J|A|C|S ARTICLES

Published on Web 05/28/2009

# Contribution of the Closing Base Pair to Exceptional Stability in RNA Tetraloops: Roles for Molecular Mimicry and Electrostatic Factors

Joshua M. Blose,[†] David J. Proctor,[‡] Narayanan Veeraraghavan,[⊥] Vinod K. Misra,*[,§] and Philip C. Bevilacqua*[,†]

*Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, Cancer Research UK Nucleic Acid Structure Research Group, University of Dundee, Dundee DD1 5EH, United Kingdom, and Department of Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, Michigan 48109*

Received January 5, 2009; E-mail: pcb@chem.psu.edu (P.C.B.); vmisra@umich.edu (V.K.M.)

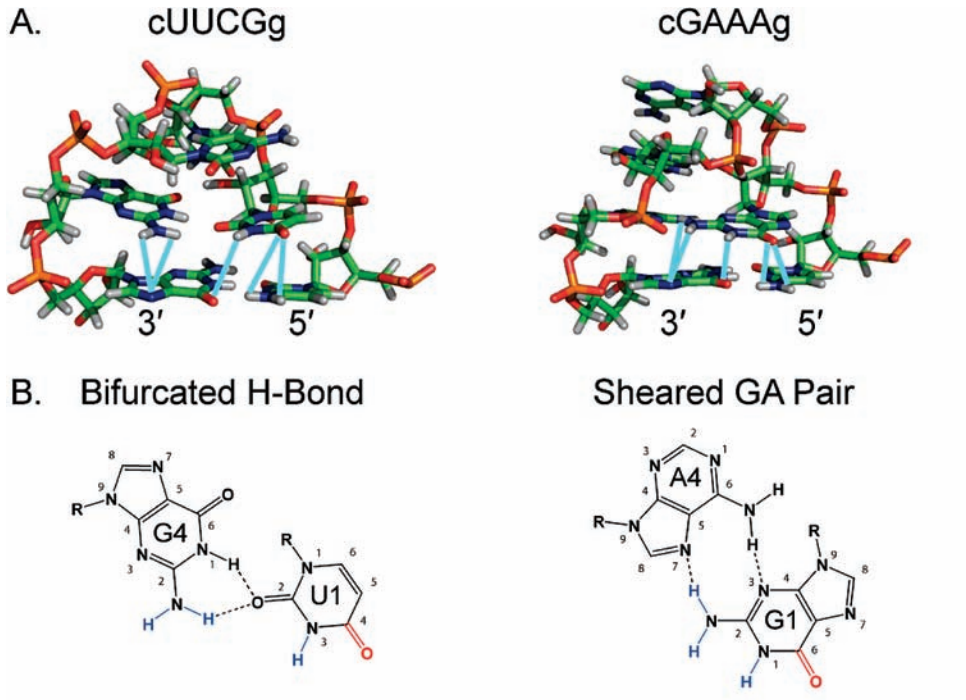
***Abstract:*** Hairpins are common RNA secondary structures that play multiple roles in nature. Tetraloops are the most frequent RNA hairpin loops and are often phylogenetically conserved. For both the UNCG and GNRA families, CG closing base pairs (cbps) confer exceptional thermodynamic stability but the molecular basis for this has remained unclear. We propose that, despite having very different overall folds, these two tetraloop families achieve stability by presenting the same functionalities to the major groove edge of the CG cbp. Thermodynamic contributions of this molecular mimicry were investigated using substitutions at the nucleobase and functional group levels. By either interrupting or deleting loop−cbp electrostatic interactions, which were identified by solving the nonlinear Poisson−Boltzmann (NLPB) equation, stability changed in a manner consistent with molecular mimicry. We also observed a linear relationship between $\Delta G^\circ_{37}$ and log[$Na^+$] for both families, and loops with a CG cbp had a decreased dependence of stability on salt. NLPB calculations revealed that, for both UUCG and GAAA tetraloops, the GC cbp form has a higher surface charge density, although it arises from changes in loop compaction for UUCG and changes in loop configuration for GAAA. Higher surface charge density leads to stronger interactions of GC cbp loops with solvent and salt, which explains the correlation between experimental and calculated trends of free energy with salt. Molecular mimicry as evidenced in these two stable but otherwise unrelated tetraloops may underlie common functional roles in other RNA and DNA motifs.

The hairpin is the most common RNA secondary structure motif and has diverse structures and biological and physical functions.[1] Hairpins often participate in RNA tertiary contacts[2−4] and play roles in cellular processes such as transcription regulation,[5,6] mRNA degradation,[7−9] and RNA interference.[10−12] Tetraloops are the most common RNA hairpins, and phylogenetic analyses show that ∼50% of the hairpin structures in rRNA are tetraloops.[13−17] Many tetraloops, including some rRNA loops, belong to the UNCG or GNRA motifs, where N is any nucleotide and R is a purine (Figure 1). We found that the UNCG motif is a subset of an expanded family of tetraloops known as YNMG, where M is A or C and Y is a pyrimidine.[17] Tetraloops in these families are thermodynamically stable, with GNRA and UNMG loops being particularly stable in the context of a CG closing base pair (cbp) as compared with other Watson−Crick cbps.[17−24] For example, in previous studies the CG cbp was found to confer an additional −1.3 kcal/mol for

[†] Department of Chemistry, The Pennsylvania State University.
[⊥] Huck Institutes of the Life Sciences, The Pennsylvania State University.
[‡] University of Dundee.
[§] University of Michigan.

(1) Bevilacqua, P. C.; Blose, J. M. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79–103.
(2) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678–1685.
(3) Chang, K. Y.; Tinoco, I., Jr. *J. Mol. Biol.* **1997**, *269*, 52–66.
(4) Martick, M.; Scott, W. G. *Cell* **2006**, *126*, 309–320.
(5) Babitzke, P.; Schaak, J.; Yakhnin, A. V.; Bevilacqua, P. C. *Methods Enzymol.* **2003**, *371*, 392–404.
(6) McGraw, A. P.; Bevilacqua, P. C.; Babitzke, P. *RNA* **2007**, *13*, 2020–2033.
(7) Adams, C. C.; Stern, D. B. *Nucleic Acids Res.* **1990**, *18*, 6003–6010.
(8) Klug, G.; Cohen, S. N. *J. Bacteriol.* **1990**, *172*, 5140–5146.
(9) Khan, I. M.; Coulson, J. M. *Nucleic Acids Res.* **1993**, *21*, 2957–2958.
(10) McManus, M. T.; Petersen, C. P.; Haines, B. B.; Chen, J.; Sharp, P. A. *RNA* **2002**, *8*, 842–850.
(11) Lee, Y.; Jeon, K.; Lee, J. T.; Kim, S.; Kim, V. N. *EMBO J.* **2002**, *21*, 4663–4670.
(12) Boden, D.; Pusch, O.; Silbermann, R.; Lee, F.; Tucker, L.; Ramratnam, B. *Nucleic Acids Res.* **2004**, *32*, 1154–1158.
(13) Woese, C. R.; Winker, S.; Gutell, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467–8471.
(14) Wolters, J. *Nucleic Acids Res.* **1992**, *20*, 1843–1850.
(15) Gutell, R. R. *Nucleic Acids Res.* **1993**, *21*, 3051–3054.
(16) Gutell, R. R.; Gray, M. W.; Schnare, M. N. *Nucleic Acids Res.* **1993**, *21*, 3055–3074.
(17) Proctor, D. J.; Schaak, J. E.; Bevilacqua, J. M.; Falzone, C. J.; Bevilacqua, P. C, *Biochemistry* **2002**, *41*, 12062–12075.
(18) Antao, V. P.; Lai, S. Y.; Tinoco, I., Jr. *Nucleic Acids Res.* **1991**, *19*, 5901–5905.
(19) Antao, V. P.; Tinoco, I., Jr. *Nucleic Acids Res.* **1992**, *20*, 819–824.
(20) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 1045–1061.
(21) Moody, E. M.; Feerrar, J. C.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 7992–7998.

 **10.1021/ja900065e CCC: $40.75** 




C. Partial Charge Comparison

cUUCGg

| Loop Nt. | Atom | Charge | Distance to CBP Atom (Å) |
|---|---|---|---|
| U1 | O4 | -0.4740 | 3.32 (H41), 3.83 (H42) |
| U1 | H3 | +0.3340 | 3.56 (O6) |
| G4 | H21 | +0.3250 | 3.37 (N7) |
| G4 | H22 | +0.3390 | 3.40 (N7) |

cGAAAg

| Loop Nt. | Atom | Charge | Distance to CBP Atom (Å) |
|---|---|---|---|
| G1 | O6 | -0.4580 | 3.17 (H41), 3.19 (H42) |
| G1 | H1 | +0.3360 | 3.39 (O6) |
| G1 | H21 | +0.3250 | 3.70 (N7) |
| G1 | H22 | +0.3390 | 3.43 (N7) |

***Figure 1.*** Molecular mimicry at positions L1 and L4 of GNRA and UNMG tetraloops. (A) 3D-structures highlighting interactions between the loop and the cbp, with key stacking interactions with the cbp shown in light blue. Structures were drawn according to pdb coordinates in Table 1. (B) Characteristic pairing interactions between L1 and L4 of the loop. The atoms that interact with the cbp and are conserved by molecular mimicry are colored according to partial charges: red for negative, blue for positive. (C) Listing of the atoms and partial charges (Discover charge set) of the loop nucleotides that stack with the cbp. Note that each loop presents the same partial charges to the major groove edge of the cbp using unique nucleobases. Also provided are distances between the loop and cbp functionalities.

