RNA Tetraloop Folding Reveals Tension between Backbone Restraints and Molecular Interactions

Srividya Mohan,[†,§] Chiaolong Hsiao,[†,§] Jessica C. Bowman,[†,§] Roger Wartell,[‡,§] and Loren Dean Williams*,[†,§]

School of Chemistry and Biochemistry, School of Biology, and Parker H. Petit Institute of Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332-0400

Received May 20, 2010; E-mail: loren.williams@chemistry.gatech.edu

Abstract: In RNA, A-form helices are commonly terminated by tetraloops or 3′ dangling ends. Aside from helices themselves, these helix-breaking motifs appear to be among the most frequent and repetitive structural elements of large folded RNAs. We show here that within a frequent type of tetraloop, cGNRAg (G is guanine, N is any base, R is purine, A is adenine), a tension exists between the backbone torsional energy of the loop and the energy contributed by molecular interactions (stacking and pairing). A model in which favorable bond rotamers are opposed by favorable stacking and pairing interactions is consistent with our observation that release of torsional restraints upon conversion of one or more loop riboses to more flexible trimethylene phosphate(s) contributes favorably to the enthalpy of folding. This effect presumably results from improved stacking and hydrogen-bonding interactions upon release of torsional restraints. The most obvious possibility for improving molecular interactions is a repositioning of A, which is proximal to the unfavorable torsion angles in native cGNRAg tetraloops, and which is unstacked on the 3′ side and unpaired (it forms a single hydrogen bond with the opposing G). This tension between favorable bond rotamers and favorable molecular interactions may be representative of a general evolutionary strategy to prevent achievement of deep and irreversible thermodynamic wells in folded RNAs. Finally, we observe a simple stacking substructure with conserved geometry and sequence that forms a scaffold for both tetraloops and 3′ dangling ends. It seems that simple substructures can build RNA motifs, which combine to establish the fundamental architecture of RNA.

Introduction

Native RNA structures are determined by base pairing, base stacking, and bond rotameric preferences and by interactions with water, cations, and proteins. It is thought that coincident minima of energy of these factors characterizes A-form helices and common motifs such as tetraloops,[1,2] kink-turns,[3] dangling ends,[4,5] and AA platforms[6] (reviewed by Doudna,[7] Moore,[8] and Westhof[9]).

In folded RNAs, helices are commonly terminated by tetraloops[2] or 3′ dangling ends.[4,5] Aside from helices themselves, these helix-breaking motifs appear to be the most frequent and repetitive structural elements within large folded RNAs. For example, in three-dimensional structures of ribosomes,[10−13] the A-form helices that constitute around 50% of the RNA[14,15] frequently terminate with tetraloops and 3′ dangling ends.[16−18]

Data here support a model in which backbone rotamers oppose optimum molecular interactions within the most frequent type of RNA tetraloop. This tension may be part of a general evolutionary strategy to prevent deep and irreversible thermodynamic wells in folded RNAs. In addition, we observe a simple stacking substructure with conserved geometry and sequence that forms a scaffold for both tetraloops and 3′ dangling ends.

