14223.
8. Burkard, M. E., and Turner, D. H. (2000) *Biochemistry 39*, 11748−11762.
9. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., Appendix 1, pp 675−680, Cold Spring Harbor Laboratory Press, Plainview, NY.
10. Leontis, N. B., and Westhof, E. (1998) *Q. Rev. Biophys. 31*, 399−455.
11. Ippolito, J. A., and Steitz, T. A. (2000) *J. Mol. Biol. 295*, 711−717.
12. Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D., and Williamson, J. R. (1996) *Science 273*, 1547−1551.
13. Peterson, R. D., and Feigon, J. (1996) *J. Mol. Biol. 264*, 863−877.
14. 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.
15. Stawinski, J., Strömberg, R., Thelin, M., and Westman, E. (1988) *Nucleosides Nucleotides 7*, 779−782.
16. Pörschke, D. (1976) *Biochemistry 15*, 1495−1499.
17. Johnson, W. C., Jr., Vipond, P. M., and Girod, J. C. (1971) *Biopolymers 10*, 923−933.
18. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) *Biochemistry 30*, 6428−6436.
19. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
20. Borer, P. N., Dengler, B., Tinoco, I., Jr., and Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843−853.
21. Hore, P. J. (1983) *J. Magn. Reson. 55*, 283−300.
22. 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.
23. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
24. Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3502−3507.
25. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
26. McDowell, J. A., He, L., Chen, X., and Turner, D. H. (1997) *Biochemistry 36*, 8030−8038.
27. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.
28. Trikha, J., Filman, D. J., and Hogle, J. M. (1999) *Nucleic Acids Res. 27*, 1728−1739.
29. Correll, C. C., Freeborn, B., Moore, P. B., and Steitz, T. A. (1997) *Cell 91*, 705−712.
30. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 233−264, Cold Spring Harbor Laboratory Press, Plainview, NY.
31. 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.
32. Gray, D. M., Tinoco, I., Jr., and Chamberlin, M. J. (1972) *Biopolymers 11*, 1235−1258.
33. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.
34. Li, Y., Zon, G., and Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26−30.
35. SantaLucia, J., Jr., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
36. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−980.
37. Gautheret, D., Konings, D., and Gutell, R. R. (1994) *J. Mol. Biol. 242*, 1−8.
38. Bakalarski, G., Grochowski, P., Kwiatkowski, J. S., Lesyng, B., and Leszczynski, J. (1996) *Chem. Phys. 204*, 301−311.
39. Dallas, A., and Moore, P. B. (1997) *Structure 5*, 1639−1653.
40. Schroeder, S. J., and Turner, D. H. (2000) *Biochemistry 39*, 9257−9274.
41. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.
42. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) *Biochemistry 25*, 5755−5759.

BI0012181