GAAA[21] and −2.3 kcal/mol for UUCG[17,18] as compared to a GC cbp. Using nearest neighbor parameters for the stem nucleotides, $\Delta\Delta G^{\circ}_{37}$ for the CG to GC swap was predicted to be only ∼0.2 kcal/mol.[17,21,25,26] Selection for heightened tetraloop stability with the CG cbp shows an inverse relationship with protein content for variable loops in 16S rRNA[14] suggesting such loops may have played an important role in an RNA world.

Numerous structural and thermodynamic studies have identified key features of GNRA and UNMG loops.[17,20,21,27−33] In particular, both loops are characterized by extensive base stacking, nucleobase−phosphate interactions, and hydrogen bonding networks, especially between loop positions L1 and L4 (Figure 1A, B). Such interactions, however, do not explain the specific thermodynamic contribution of the loop with the adjoining CG cbp. Furthermore, loop structures do not explain why the CG cbp confers stability to both GNRA and UNMG loops, but not all hairpins, as these two families differ extensively in the identity and distribution of their stabilizing interactions. For example, stacking interactions are more prevalent on the 5′ side of the loop in UNMG and on the 3′

(22) Melchers, W. J.; Zoll, J.; Tessari, M.; Bakhmutov, D. V.; Gmyl, A. P.; Agol, V. I.; Heus, H. A. *RNA* **2006**, *12*, 1671–1682.
(23) It is important to note that, in referring to stable hairpins, "stable" does not mean loop formation is itself thermodynamically favorable but that formation of stable loops is less destabilizing than formation of other loops.
(24) Although the CNMG loops possess the same bifurcated hydrogen bonding and stacking as the UNMG subset of the YNMG loops, the CNMG subset does not present the same partial charges to the cbp as UNMG and GNRA. This loss of mimicry appears to be reflected in the observation that the CG cbp is less important to the CNMG loops as compared to UNMG, with UNMG receiving an ∼1.0 kcal/mol bonus for a CG cbp over a GC cbp as compared with CNMG loops.[17] Our study therefore focuses on UNMG subgroup of the YNMG family.
(25) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719–14735.
(26) Shu, Z.; Bevilacqua, P. C. *Biochemistry* **1999**, *38*, 15369–15379.
(27) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680–682.
(28) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191–194.
(29) Varani, G.; Cheong, C.; Tinoco, I., Jr. *Biochemistry* **1991**, *30*, 3280–3289.
(30) SantaLucia, J., Jr.; Kierzek, R.; Turner, D. H. *Science* **1992**, *256*, 217–219.
(31) Varani, G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379–404.
(32) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H.; Pardi, A. *J. Mol. Biol.* **1996**, *264*, 968–980.
(33) Williams, D. J.; Boots, J. L.; Hall, K. B. *RNA* **2001**, *7*, 44–53.

side in GNRA (Figure 1A), while hydrogen bonding between L1 and L4 gives a bifurcated UG pair for UNMG loops and a sheared GA pair for GNRA loops (Figure 1B).

We noticed that, despite the aforementioned structural differences, UNMG and GNRA loops may act as molecular mimics with respect to their interactions with the cbp. As shown in Figure 1B, amino, imino, and keto groups are similarly positioned over the cbp. We therefore examined the thermodynamic consequences of atomic substitutions in GNRA and UNCG loops via UV melting experiments. Changes in sequence included swapping of the cbp, inserting 3-carbon (C3) linkers between the cbp and L1 or L4 of the loop, and altering functional groups in the cbp. In addition, the salt dependencies of hairpin free energies for the cbp swaps were measured experimentally and were calculated by nonlinear Poisson−Boltzmann (NLPB) theory. Correlation between experimental and calculated values was strong, which supports an important role for electrostatics and molecular mimicry in contributing to the stability of UNMG and GNRA loops in the context of the CG cbp.

## Materials and Methods

**RNA Preparation.** All RNA hairpins have the general sequence 5′-GGAX$L_1L_2L_3L_4$X′UCC where X and X′ are complementary nucleotides in the cbp, and L1−L4 are the tetraloop nucleotides. In the text, oligonucleotides will be referenced in the $XL_1L_2L_3L_4X'$ shorthand since the three initial and three final nucleotides are identical for all sequences. Except for 7-deazaguanine-substituted (7dzG) sequences, oligonucleotides were chemically synthesized by Dharmacon Research Inc., deprotected as per manufacturer's suggestions, and dialyzed into $P_{10}E_{0.1}$ [10 mM sodium phosphate and 0.1 mM $Na_2$EDTA (pH 7.0)] using a microdialysis system (Gibco-BRL Life Technologies). Oligonucleotides substituted with 7dzG were chemically synthesized by the HHMI-Keck Facility at Yale University. The 7dzG-substituted oligonucleotides were deprotected using TBAF in THF and desalted by Sep-Pak (Waters) C18 columns. The oligonucleotides were then purified utilizing preparative TLC (*n*-propanol/ammonium hydroxide/water, 55:35:10) and desalted using Sep-Pak C18 columns before microdialysis into $P_{10}E_{0.1}$ (Gibco-BRL Life Technologies). Purity of all oligonucleotides was confirmed by analytical TLC and estimated to be >95%.

**UV Melting Experiments.** RNA was renatured by heating to 90 °C and cooling to room temperature. All melts were performed in $P_{10}E_{0.1}$, with additional NaCl added as appropriate. UV melts were performed using a Gilford Response II spectrophotometer with a data point acquired every 0.5 °C. Absorbance was recorded at 280 nm, with select melts at 260 nm to ensure similar thermodynamics and two-state behavior. The UV melts were performed over temperature ranges of 95 to 5 °C and 5 to 95 °C and provided data consistent with reversibility of the folding transition.[34] Melt data were fit to a two-state model using sloping baselines and analyzed using a Marquadt algorithm for nonlinear curve fitting in KaleidaGraph v.3.5 (Synergy software). Thermodynamic parameters are the average of at least three independently prepared samples, and the melting temperature was found to be independent of RNA strand concentrations (10 to 100 $\mu$M), consistent with intramolecular hairpin folding.

**Structural Models and NLPB Calculations.** Structural models for NLPB calculations containing only the tetraloop and the cbp were prepared from the atomic coordinates of crystal structures deposited in the PDB files, as listed in Table 1. In some instances, calculations were performed on models with another stem base pair, as described in the Results and Discussion. Resolution of the structures in the main text and the Supporting Information ranged from 2.0 to 3.3 Å and gave similar results, as discussed in the Results and Discussion. For the gUUCGc sequence, we were able to identify only one unique structure in the pdb; additional structural files were therefore generated by running molecular dynamics on the cbp to remove a clash followed by energy minimization. In general, we selected structures that form interactions characteristic of UUCG and GAAA loops and do not participate in protein or RNA tertiary contacts. Hydrogen atoms were added to crystal structure coordinates using Reduce.[35] Details of the NLPB model as well as the finite difference procedure used to calculate electrostatic potentials and electrostatic free energies have been previously described.[36−41] Calculation of electrostatic potentials and electrostatic free energies was performed using Qnifft package 2.2 (http://crystal.med.upenn.edu/software.html). In most calculations, partial charges and radii were from the cvff91 parameter set from the Discover force field (MSI), although some calculations utilized a simple phosphate-only charge set that comprised −0.5 charges on both nonbridging phosphate oxygens with all other partial charges set to zero. For all calculations the following parameters were used: RNA dielectric constant, 2; solvent probe radius, 1.4 Å; ion exclusion radius, 2.0 Å; solvent dielectric constant, 80; and temperature, 310 K. In the calculations, the RNA was placed inside of a $65^3$ lattice, and a two-step focusing procedure from 20−60% lattice fill was used to calculate the electrostatic potentials. The final resolution was ∼1.7 grids/Å, and potentials were iterated to a convergence of $10^{-4}$ kT/e. For each loop−cbp model, the salt-dependent electrostatic free energy was calculated by integrating the ion atmosphere over the lattices from the focusing procedures with charge neutrality maintained within 3%. NLPB output including structures with 3D potential contours or surface potentials was visualized using the ABPS plug-in[42] in PyMol (http://www.pymol.org/).

***Table 1.*** PDB ID of Coordinates Used in NLPB Calculations[a]

| sequence | PDB ID | chain, nucleotides |
|---|---|---|
| cUUCGg | 1NBS[62] | B, 152−157 |
| gUUCGc | 3BWP[63] | A, 338−343 |
| cGAAAg | 2OIU[64] | Q, 10−15 |
| gGAAAc | 1NKW[65] | 0, 2353−2358 |

[a] All are crystal structures.