† School of Chemistry and Biochemistry.
‡ School of Biology.
§ Parker H. Petit Institute of Bioengineering and Biosciences.
(1) Woese, C. R.; Winker, S.; Gutell, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467.
(2) Woese, C. R.; Gutell, R.; Gupta, R.; Noller, H. F. *Microbiol. Rev.* **1983**, *47*, 621.
(3) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 4214.
(4) Martin, F. H.; Uhlenbeck, O. C.; Doty, P. *J. Mol. Biol.* **1971**, *57*, 201.
(5) Uhlenbeck, O. C.; Martin, F. H.; Doty, P. *J. Mol. Biol.* **1971**, *57*, 217.
(6) Basu, S.; Rambo, R. P.; Strauss-Soukup, J.; Cate, J. H.; Ferre-D'Amare, A. R.; Strobel, S. A.; Doudna, J. A. *Nat. Struct. Biol.* **1998**, *5*, 986.
(7) Batey, R. T.; Rambo, R. P.; Doudna, J. A. *Angew. Chem., Int. Ed.* **1999**, *38*, 2326.
(8) Moore, P. B. *Annu. Rev. Biochem.* **1999**, *68*, 287.
(9) Leontis, N. B.; Westhof, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300.
(10) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.
(11) Selmer, M.; Dunham, C. M.; Murphy, F. V. t.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935.
(12) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679.
(13) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. *Science* **2005**, *310*, 827.
(14) Woese, C. R.; Magrum, L. J.; Gupta, R.; Siegel, R. B.; Stahl, D. A.; Kop, J.; Crawford, N.; Brosius, J.; Gutell, R.; Hogan, J. J.; Noller, H. F. *Nucleic Acids Res.* **1980**, *8*, 2275.
(15) Noller, H. F.; Kop, J.; Wheaton, V.; Brosius, J.; Gutell, R. R.; Kopylov, A. M.; Dohme, F.; Herr, W.; Stahl, D. A.; Gupta, R.; Waese, C. R. *Nucleic Acids Res.* **1981**, *9*, 6167.
(16) Isaksson, J.; Chattopadhyaya, J. *Biochemistry* **2005**, *44*, 5390.
(17) Mohan, S.; Hsiao, C.; VanDeusen, H.; Gallagher, R.; Krohn, E.; Wartell, R.; Williams, L. D. *J. Phys. Chem. B* **2009**, *113*, 2614.
(18) Hsiao, C.; Mohan, S.; Kalahar, B. K.; Williams, L. D. *Mol. Biol. Evol.* **2009**, *26*, 2415.

Thus, it seems that simple substructures can build RNA motifs, which combine to establish the fundamental architecture of RNA.

Tetraloops. Helix termination by short loops is accomplished by (a) GNRA tetraloops,[1,2] (b) CUYG tetraloops,[1] (c) UNCG tetraloops,[19] (d) GANC tetraloops,[20] (e) AGNN tetraloops,[21,22] (f) UNR triloops,[23] and (g) lone-pair triloops.[24] GNRA tetraloops (G is guanine, N is any base, R is purine, A is adenine, Figure 1) are small and stable and fold as independent units.[19,25−27] GNRA tetraloops are found in ribosomes (above), the Tetrahymena group I intron,[28,29] the Signal Recognition Particle,[30] and other RNAs. GNRA tetraloops are thought to (a) initiate folding of complex RNA molecules,[19] (b) stabilize helical stems,[19,31] and (c) provide recognition elements for binding to RNA and proteins.[28,32−35] Three-dimensional analysis indicates that there are 14 GNRA tetraloops in the 23S rRNA of *Haloarcula marismortui* and 10 in the 23S rRNA of *Thermus thermophilus*.[18,36] GNRA tetraloops are usually closed by c·g base-pairs[1,2] (and are henceforth called cGNRAg tetraloops, where lowercase "c" and "g" indicate the c·g base-pair at the helix termini and uppercase "GNRA" indicates the loop nucleotides).

Several features of cGNRAg tetraloops have remained unexplained until now. The fourth residue of the loop, A, is unstacked on the 3′ side[36,37] and is unpaired (it forms only a single hydrogen bond with the opposing G). A slight change in position of A would allow it to pair and stack. Why does it not? Indeed the original NMR models of tetraloops, which are inconsistent with X-ray derived structures, incorrectly showed A to be paired and stacked.

Dangling Ends. A second broadly used strategy for helix termination in RNA is the dangling end[4,5] (Figure 2). 3′ Dangling ends are observed in siRNAs,[38] tRNAs,[39] and rRNAs.[17,40] There are 14 3′ dangling ends in 16S rRNA of *T. thermophilus* and 12 in the 23S rRNA of *H. marismortui*.[17] A dangling end is an unpaired nucleotide linked to, and stacked on, the terminus of a double-stranded helix. In RNA, a dangling end confers significantly greater stability when attached to the 3′ helix terminus than when attached to the 5′ terminus. The stability of a 3′ dangling end is modulated by sequence,[41−44] stack length,[44,45] and geometry.[16,46] A dangling end does not require strand termination; dangling ends can be embedded in large RNAs.

Materials and Methods

Thermodynamic Characterization. Melting curves for tetraloops (native and modified, Figures 3 and 4) were obtained by monitoring the UV absorbance at 260 nm as a function of temperature with a Varian Cary-1E UV spectrophotometer with a temperature controller. HPLC-purified RNA oligomers, characterized by mass spectrometry, were obtained from IDT (www.idtdna.com). Each tetraloop evaluated here (Table 1) contains a three-base-pair stem, with a c·g closing base-pair. The parent sequence, which forms a native tetraloop, is ggcGCAAgcc. This consensus sequence, obtained by three-dimensional data-mining (below), was previously characterized in solution by Santa Lucia and Turner.[47]

Melting transitions (Figure 4G) were determined in 0.010 M sodium phosphate (pH 7.0), 1.0 mM EDTA, 1 M NaCl. Prior to melting, samples were heated to 85 °C for several minutes and then quick-cooled on ice. The concentration of the RNA strand (assessed from OD_{260nm} at 90 °C) was 5 μM. In addition, RNA melting curves were obtained over a 50-fold range in concentration to determine if the melting transitions are unimolecular.

Thermodynamic parameters were determined from reversible, concentration-independent melting curves by fitting the data to a two-state model.[48,49] Absorbance vs temperature data were expressed as the fraction of total strand in the single-stranded state, $\theta_s(T)$, as shown in eq 1.

$$\theta_s(T) = [A(T) - A_{pre}(T)]/[A_{post}(T) - A_{pre}(T)] \qquad (1)$$

where $A(T)$ is the absorbance at temperature T, and $A_{pre}(T)$ and $A_{post}(T)$ are the pre- and post-transition linear baselines of the melting curves. The enthalpy and entropy changes of the transition were evaluated from the $\theta_s(T)$ curves using eq 2.[49]

$$-\Delta H + T\Delta S = RT \ln[\theta_s(T)/(1 - \theta_s(T))] \qquad (2)$$

ΔH and ΔS values that best fit eq 2 were determined using nonlinear regression (SigmaPlot). Values of ΔG at 37 °C were calculated from ΔH and ΔS.

Three-Dimensional Data-Mining. Three-dimensional structures of tetraloops and 3′ dangling ends were extracted from high-resolution crystal structures. The 23S rRNA from the *H. marismortui* large subunit (LSU)[3,10] and the 16S rRNA from the *T. thermophilus* assembled ribosome[11] are the highest resolution and largest independent RNA structures in the database. The LSU of *H. marismortui*, with 2914 observable 23S rRNA residues, has a

(19) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; d'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I., Jr.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1364.
(20) Keating, K. S.; Toor, N.; Pyle, A. M. *J. Mol. Biol.* **2008**, *383*, 475.
(21) Butcher, S. E.; Dieckmann, T.; Feigon, J. *J. Mol. Biol.* **1997**, *268*, 348.
(22) Wu, H.; Yang, P. K.; Butcher, S. E.; Kang, S.; Chanfreau, G.; Feigon, J. *EMBO J.* **2001**, *20*, 7240.
(23) Gutell, R. R.; Cannone, J. J.; Konings, D.; Gautheret, D. *J. Mol. Biol.* **2000**, *300*, 791.
(24) Lee, J. C.; Cannone, J. J.; Gutell, R. R. *J. Mol. Biol.* **2003**, *325*, 65.
(25) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680.
(26) Varani, G.; Cheong, C.; Tinoco, I., Jr.; Wimberly, B. *Biochemistry* **1991**, *30*, 3280.
(27) Antao, V. P.; Tinoco, I., Jr. *Nucleic Acids Res.* **1992**, *20*, 819.
(28) 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.
(29) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1696.
(30) Batey, R. T.; Rambo, R. P.; Lucast, L.; Rha, B.; Doudna, J. A. *Science* **2000**, *287*, 1232.
(31) Selinger, D.; Liao, X.; Wise, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5409.
(32) Michel, F.; Westhof, E. *J. Mol. Biol.* **1990**, *216*, 585.
(33) Jaeger, L.; Michel, F.; Westhof, E. *J. Mol. Biol.* **1994**, *236*, 1271.
(34) Puglisi, J. D.; Tan, R. Y.; Calnan, B. J.; Frankel, A. D.; Williamson, J. R. *Science* **1992**, *257*, 76.
(35) Qin, P. Z.; Butcher, S. E.; Feigon, J.; Hubbell, W. L. *Biochemistry* **2001**, *40*, 6929.
(36) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481.
(37) Correll, C. C.; Swinger, K. *RNA* **2003**, *9*, 355.
(38) Elbashir, S. M.; Lendeckel, W.; Tuschl, T. *Genes Dev.* **2001**, *15*, 188.
(39) Quigley, G. J.; Rich, A. *Science* **1976**, *194*, 796.