## Results and Discussion

**Molecular Mimicry Between UNMG and GNRA Tetraloops.** Upon inspection of UNMG and GNRA loops, we noticed that both structures present the same functionalities—keto, imino, and amino groups—and thus the same partial charges to the major groove edge of the CG cbp (Figure 1). Distances between loop and cbp functionalities are similar for both cUUCGg and cGAAAg loops and range from 3.2 to 3.8 Å (tabulated in Figure 1C). These values are the same, on average, as distances between nearest functionalities in stacked bases in the helical stem of these structures, consistent with significant stacking between the loop and cbp. In the GAAA loop, all three functionalities are contributed by the Watson−Crick base pairing face of the G at L1, while in the UUCG loop the keto and imino come

(34) Nakano, S.; Cerrone, A. L.; Bevilacqua, P. C. *Biochemistry* **2003**, *42*, 2982–2994.
(35) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. *J. Mol. Biol.* **1999**, *285*, 1735–1747.
(36) Jayaram, B.; Sharp, K. A.; Honig, B. *Biopolymers* **1989**, *28*, 975–993.
(37) Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1990**, *94*, 7684–7692.
(38) Misra, V. K.; Sharp, K. A.; Friedman, R. A.; Honig, B. *J. Mol. Biol.* **1994**, *238*, 245–263.
(39) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
(40) Misra, V. K.; Honig, B. *Biochemistry* **1996**, *35*, 1115–1124.
(41) Misra, V. K.; Hecht, J. L.; Yang, A. S.; Honig, B. *Biophys. J.* **1998**, *75*, 2262–2273.
(42) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

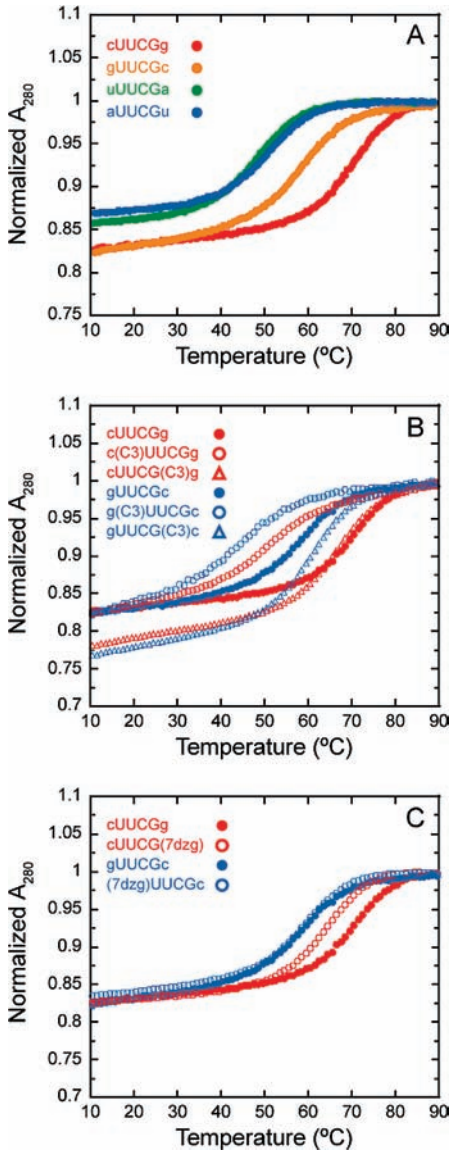



***Figure 2.*** Absorbance vs temperature curves from thermal denaturations of representative UUCG sequences. Shown are plots of normalized absorbance at 280 nm versus temperature. (A) Changing the cbp from CG to another canonical pair. (B) Insertion of C3 linkers. (C) 7dzG-substitution of the cbp. In B and C, only every other data point is shown for clarity of individual curves. Sample absorbance vs temperature curves for thermal denaturations of representative GAAA sequences are provided in the Supporting Information.

from the Watson−Crick base pairing face of the U at L1 with the amino contributed by the G at L4. Observation of this molecular mimicry was the impetus for examining the thermodynamics of loop−cbp interactions for GNRA and UNMG tetraloops. The next four sections describe effects of cbp changes, C3 linker insertions, 7dzG substitutions, and salt concentration increases on loop stability (Figures 2 and 3). These data are then interpreted in terms of NLPB models and electrostatic calculations in the two subsequent sections (Figures 4 and 5).

**Thermodynamic Effects of Changing the Closing Base Pair in UUCG and GAAA Tetraloops.** The extra stability of UUCG and GAAA loops in the context of a CG cbp has been reported but under variable experimental conditions, including synthetic or transcribed RNA and differences in stem sequence, salt concentration, buffer identity, and pH. For example, Proctor et al. used transcribed and synthetic RNAs with four base pair stems in $P_{10}E_{0.1}$, pH 7.0,[17] while Melchers et al. used synthetic RNA with five base pair stems in $P_{10}E_{0.1}$, pH 6.8, and 1 M NaCl.[22] To permit comparison with substituted loops, we conducted melting experiments of unmodified GAAA and UUCG loops under a standardized set of buffer and salt conditions.

The effects of CG and GC cbps on stability were examined for both GAAA and UUCG loops, with all four canonical cbps examined for the latter. Sample melts for UUCG are shown in Figure 2A, and thermodynamic parameters for all sequences are provided in Table 2 under the "Unmodified" subheading. Representative GAAA melts are supplied in the Supporting Information (Figure S1). As expected, significant stabilization is afforded for UUCG and GAAA loops in the context of the CG cbp. In particular, changing the CG cbp to GC is destabilizing, with $\Delta\Delta G^\circ_{37}$ values of +1.86 and +1.54 kcal/mol for UUCG and GAAA, respectively, under the conditions of $P_{10}E_{0.1}$ (Table 2, column 5). The $\Delta\Delta G^\circ_{37}$ values for the CG to GC change are especially significant when compared to nearest neighbor predictions in 1 M NaCl of ~0.2 kcal/mol[25] (Table 2, column 6). Correcting for the slight effect from nearest neighbor parameters for the CG to GC change in the stem leaves an approximately −1.7 and −1.4 kcal/mol bonus for the cbp with UUCG and GAAA loops, respectively (Table 2, column 7).

This extraordinary stability of a CG cbp also holds for comparisons to UA and AU base pairs made in UUCG, where even though a significant destabilization would be expected from nearest neighbor parameters (Table 2, column 6), there remains penalties of approximately +1.5 and +1.8 kcal/mol, respectively (Table 2, column 7). These net free energies are similar to values determined relative to a GC cbp for both UUCG and GAAA loops (Table 2, column 7). Thus, CG is the only one of the four canonical base pairs that affords extra stability to these hairpin loops.

**Thermodynamic Effects of C3 Linker Insertion between the Loop and the Closing Base Pair.** Insertions of C3 linkers have been used previously to investigate energetic coupling between positions within closely related DNA and RNA hairpin loops, as well as between the loop and the cbp.[21,43−45] In those studies, the majority of C3 linker insertions led to minimal perturbation of free energy, which suggested little alteration of the loop structure and therefore extrusion of a flexible C3 linker. Consistent with this interpretation, probing of loop−loop interactions by thermodynamic double mutant cycles revealed that most C3-linker-containing loops had an intact loop structure, with little or no cooperativity between the C3 linker and structure-forming functional groups in the loop. In addition, model building of the C3 linker revealed that the linker can be readily extruded from the loops (data not shown). Such flexibility of the C3 linker is consistent with it having six rotatable single bonds in its backbone. Nonetheless, C3 linker insertions before position 1 of the loop with CG cbps gave a significant thermodynamic penalty to DNA and RNA loop stability, suggesting that when important stacking interactions are present between the loop and cbp the C3 linker can alter them.

(43) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 2032–2033.
(44) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 16285–16293.
(45) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2004**, *126*, 9570–9577.

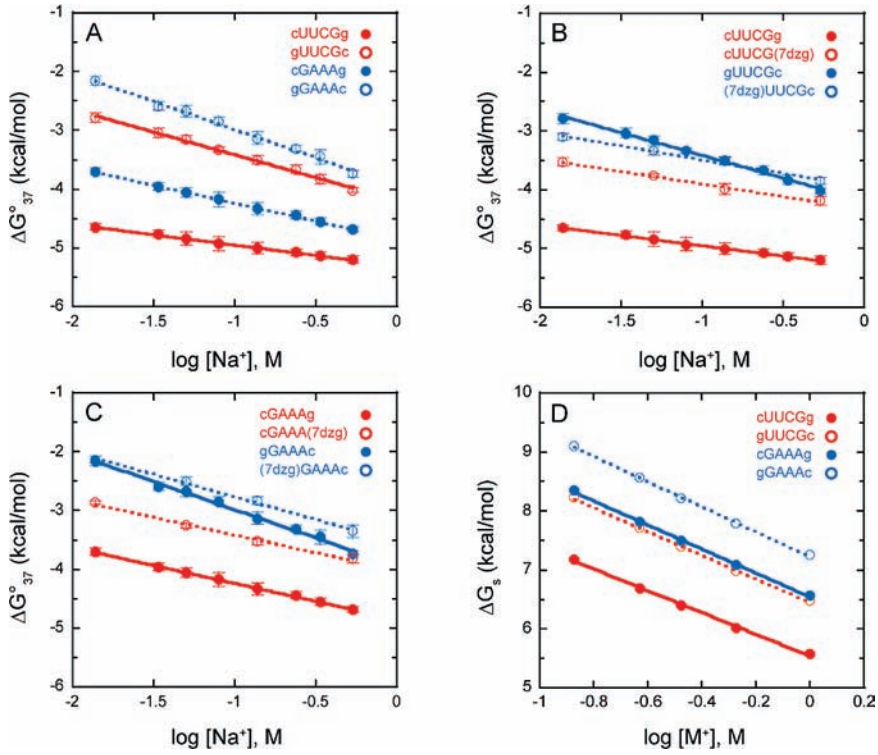



***Figure 3.*** Dependencies of stability on salt concentration for UUCG and GAAA sequences. Shown are plots of $\Delta G^{\circ}_{37}$ vs log[$Na^+$]. Slopes and $R^2$ values for each plot are provided in Table 3. (A) Salt dependence of $\Delta G^{\circ}_{37}$ for UUCG and GAAA with CG and GC cbps. (B) Salt dependence of $\Delta G^{\circ}_{37}$ for UUCG with CG and GC cbps with each cbp substituted with 7dzG. (C) Salt dependence as in panel B but for GAAA loops. (D) Salt-dependent component of electrostatic free energy ($\Delta G_s$) from NLPB calculations.

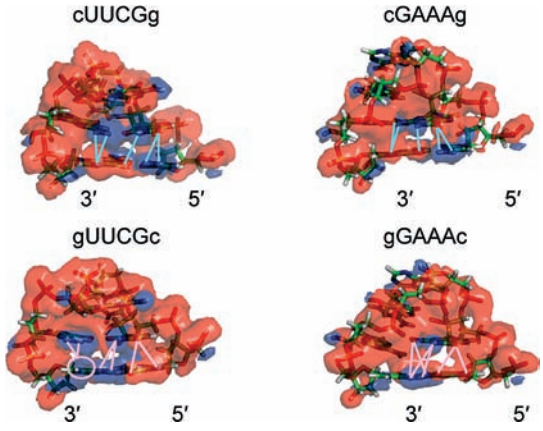



***Figure 4.*** Three-dimensional potential contour maps for UUCG and GAAA tetraloops. Structures were visualized according to pdb coordinates in Table 1. Each structure shows the electrostatic potentials calculated at 0.23 M salt at 40% transparency with the underlying crystal structure in stick representation. UUCG potential contours are −10 kT/e (red) and +10 kT/e (blue), and GAAA potential contours are −20kT/e (red) and +10 kT/e (blue). The gUUCGc structure originated from the pdb coordinates in Table 1, and a clash in the cbp was removed by MD of the cbp for 1 ps, followed by energy minimization and structure cleaning to ensure planarity of bases. Additional MD structures are shown in the Supporting Information (Figure S3). For cUUCGg and cGAAAg, most interactions between L1 and L4 of the loops and the major groove edge of the CG cbp contribute to stability, highlighted in light blue, as per Figure 1A. For gUUCGc and gGAAAc, most interactions between L1 and L4 of the loop and the major groove edge of the GC cbp oppose stability, highlighted in pink.

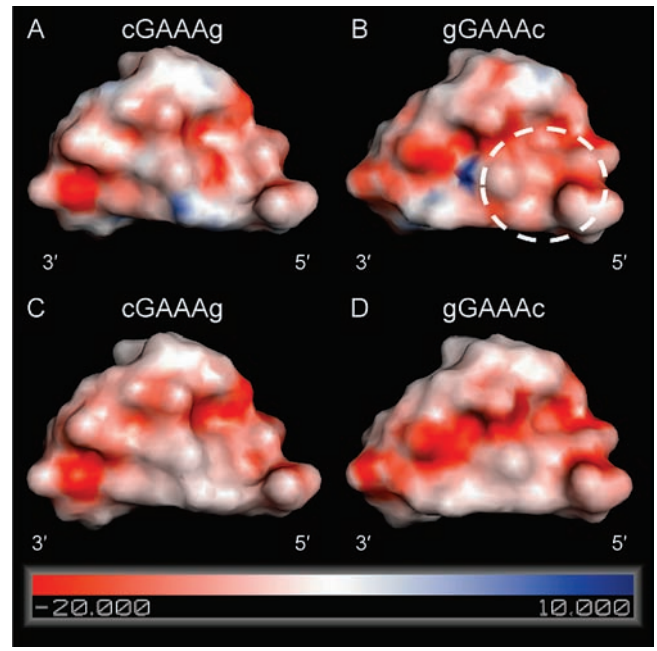



***Figure 5.*** Surface potential maps of GAAA tetraloops. Surface potentials maps from NLPB shown for calculations using the full partial charge parameter set from Discover (panels A and B) or only −0.5 charges on the nonbridging phosphoryl oxygens, with all other charges set to zero (panels C and D). Each calculation is at 0.23 M salt. Notice the increase in negative potential forming away from the phosphates (highlighted in white dashed circle) for gGAAAc with a complete partial charge set (panel B) as compared to the phosphate-only charge set (D). Surface potential maps of UUCG tetraloops are provided in the Supporting Information.

Thermodynamic consequences of C3 linker insertions in GAAA[21] were repeated and extended herein with the gGAAA(C3)c sequence and were also determined for cUUCGg and gUUCGc sequences. Sample UUCG melts with C3 linkers are shown in Figure 2B, and the thermodynamic data are provided in Table 2 under the "C3 Linker" subheading. Representative GAAA melts are supplied in the Supporting Information (Figure S1).

***Table 2.*** Thermodynamic Parameters for Folding of Unmodified and Modified Stable RNA Hairpins[a]

| sequence | $\Delta H^\circ$ (kcal/mol) | $\Delta S^\circ$ (eu) | $\Delta G^\circ_{37}$ (kcal/mol) | $\Delta\Delta G^\circ_{37}$ (kcal/mol) | $\Delta\Delta G^\circ_{37}$ NN (kcal/mol)[b] | $\Delta\Delta\Delta G^\circ_{37}$[c] (kcal/mol) | $T_M$ (°C) | $\Delta T_M$ (°C) |
|---|---|---|---|---|---|---|---|---|
| | | | | Unmodified | | | | |
| cUUCGg | −46.0 ± 0.5 | −133.4 ± 1.5 | −4.64 ± 0.05 | | | | 71.8 ± 0.2 | |
| gUUCGc | −40.8 ± 1.3 | −122.5 ± 4.1 | −2.78 ± 0.08 | 1.86[d] ± 0.09 | 0.16 | 1.70 | 59.7 ± 0.4 | −12.1 ± 0.4[d] |
| uUUCGa | −40.5 ± 1.6 | −124.5 ± 4.8 | −1.85 ± 0.08 | 2.79[d] ± 0.09 | 1.31 | 1.48 | 51.9 ± 0.4 | −19.9 ± 0.4[d] |
| aUUCGu | −40.4 ± 1.6 | −125.0 ± 4.9 | −1.59 ± 0.07 | 3.05[d] ± 0.09 | 1.24 | 1.81 | 49.7 ± 0.4 | −22.1 ± 0.4[d] |
| cGAAAg | −42.3 ± 0.8 | −124.6 ± 2.3 | −3.70 ± 0.06 | | | | 66.7 ± 0.3 | |
| gGAAAc | −37.4 ± 1.0 | −113.6 ± 3.0 | −2.16 ± 0.08 | 1.54[e] ± 0.10 | 0.16 | 1.38 | 56.0 ± 0.8 | −10.7 ± 0.9[e] |
| | | | | C3 Linker | | | | |
| c(C3)UUCGg | −37.6 ± 1.1 | −115.7 ± 3.4 | −1.73 ± 0.07 | 2.91[d] ± 0.09 | | | 52.0 ± 0.5 | −19.9 ± 0.6[d] |
| cUUCG(C3)g | −44.9 ± 1.2 | −131.7 ± 3.4 | −4.05 ± 0.13 | 0.59[d] ± 0.14 | | | 67.7 ± 0.4 | −4.1 ± 0.4[d] |
| g(C3)UUCGc | −31.7 ± 2.6 | −99.4 ± 8.3 | −0.83 ± 0.10 | 1.95[f] ± 0.13 | | | 45.3 ± 0.8 | −14.4 ± 0.9[f] |
| gUUCG(C3)c | −45.1 ± 0.8 | −135.1 ± 2.3 | −3.23 ± 0.08 | −0.45[f] ± 0.11 | | | 60.9 ± 0.5 | 1.2 ± 0.6[f] |
| c(C3)GAAAg | −36.1 ± 1.5 | −110.5 ± 4.6 | −1.83 ± 0.09 | 1.87[e] ± 0.11 | | | 53.5 ± 0.4 | −13.2 ± 0.5[e] |
| cGAAA(C3)g | −43.9 ± 0.6 | −129.2 ± 1.7 | −3.83 ± 0.05 | −0.13[e] ± 0.08 | | | 66.6 ± 0.2 | −0.1 ± 0.4[e] |
| g(C3)GAAAc | −30.5 ± 1.2 | −95.2 ± 3.8 | −0.93 ± 0.03 | 1.23[g] ± 0.09 | | | 46.8 ± 0.4 | −9.2 ± 0.9[g] |
| gGAAA(C3)c | −41.3 ± 0.4 | −125.1 ± 2.0 | −2.52 ± 0.02 | −0.36[g] ± 0.08 | | | 57.1 ± 0.2 | 1.1 ± 0.9[g] |
| | | | | 7-Deazaguanine | | | | |
| cUUCG(7dzg) | −41.9 ± 0.9 | −123.8 ± 2.7 | −3.53 ± 0.07 | 1.11[d] ± 0.09 | | | 65.5 ± 0.4 | −6.3 ± 0.4[d] |
| (7dzg)UUCGc | −44.1 ± 1.0 | −132.3 ± 3.2 | −3.10 ± 0.06 | −0.32[f] ± 0.10 | | | 60.4 ± 0.3 | 0.7 ± 0.5[f] |
| cGAAA(7dzg) | −38.2 ± 1.6 | −113.9 ± 4.1 | −2.86 ± 0.03 | 0.84[e] ± 0.07 | | | 62.2 ± 0.7 | −4.5 ± 0.8[e] |
| (7dzg)GAAAc | −36.4 ± 1.6 | −110.6 ± 4.9 | −2.13 ± 0.06 | 0.03[g] ± 0.10 | | | 56.3 ± 0.8 | 0.3 ± 1.1[g] |

[a] All of these melts were performed in $P_{10}E_{0.1}$, as described in Materials and Methods. [b] This column denotes the portion of $\Delta\Delta G^\circ_{37}$ attributable to changes in the stem, as determined from nearest neighbor parameters.[25] [c] This column provides the difference between $\Delta\Delta G^\circ_{37}$ and $\Delta\Delta G^\circ_{37}$ NN. [d] Values are in comparison to cUUCGg. [e] Values are in comparison to cGAAAg. [f] Values are in comparison to gUUCGc. [g] Values are in comparison to gGAAAc.