(40) Acharya, P.; Acharya, S.; Cheruku, P.; Amirkhanov, N. V.; Foldesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2003**, *125*, 9948.
(41) Romaniuk, P. J.; Hughes, D. W.; Gregoire, R. J.; Neilson, T.; Bell, R. A. *J. Am. Chem. Soc.* **1978**, *100*, 3971.
(42) Freier, S. M.; Alkema, D.; Sinclair, A.; Neilson, T.; Turner, D. H. *Biochemistry* **1985**, *24*, 4533.
(43) Sugimoto, N.; Kierzek, R.; Turner, D. H. *Biochemistry* **1987**, *26*, 4554.
(44) O'Toole, A. S.; Miller, S.; Serra, M. J. *RNA* **2005**, *11*, 512.
(45) Ohmichi, T.; Nakano, S.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2002**, *124*, 10367.
(46) Burkard, M. E.; Kierzek, R.; Turner, D. H. *J. Mol. Biol.* **1999**, *290*, 967.
(47) Santa Lucia, J., Jr.; Kierzek, R.; Turner, D. *Science* **1992**, *256*, 217.
(48) Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601.
(49) Puglisi, J. D.; Tinoco, I., Jr. *Methods Enzymol.* **1989**, *180*, 304.

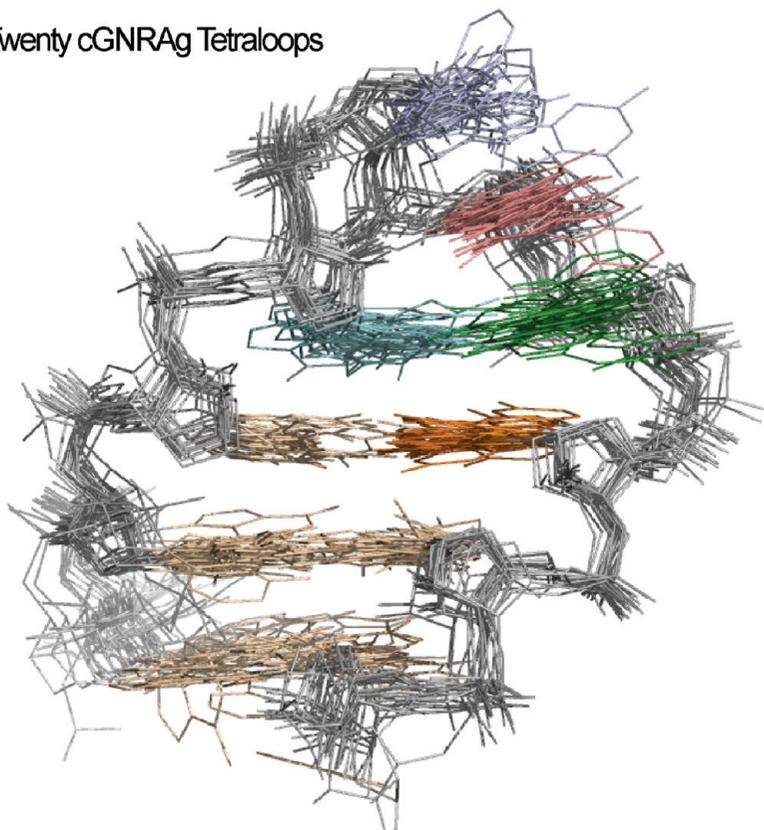## A) One cGNRAg Tetraloop
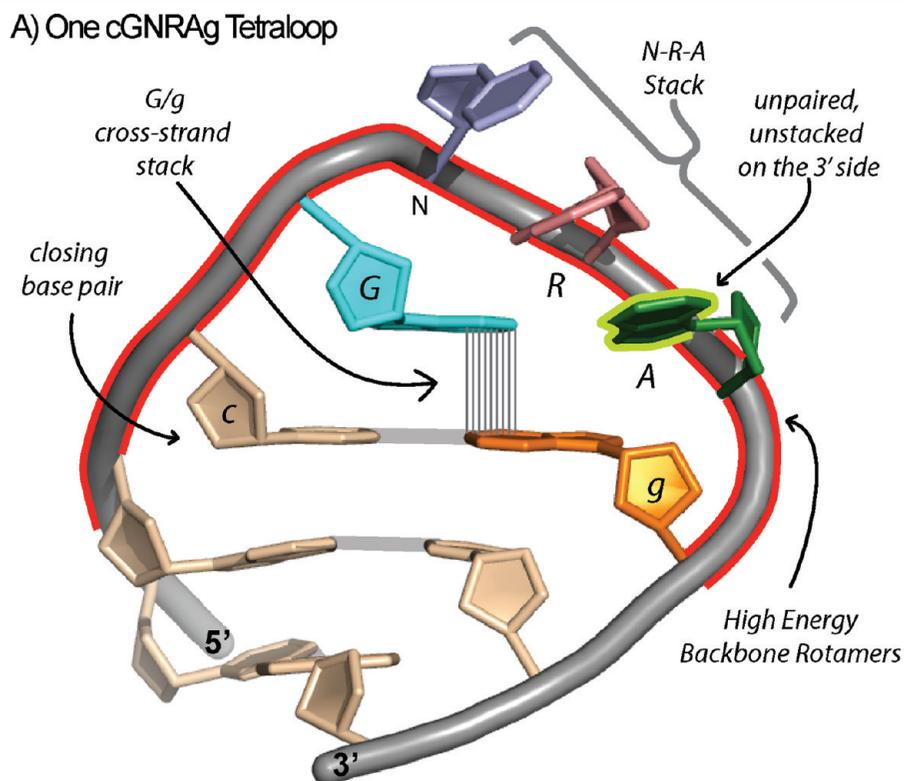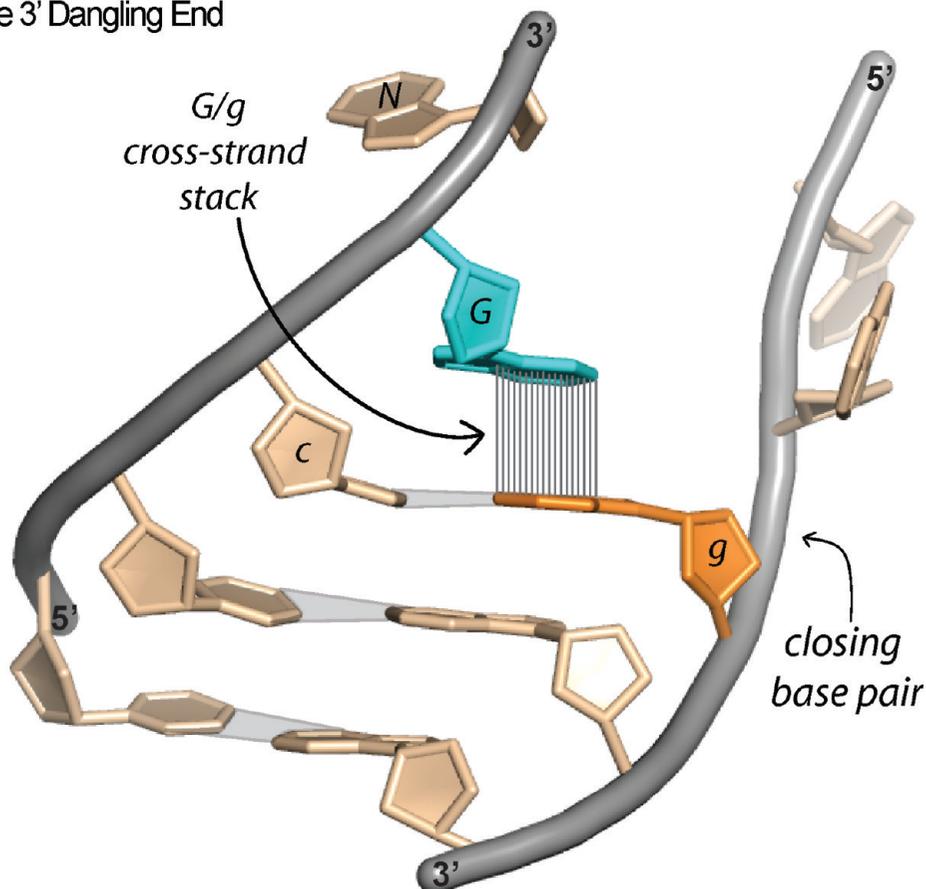