There is a large destabilization for 5′-insertion of the C3 linker in c(C3)UUCGg of almost 3 kcal/mol, while a 3′-insertion in cUUCG(C3)g is less destabilizing, with a $\Delta\Delta G^\circ_{37}$ of just 0.6 kcal/mol. Similar trends hold for GAAA loops in the context of a CG cbp, although the magnitude of the destabilizing effect is somewhat less for the 5′-insertion of C3, and slightly stabilizing for the 3′-insertion. For both cUUCGg and cGAAAg, we propose that the 5′-C3 linker interrupts favorable interactions between the cbp and the loop, while the 3′-insertion interrupts somewhat favorable interactions in cUUCGg but not in cGAAAg, where L4 does not interact appreciably with the cbp (Figure 1A and see NLPB section).

In the context of the GC cbp, the 5′-insertion of the C3 linker is destabilizing for both UUCG and GAAA, and again the effect is greater for UUCG. In both instances, the effect is smaller in magnitude than that for a CG cbp by ∼0.6−1.0 kcal/mol. Surprisingly, the 3′-C3 linker insertion is *stabilizing* for both gUUCG(C3)c and gGAAA(C3)c (Figures 2B and S1A; Table 2) with $\Delta\Delta G^\circ_{37}$ values of −0.45 and −0.36 kcal/mol, respectively. These insertions appear to relieve electrostatic *repulsions* between the cbp and loop nucleotides (see NLPB section) consistent with these effects. Opposing effects of 5′- and 3′-C3 linker insertions in gUUCGc are especially apparent in the raw data in Figure 2B, where the 5′-insertion gives a shift to a lower $T_M$ while the 3′-insertion gives a shift to a higher $T_M$.

**Thermodynamic Effects of 7dzG Substitution in the Closing Base Pair.** SantaLucia and co-workers measured the energetic consequences of functional group substitutions throughout a cGCAAg hairpin.[30] Relevant to the present study, they determined that changing the G of the cbp to inosine (I) resulted in a free energy penalty of 1.3 kcal/mol at 70 °C. This functional group substitution deletes the 2-amino group of G, which is in the minor groove of a Watson−Crick base pair. The free energy change of 1.3 kcal/mol is similar to values reported for CG to CI changes at the end of model helices.[46] Thus, the simplest interpretation of these data is that the minor groove face of the cbp does not make significant contributions to stability beyond simple helical interactions. This is congruent with our molecular mimicry model, which involves interactions with the major groove edge of the CG cbp (Figure 1). We therefore focused our experimental efforts on functional groups that interact with the major groove of the CG cbp.

To probe the energetics of the loop−cbp interaction further, we made 7dzG substitutions in the G of the cbp for UUCG and GAAA loops, which change the N7 to a CH. Importantly, these functional group substitutions probe loop stability, but they do so without changes in hydrogen bonding. Substitutions of 7dzG have been made previously in the stem of a stable DNA triloop, which caused a destabilization of +0.34 kcal/mol in $P_{10}E_{0.1}$.[44] Substitution of 7dzG in RNA duplexes has similar, small destabilizations of +0.14 to + 0.36 kcal/mol in 1 M NaCl.[47]

Sample melts for cbp substituted UUCG sequences are shown in Figure 2C, and the thermodynamic parameters are provided in Table 2 under the "7-deazaguanine" subheading. Representative GAAA melts are provided in the Supporting Information (Figure S1). For cUUCG(7dzg), there was a significant destabilization of +1.1 kcal/mol relative to the unsubstituted cbp, which is similar in value to the C3 linker destabilization of cUUCG(C3)g. In contrast, for (7dzg)UUCGc there was a *stabilization* of −0.32 kcal/mol relative to the unsubstituted cbp; moreover, the magnitude of this stabilization could be underestimated because of the aforementioned intrinsic destabilization due to the 7dzG substitution.[47] Thus, 7dzG substitution gives opposing effects for the CG and GC cbps.

The 7dzG substitution to give cGAAA(7dzg) is destabilizing relative to the unsubstituted cbp, as it was for cUUCG(7dzg), but slightly smaller in magnitude, as was the case for both the

(46) Turner, D. H.; Sugimoto, N.; Kierzek, R.; Dreiker, S. D. *J. Am. Chem. Soc.* **1987**, *109*, 3783–3785.
(47) Burkard, M. E.; Turner, D. H. *Biochemistry* **2000**, *39*, 11748–11762.

***Table 3.*** Linear Fit Parameters for $\Delta G^\circ_{37}$ vs log[$Na^+$] for UUCG and GAAA Tetraloops[a]

**Plot 3A: Effects of cbp swaps (from experiments)**

| *Sequence* | *Intercept*[b] *(kcal/mol)* | *Slope (kcal/mol)* | *ΔSlope (kcal/mol)* | $R^2$ |
|---|---|---|---|---|
| cUUCGg ● | -5.30 ± 0.01 | -0.35 ± 0.01 | - | 0.9974 |
| gUUCGc ○ | -4.18 ± 0.03 | -0.77 ± 0.02 | -0.42 ± 0.02[c] | 0.9950 |
| cGAAAg ● | -4.84 ± 0.01 | -0.61 ± 0.01 | - | 0.9985 |
| gGAAAc ○ | -3.94 ± 0.04 | -0.96 ± 0.03 | -0.35 ± 0.03[d] | 0.9934 |

**Plot 3B: Effects of 7dzG in cbp for UUCG (from experiments)**

| *Sequence* | *Intercept*[b] *(kcal/mol)* | *Slope (kcal/mol)* | *ΔSlope (kcal/mol)* | $R^2$ |
|---|---|---|---|---|
| cUUCGg ● | -5.30 ± 0.01 | -0.35 ± 0.01 | - | 0.9974 |
| cUUCG(7dzg) ○ | -4.32 ± 0.03 | -0.43 ± 0.02 | -0.08 ± 0.02[c] | 0.9937 |
| gUUCGc ● | -4.18 ± 0.03 | -0.77 ± 0.01 | - | 0.9950 |
| (7dzg)UUCGc ○ | -3.95 ± 0.04 | -0.47 ± 0.03 | +0.30 ± 0.04[e] | 0.9921 |

**Plot 3C: Effects of 7dzG in cbp for GAAA (from experiments)**

| *Sequence* | *Intercept*[b] *(kcal/mol)* | *Slope (kcal/mol)* | *ΔSlope (kcal/mol)* | $R^2$ |
|---|---|---|---|---|
| cGAAAg ● | -4.84 ± 0.01 | -0.61 ± 0.01 | - | 0.9985 |
| cGAAA(7dzg) ○ | -4.01 ± 0.05 | -0.60 ± 0.04 | +0.01 ± 0.04[d] | 0.9922 |
| gGAAAc ● | -3.94 ± 0.04 | -0.96 ± 0.03 | - | 0.9934 |
| (7dzg)GAAAc ○ | -3.53 ± 0.04 | -0.77 ± 0.04 | +0.19 ± 0.05[f] | 0.9949 |

**Plot 3D: Effects of cbp swaps (from calculations)**

| *Sequence* | *Intercept*[b] (kcal/mol) | *Slope (kcal/mol)* | *ΔSlope (kcal/mol)* | $R^2$ |
|---|---|---|---|---|
| cUUCGg ● | 5.54 | -1.85 | - | 0.9982 |
| gUUCGc ○ | 6.45 | -2.01 | -0.16[c] | 0.9988 |
| cGAAAg ● | 6.55 | -2.03 | | 0.9989 |
| gGAAAc ○ | 7.23 | -2.13 | -0.10[d] | 0.9992 |

[a] Results are from fits to data plotted in Figure 3, and the appropriate panel of Figure 3 is indicated in the table. The Δslope error is propagated as the square root of the sum of squares from curve fit errors. [b] The *y*-intercept value is defined as $\Delta G^\circ_{37}$ extrapolated to 1 M [$Na^+$]. [c] The difference in slopes is relative to cUUCGg. [d] The difference in slopes is relative to cGAAAg. [e] The difference in slopes is relative to gUUCGc. [f] The difference in slopes is relative to gGAAAc.

cbp and C3 linker substitutions (Table 2). For (7dzg)GAAAc there was no significant effect of the substitution ($\Delta\Delta G^\circ_{37}$ = 0.03 kcal/mol), although there could be a slight stabilization as with (7dzg)UUCGc, owing to the aforementioned intrinsic 7dzG destabilization of the stem. Results from C3 linker and 7dzG insertions in this and the preceding sections can be accommodated by NLPB calculations as described below.

**UV Melting Experiments: Salt Dependence of Δ*G*.** For UUCG and GAAA loops, hairpin free energies were determined as a function of sodium ion concentration for CG and GC cbps with and without 7dzG substitutions.[48] Results of the salt-dependent UV melting experiments are shown in Figure 3, and linear fitting parameters are provided in Table 3. For both loops, the lowest salt dependence of the hairpin free energy was in the context of the CG cbp (Figure 3A; Table 3). In addition, the salt dependence for cUUCGg was lower than that for cGAAAg (−0.35 ± 0.01 vs −0.61 ± 0.01 kcal/mol), and the difference in slope between CG and GC cbp was greater for UUCG than GAAA (−0.42 ± 0.02 vs −0.35 ± 0.03). These results mirror the pattern of slightly more pronounced thermodynamic effects in UUCG than GAAA loops. The salt dependencies for UUCG and GAAA loops were also determined as a function of potassium ion concentration (data not shown). Although the magnitude of the slopes was slightly smaller for both UUCG and GAAA loops, the Δslope values for the CG to GC swap were very similar to those in sodium (−0.41 and −0.36 kcal/mol for UUCG and GAAA, respectively), consistent with the salt effect not being specific to a particular monovalent cation. In summary, there is an inverse correlation between thermodynamic stability and salt dependence for these four sequences. As a consequence, the CG cbp contributes somewhat less to hairpin stability at higher salt concentration.

For cbps with 7dzG substitutions, the slope of the salt dependence plots also correlated inversely with the thermodynamic effect of the substitution. For example, when 7dzG stabilized the loop, the salt dependence of $\Delta G^\circ_{37}$ decreased (i.e., the slope became less negative), whereas when 7dzG destabilized the loop, the salt dependence increased slightly (i.e., the slope became more negative) or changed very little. In addition, it appears that the 7dzG substitutions had greater effects on slope in the context of the GC cbp, where 7dzG changed the magnitude of the slope by +0.30 (from −0.77 to −0.47 kcal/mol) for (7dzg)UUCGc and by +0.19 (from −0.96 to −0.77 kcal/mol) for (7dzg)GAAAc. In contrast, the 7dzG substitutions in the context of the CG cbp increased the magnitude of the slope by only 0.08 kcal/mol for UUCG, and had no experimentally significant effect in GAAA. One possibility is that for

(48) The salt dependence for the C3 linkers was not conducted, as the salt dependence could be artificially higher due to the presence of an extra phosphate group in the hairpin with the addition of the C3 linker.

the CG cbps, while the 7dzG substitution eliminates favorable interactions leading to destabilization (Table 2), numerous other cbp−loop interactions (Figure 1A) maintain the overall loop structure leading to no significant change in the salt dependence (Table 3).

**Calculations of NLPB Potentials Support Experimental Observations.** In the preceding four sections, experimental data were presented for effects of cbp swaps, C3 linker insertions, 7dzG substitutions, and ionic strength on UUCG and GAAA loops. In the next two sections we integrate these data into NLPB models and electrostatic potential calculations. Three-dimensional potential contours were mapped onto UUCG and GAAA loop−cbp structures at 40% transparency (Figure 4 and Supporting Information, Figures S2−S5). The electrostatic potential contours were produced from the potential map output of the NLPB calculations. The vertical (stacking) juxtaposing of loop and cbp potentials is congruent with the thermodynamic parameters from UV melting experiments of model sequences with cbp swaps, C3 linker insertions, and 7dzG substitutions in the following ways.

For cUUCGg and cGAAAg, the favorable loop−cbp electrostatics highlighted in light blue in Figure 1 are also highlighted in light blue in Figure 4 in the context of the 3D potential contour maps. Notable is the electrostatic interaction of positive (blue) and negative (red) potentials between regions of mimicry in L1 and L4 and the CG cbp. For example the O4 of U1 (negative) in UUCG and the O6 of G1 (negative) in GAAA are attracted to the H4's of C (positive) in the CG cbp. Turning to the GC counterparts, there is a general increase in the number of repulsive interactions between L1/L4 and the cbp, with more like-potentials positioned over each other. For example, the O4 of U1 in UUCG and the O6 of G1 in GAAA are repelled by the negative potentials at the O6 and N7 of G in the GC cbp. Examination of additional structures for a given sequence revealed similar loop−cbp interactions, which is consistent with an important role for molecular mimicry in providing stability to loops with a CG cbp (see Supporting Information, Figures S2−S5).

As described, insertions of C3 linkers into cUUCGg in c(C3)UUCGg and in cUUCG(C3)g were both destabilizing, +2.9 and +0.6 kcal/mol, respectively (Table 2). As can be seen from the electrostatic potential contours (Figures 4 and S2), both insertions have the potential to disrupt attractive interactions. The 5′-C3 linker insertion is more destabilizing, perhaps because it interrupts the interactions of *two* functional groups, the O4 and H3 of U1, with oppositely charged functionalities in the cbp, while the insertion in cUUCG(C3)g interrupts just the interaction of the positive H2's of G4 with the negative potential of N7 of the cbp G (Figure 1). Substitution of 7dzG in the CG cbp has a similar effect (+1.1 kcal/mol) as the 3′-C3 linker insertion, consistent with perturbation of the same attractive interaction.

For cGAAAg, the 5′-C3 insertion was also destabilizing (+1.9 kcal/mol) and the 3D contours qualitatively agree with this result (Figures 4 and S4). This C3 insertion has the ability to interrupt attractions of the cbp to the three functional groups, all at G1, that act as molecular mimics of the functionalities in U1 and G4 of UUCG. The 3′-C3 insertion in cGAAA(C3)g, on the other hand, is slightly stabilizing (−0.13 kcal/mol), probably because L4 does not participate in molecular mimicry for the GAAA loops (Figure 1B). In the case of cGAAA(7dzg), the loop is destabilized by 0.84 kcal/mol, similar to what was observed for cUUCG(7dzg). For GAAA, the attractions between H2's of G at L1 and the N7 of the cbp G are deleted (Figure 4).

For gUUCGc and gGAAAc, repulsions between the loop and the cbp are highlighted with pink lines in the 3D potential contours (Figures 4, S3, S5). For gUUCGc, C3 insertion is destabilizing 5′ of the loop (+1.95 kcal/mol) and stabilizing 3′ of the loop (−0.45 kcal/mol), as described above (Table 2). The structure and electrostatics support the first result by revealing that an attraction between H3 of U at L1 and the O6 of G of the cbp (not highlighted) could be removed. However, this destabilization is ∼1 kcal/mol lesser in magnitude than that observed in c(C3)UUCGg, perhaps because while breaking a favorable interaction the C3 linker also relieves a repulsion of negative potentials between the O4 of U at L1 and O6/N7 of G in the cbp (Figures 4, S3). The same interaction was probed via 7dzG substitution, and removal of the N7 alone is stabilizing by −0.3 kcal/mol, supporting this interpretation. The 3′-C3 linker insertion is stabilizing for gUUCG(C3)c (−0.45 kcal/mol), and the structure suggests that the basis for this is relief of the repulsive interaction between the positive potential of G4 H2's and multiple atoms with positive potential in C of the GC cbp. Overall, 7dzG effects and C3 linker insertions in cUUCGg and gUUCGc are compatible with NLPB calculations.

For gGAAAc, the 5′-C3 insertion is destabilizing (+1.2 kcal/mol) but less so than in cGAAAg (+1.9 kcal/mol). A favorable interaction between H1 of G1 and O6 of G in the cbp (not highlighted) could be removed, but so could repulsions between G1 and the cbp, including O6 of G1 and O6/N7 of G in the cbp (Figures 4 and S5). This result and structural reasoning are similar to those for the 5′-C3 insertion in gUUCGc. The 7dzG substitution for (7dzg)GAAAc has a lessened effect (+0.03 kcal/mol) compared to the stabilization observed for (7dzg)UUCGc (−0.3 kcal/mol), which could be due to elimination of a smaller repulsion, although possible rearrangement of the loop limits our ability to be fully quantitative. In addition, we may be underestimating the stabilization of the substitution due to intrinsic destabilization of the stem by 7dzG,[47] as mentioned above. As with gUUCG(C3)c, 3′-linker insertion in gGAAA(C3)c is stabilizing (−0.4 kcal/mol), and from the structure this could arise from relieving repulsive positive potentials from H1 and H2's of G1 and H4's of C and H1 of G in the cbp. Overall, interacting potentials have the ability to explain the stacking stability in UUCG and GAAA loops. This is qualitatively similar to effects described by Burkard and Turner for 3′-end stacking.[49]

Lastly, it is instructive to consider previously published experimental observations on the dynamics of cUUCGg and gUUCGc loops.[20] Williams and Hall measured imino proton spectra for these two loops as a function of temperature and found that the resonance for G at L4 broadens at temperatures below the melting temperature for the GC cbp loop but not the CG cbp loop. These observations were interpreted as likely being due to solvent exchange of the N1 imino proton at L4 in the bifurcated hydrogen bond (see Figure 1B) of the GC cbp loop. The thermodynamic measurements and electrostatic calculations performed herein do not speak directly to loop dynamics. However, it is often observed that less thermodynamically stable structures have greater intermolecular motions. In this context electrostatic repulsions observed in gUUCGc between the loop and cbp (Figure 4) may prevent the formation of locally stabilizing attractions resulting in a more dynamic structure.

(49) Burkard, M. E.; Kierzek, R.; Turner, D. H. *J. Mol. Biol.* **1999**, *290*, 967–982.

***Table 4.*** NLPB Calculated Thermodynamic Parameters[a]

| sequence | $\Delta\Delta G_c$ (kcal/mol) | $\Delta\Delta G_p$ (kcal/mol) | $\Delta\Delta G_{ns}$[c] (kcal/mol) | $\Delta\Delta G_s$ (1 M) (kcal/mol) | $\Delta\Delta G_{el}$ (1 M)[d] (kcal/mol) |
|---|---|---|---|---|---|
| Full Charge Set[b] | | | | | |
| cUUCGg | – | – | – | – | – |
| gUUCGc | 40.71 | −28.56 | 12.15 | 0.91 | 13.06 |
| cGAAAg | – | – | – | – | – |
| gGAAAc | 31.88 | −14.02 | 17.86 | 0.68 | 18.54 |
| Phosphate-Only Charge Set[e] | | | | | |
| cUUCGg | – | – | – | | |
| gUUCGc | 44.66 | −29.50 | 15.16 | | |
| cGAAAg | – | – | – | | |
| gGAAAc | 11.90 | −8.68 | 3.22 | | |

[a] The following parameters are differences in energies outputted from NLPB calculations. [b] Charges on all atoms according to the Discover force field. [c] $\Delta\Delta G_{ns}$ calculated as the sum of $\Delta\Delta G_c$ and $\Delta\Delta G_p$. [d] The difference in electrostatic free energies includes the difference in the salt-dependent free energy at 1 M monovalent salt. $\Delta\Delta G_{el}$ is calculated as the sum of $\Delta\Delta G_{ns}$ and $\Delta\Delta G_s$ (1 M) according to eq 1. [e] Nonbridging phosphoryl oxygens assigned −0.5 charge with all other charges set to zero.

**Calculations of Electrostatic Free Energy Support Experimental Observations.** The total electrostatic free energy can be written as follows

$$\Delta G_{el} = \Delta G_{ns} + \Delta G_s \quad (1)$$

where the electrostatic free energy ($\Delta G_{el}$) is partitioned into nonsalt-dependent ($\Delta G_{ns}$) and salt-dependent ($\Delta G_s$) and terms.[40] The salt-dependent free energy is the interaction of macromolecular charges with mobile ions, while $\Delta G_{ns}$ can be further partitioned

$$\Delta G_{ns} = \Delta G_c + \Delta G_p \quad (2)$$

where $\Delta G_c$ is the energy of the intramolecular Coulombic interactions and $\Delta G_p$ is the aqueous solvation free energy of the nucleic acid structure.[40] Substitution of eq 2 into eq 1 gives

$$\Delta G_{el} = \Delta G_c + \Delta G_p + \Delta G_s \quad (3)$$

$$\text{or } \Delta\Delta G_{el} = \Delta\Delta G_c + \Delta\Delta G_p + \Delta\Delta G_s \quad (4)$$

Values for $\Delta\Delta G_c$, $\Delta\Delta G_p$, $\Delta\Delta G_{ns}$, $\Delta\Delta G_s$ (1 M NaCl), and $\Delta\Delta G_{el}$ (at 1 M NaCl) using the complete Discover partial charge parameter set are provided in Table 4 for the CG-GC cbp swap for UUCG and GAAA. These calculations were first performed on a representative structure (Tables 1 and 4, Figure 5) and then repeated on additional structures (Table S1, Figures S2−S5), which were in agreement (see Materials and Methods for further details). In addition, values for $\Delta\Delta G_c$, $\Delta\Delta G_p$, $\Delta\Delta G_{ns}$ from calculation using only −0.5 charges on each nonbridging phosphoryl oxygen are provided for comparison (Table 4).

For UUCG, the nonsalt-dependent contribution to the electrostatic free energy, $\Delta\Delta G_{ns}$, destabilizes structures with a GC cbp relative to their CG counterparts by ∼12 kcal/mol (Table 4, "Full Charge Set"). Partitioning $\Delta\Delta G_{ns}$ into $\Delta\Delta G_c$ and $\Delta\Delta G_p$ reveals that the intramolecular Coulombic interactions ($\Delta\Delta G_c$) strongly destabilize structures with GC cbps relative to their CG counterparts for the UUCG loops (∼+41 kcal/mol) (Table 4). In an effort to uncover the origin of the large $\Delta\Delta G_c$ effect, we calculated the accessible surface area of the two UUCG loops (Table 5). The accessible surface area of the UUCG loop decreases by 48 Å$^2$ for a GC cbp relative to the CG cbp; in addition, the interphosphorus distances in the GC cbp loop were closer on average by 0.7 Å than in the CG cbp loop (data not shown). Thus, gUUCGc is significantly more compact than cUUCGg.

***Table 5.*** Accessible Surface Area of Tetraloop Sequences[a]

| squence | surface area (Å$^2$) | Δsurface area (Å$^2$) |
|---|---|---|
| cUUCGg | 1488 | – |
| gUUCGc | 1440 | −48 |
| cGAAAg | 1417 | – |
| gGAAAc | 1433 | +16 |

[a] Surface areas from NLPB output.

To test whether compaction of the UUCG with a GC cbp leads to phosphate repulsions, we repeated the NLPB calculations with a simplified potential set, in which we simply assigned half of the −1 charge to each nonbridging phosphoryl oxygen (Table 4, "Phosphate-Only Charge Set"). The $\Delta\Delta G_c$ values for CG to GC cbp calculated using the simplified potential set were very similar to those calculated using the full Discover potential set, +40.71 versus +44.66 kcal/mol for the full and phosphate-only charge set, respectively. Thus, the intramolecular Coulombic destabilization of UUCG loops with GC cbps rises primarily from interphosphate repulsions in the more compact structures with a GC closing pair. Contributions from less favorable base−base interactions such as repulsion between the positive potential of G4 H2's and multiple atoms with positive potential in C of the cbp for gUUCGc are therefore small (Supporting Information, Figure S3).

In contrast to $\Delta\Delta G_c$, the solvation free energy ($\Delta\Delta G_p$) favors structures with GC cbps, with $\Delta\Delta G_p \approx -29$ kcal/mol (Table 4, "Full Charge Set"). Thus, solvation is more favorable for the more compact gUUCGc loop, but the magnitude of the favorable $\Delta\Delta G_p$ term is not large enough to compensate for $\Delta\Delta G_c$, as $\Delta\Delta G_{ns} \approx 12$ kcal/mol. Moreover, the salt-dependent contribution, $\Delta\Delta G_s$, also destabilizes gUUCGc relative to cUUCGg, even at 1 M monovalent salt, such that $\Delta\Delta G_{el}$ favors the CG cbp ($\Delta\Delta G_{el} \approx 13$ kcal/mol), qualitatively agreeing with experimental thermodynamics parameters. We carried out the same calculations on a set of four structures for each sequence. These calculations demonstrated a discrete range of $\Delta\Delta G_{el}$ values (Supporting Information, Table S1) and an identical description of the electrostatic forces underlying the stability of the different structures. These results demonstrate the consistency and generality of the electrostatic description of our system. In summary, UUCG loops follow a relatively simple model, in which the GC cbp gives more compact loops that are better solvated but suffer greater repulsion among the compressed phosphates.

We now turn to NLPB calculations on GAAA loops. For GAAA, the nonsalt-dependent contribution to the electrostatic free energy, $\Delta\Delta G_{ns}$, destabilizes structures with a GC cbp relative to their CG counterparts (Table 4, "Full Charge Set"), a similar result to the UUCG loops. However, partitioning $\Delta\Delta G_{ns}$ into $\Delta\Delta G_c$ and $\Delta\Delta G_p$ terms reveals that phosphates and bases contribute differently for GAAA and UUCG loops. The intramolecular Coulombic interactions $\Delta\Delta G_c$ also strongly destabilize structures with GC cbps relative to their CG counterparts, but $\Delta\Delta G_c$ is significantly larger for calculations that assign potentials to all atoms (Table 4, "Full Charge Set") rather than to the nonbridging phosphoryl oxygens alone (Table 4, "Phosphate-Only Charge Set"), ∼32 and ∼12 kcal/mol, respectively. This trend contrasts with that observed for UUCG loops, wherein both parameter sets gave similar values. Moreover, the GAAA loop with a GC cbp is somewhat *less* compact

than its CG counterpart, with an increase of 16 Å$^2$ (Table 5); this observation again contrasts with that for UUCG loops.

Examining the surface potentials for cGAAAg and gGAAAc reveals that although cGAAAg is more compact than gGAAAc, the phosphate potentials of gGAAAc are directed toward the center of the structure, forming a diagonal "ring" of negative potential from the 5′ to 3′ end of the structure, not present in cGAAAg (Figure 5). In addition, a significant buildup of negative potential is observed near the base of the loop structure in gGAAAc calculated with full charges (Figure 5B, dashed lines) that is not present in cGAAAg (Figure 5A). This effect was not seen in calculations on UUCG loops (Supporting Information, Figure S6). Thus, repulsive interactions of the bases contribute significantly to $\Delta\Delta G_c$ in GAAA loops.

The increase in negative potential in gGAAAc likely contributes to stabilization relative to cGAAAg by solvation, since $\Delta\Delta G_p \approx -14$ kcal/mol for the CG to GC cbp switch (Table 4). Although gGAAAc is not more compact than cGAAAg (Table 5), the ring of negative potential leads to a region of high surface charge density (Figure 5), which explains its more favorable interactions with aqueous solvent. Nonetheless, the greater solvation of gGAAAc is not enough to compensate for the intramolecular Coulombic interactions, as $\Delta\Delta G_{ns} \approx +18$ kcal/mol for the CG to GC cbp switch. In addition, $\Delta\Delta G_s$ is destabilizing for gGAAAc relative to cGAAAg, even at 1 M monovalent salt. As a result, $\Delta\Delta G_{el}$ favors the CG cbp for GAAA loops ($\Delta\Delta G_{el} = +18.5$ kcal/mol), qualitatively agreeing with experimental parameters. These calculations were also repeated on a set of four structures for each sequence. The calculated $\Delta\Delta G_{el}$ values were again found to fall into a well-defined range (Supporting Information, Table S1) and provided a consistent description of the electrostatic forces involved in stabilizing the different structures. These results further support the general nature of our description of the electrostatic free energies involved in mimicry. In summary, NLPB calculations show that, for both UUCG and GAAA loops, the GC cbp form has a higher surface charge density, although it arises from different sources: for UUCG it is due to loop compaction, while for GAAA it is due to a change in loop configuration.

To afford comparison with experimental data, the salt-dependent contribution to the electrostatic free energy, $\Delta\Delta G_s$, was calculated for representative structures (Table 1) for monovalent salt concentrations between 0.13 and 1.0 M. The calculated value of $\Delta\Delta G_s$ was plotted as a function of log[M$^+$] between 0.13 and 1.0 M bulk monovalent salt concentration and compared to experimental data. As shown in Figure 3D, $\Delta G_s$ decreases linearly with log[M$^+$], and the loops with GC cbps have a greater dependence on log[M$^+$], as in the experimental results. Furthermore the calculated change in slope between GC and CG cbp is on the same order of magnitude and has the same algebraic sign as the experimentally observed change in slope for both UUCG and GAAA (Table 3). As mentioned above, GC cbp loops have a higher surface charge density, as revealed by NLPB calculations, which is consistent with an accumulation of salt at the surface and a concomitant greater dependence of free energy on bulk salt concentration. This is similar to trends that were observed in the salt-dependent transition of DNA from the B to Z forms.[40]

To address whether the salt dependence of the tetraloops was affected by the presence of additional stem nucleotides, we included a stem base pair beyond the cbp and recalculated the salt-dependent contribution to the electrostatic free energy. Although the slope values changed as expected, the Δslope values were virtually identical to those obtained using just the loop and cbp (data not shown). This result suggests that slope effects are attributable to interactions between the loop and cbp rather than within the stem.

The calculations in this study are meant to provide a semiquantitative description of the role of electrostatics in the mimicry interactions of a set of tetraloops under specific experimental conditions. We expect that $Mg^{2+}$ may also stabilize the folded structure of these tetraloops, since divalent cations tend to accumulate in and bind to pockets of high negative electrostatic potential on the surface of RNAs.[50−52] For example, the region of negative potential near the phosphates for L2 and L3 of the cGAAAg loop (upper right-hand corner of the display in Figure 5A) is consistent with a $Mg^{2+}$ binding site that was modeled from $^{31}$P NMR and phosphorothioate substitutions.[53] Observation of more of a negative ring in the gGAAAc potential calculations (Figure 5B), as described above, is suggestive of another region where $Mg^{2+}$ may accumulate and bind. While our calculations generally include the effects of a "diffusely bound" monovalent cation layer, we do not directly address how the ionic size and valence may modulate the interactions we observe, in part because thermodynamic data for such comparisons are lacking.

## Conclusions

Prior studies established the phylogenetic conservation, three-dimensional structure, and thermodynamic stability of stable tetraloops.[17,20,21,27−33] However, the molecular basis for preference of a CG cbp has remained unclear. We therefore attempted to define this preference. Although they have extensive structural differences, we observed that the UNMG and GNRA tetraloop families are molecular mimics with respect to the functional groups and partial charges stacked onto the cbp. Such mimicry is observed on the major groove edge of the cbp, but not its minor groove edge, consistent with prior functional group changes in the minor groove not having contributions beyond simple helical interactions.[30,46] The mimicry interactions were therefore probed by experiments and calculations and ultimately helped explain why CG is the thermodynamically preferred cbp for these loops. Lesser roles played by atoms on the minor groove edge of the cbp or the middle of the bases cannot be excluded.

Apparently, these disparate loops achieve stability by arranging potentials at the base of their loops in a very similar fashion. Both sequences use the free portions of the Watson−Crick base pairing face of L1, which is similar between U and G with their high imino nitrogen p$K_a$ values, to interact with the cbp. This behavior is reminiscent of the ability of A and C, which have similar *low* imino nitrogen p$K_a$ values, to substitute for each other in general acid−base catalysis in the HDV ribozyme.[54,55]

The CG closing pair is critical for the extra stability of the GNRA and UNMG loops and makes a contribution to stability that is much greater than can be explained by nearest neighbor parameters. Moreover, the stabilizing effect of the CG is not based simply on the positioning of the pyrimidine and purine

(50) Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. *Nat. Struct. Biol.* **1999**, *6*, 1055–1061.
(51) Misra, V. K.; Draper, D. E. *J. Mol. Biol.* **2000**, *299*, 813–825.
(52) Misra, V. K.; Draper, D. E. *J. Mol. Biol.* **2002**, *317*, 507–521.
(53) Maderia, M.; Horton, T. E.; DeRose, V. J. *Biochemistry* **2000**, *39*, 8193–8200.
(54) Perrotta, A. T.; Shih, I.; Been, M. D. *Science* **1999**, *286*, 123–126.
(55) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493–1497.

bases in the cbp, as a UA base pair is 1.5 kcal/mol less stable than can be accounted for by stem changes for UUCG loops (Table 2, column 7). Calculations support experiments by showing that loops with a CG cbp are electrostatically more stable but have a shallower dependence of stability on salt.

Insertion of C3 linkers and substitution of 7dzG into the cbp probed the loop−cbp interaction at the nucleotide and functional group levels. In accord with NLPB calculations, interruption or deletion of attractive interactions destabilized hairpins, while interruption or deletion of repulsive interactions stabilized hairpins. Loops less favored by electrostatics exhibited a steeper dependence of free energy on salt concentration, as determined by both experiments and NLPB calculations. The Δslope values between CG and GC cbps from the UV melts and the NLPB calculations are in good qualitative agreement, which shows the utility of NLPB calculations in studying smaller RNA systems.

Overall, substitutions exhibited the same trends for both tetraloop families but were slightly more enhanced for UUCG than GAAA loops, as also revealed in a larger enthalpy gain, albeit at a somewhat larger entropy loss; this may occur because the former arrange potentials from both L1 and L4 to interact with the cbp. Loop compaction upon introduction of a GC cbp is important for UUCG loop behavior, as evidenced by calculations, while changes in loop configuration are important for GAAA loops, with a "ring" of negative potential present in gGAAAc but not in cGAAAg (Figure 5). Formation of this potential may be favored in biology as it may facilitate tertiary interactions, such as interactions of tetraloops with tetraloop receptors.[2,56,57] The ability of tetraloop receptors to tune tetraloop stability is also possible.

Molecular mimicry can be extended to other RNA motifs such as tandem GA mismatches. When the symmetric tandem GA mismatches are closed by CG base pairs, the GA adopts a sheared pairing conformation, similar to positions L1 and L4 in a GNRA loop.[58] However, when the GA mismatches are closed by GC base pairs, the GA adopts an imino-hydrogen bonding arrangement.[59] These trends are consistent with the many repulsive interactions of a sheared GA with a GC cbp found herein.

Nature often employs molecular mimicry in nucleic acids and proteins on a grand scale,[60,61] but molecular mimicry on a smaller scale, such as found in UNMG and GNRA, may be a more common method of stabilizing otherwise unrelated structural motifs in RNA and DNA. Given that stable tetraloops motifs are most common in regions of the ribosome devoid of proteins, it is possible that molecular mimicry may have been especially advantageous in the RNA world.

**Acknowledgment.** This work was supported by NSF Grant MCB-0527102 to P.C.B.

**Supporting Information Available:** Melting curves for representative GAAA sequences, additional 3D contour potential maps for UUCG and GAAA sequences, surface potential maps of UUCG tetraloops, and average $\Delta\Delta G_{el}$ for UUCG and GAAA loops with CG and GC cbp. This information is available free of charge via the Internet at http://pubs.acs.org/.

JA900065E

(56) Costa, M.; Deme, E.; Jacquier, A.; Michel, F. *J. Mol. Biol.* **1997**, *267*, 520–536.
(57) Davis, J. H.; Tonelli, M.; Scott, L. G.; Jaeger, L.; Williamson, J. R.; Butcher, S. E. *J. Mol. Biol.* **2005**, *351*, 371–382.
(58) SantaLucia, J., Jr.; Turner, D. H. *Biochemistry* **1993**, *32*, 12612–12623.
(59) Wu, M.; Turner, D. H. *Biochemistry* **1996**, *35*, 9677–9689.
(60) Nissen, P.; Kjeldgaard, M.; Nyborg, J. *EMBO J.* **2000**, *19*, 489–495.
(61) Tsonis, P. A.; Dwivedi, B. *Biochim. Biophys. Acta* **2008**, *1783*, 177–187.
(62) Krasilnikov, A. S.; Yang, X.; Pan, T.; Mondragon, A. *Nature* **2003**, *421*, 760–764.
(63) Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. *Science* **2008**, *320*, 77–82.
(64) Robertson, M. P.; Scott, W. G. *Science* **2007**, *315*, 1549–1553.
(65) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679–688.