B) Twenty cGNRAg Tetraloops

Figure 1. (A) A representative cGNRAg tetraloop: G, cyan; N, purple; R, pink; and A, green. The tetraloop contains four unpaired loop nucleotides (GNRA) that link the antiparallel helical strands. The closing base-pair of the helix is the consensus c·g base-pair (lowercase letters: c, brown; g, orange). The first residue of the loop (G) is stacked predominantly on the cross-strand g of the closing base-pair and forms a single hydrogen bond to A (the hydrogen bond is not shown). The cross-strand G/g stack is denoted by vertical lines. The rotameric stress of the backbone is indicated by red highlight. The unstacked and unpaired A is highlighted in green. Gray shading indicates base pairing. (B) Superimposition of backbone atoms of 20 cGNRAg tetraloops identified in 1JJ2 (LSU) and 2J00 (SSU only) ribosomal PDB entries.
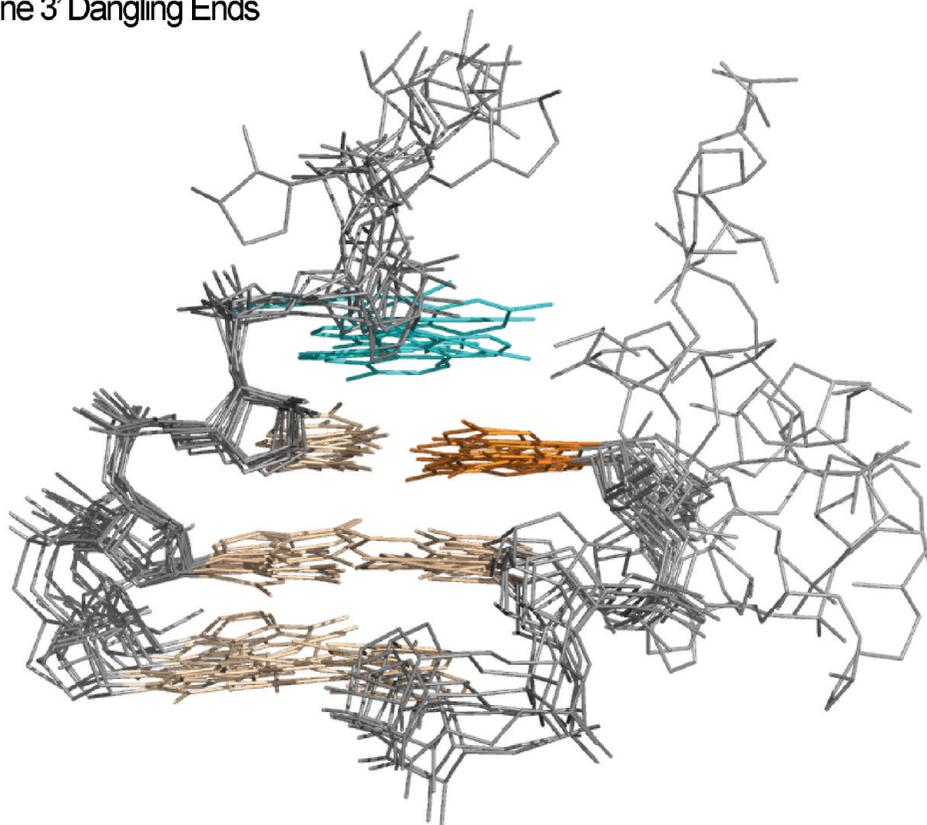
Figure 2. (A) A representative 3′ dangling end, which is composed of a helical stem with a 3′ stacked nucleotide and two single-stranded regions. The consensus closing base-pair is c•g (lowercase letters: c, brown; g, orange, as in Figure 1A). The dangling consensus G (cyan) is stacked predominantly on the cross-strand g of the closing base-pair. The cross-strand G/g stack is denoted by vertical lines. Gray shading indicates base pairing. (B) Superimposition of nine 3′ dangling ends identified in 1JJ2 (LSU) and 2J00 (SSU only) ribosomal PDB entries.
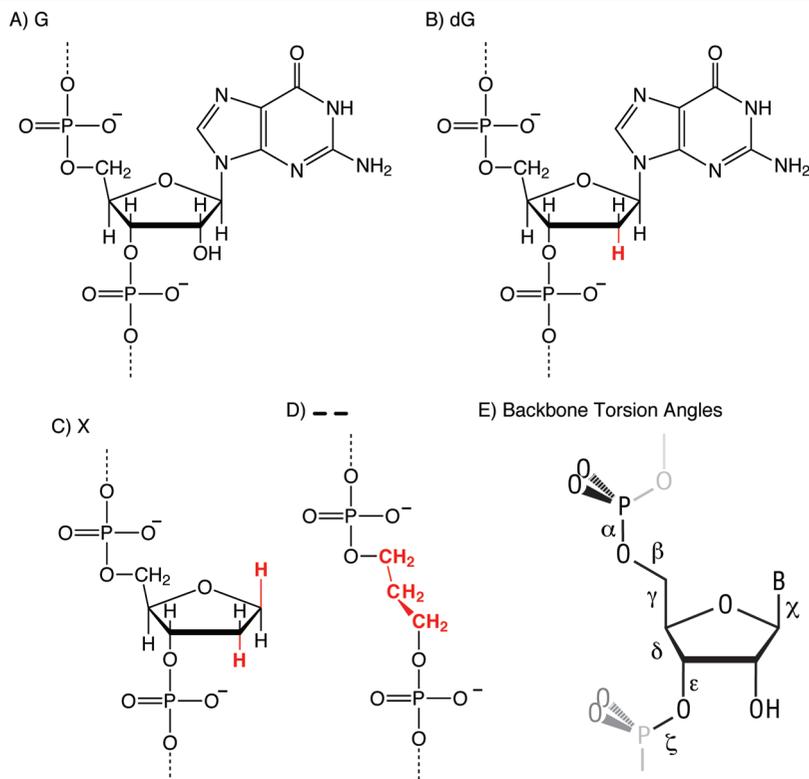
Figure 3. Modified nucleotides incorporated into tetraloops: (A) native guanosine ribonucleotide, (B) guanosine 2′-deoxyribonucleotide, (C) abasic 2′-deoxyribonucleotide, (D) trimethylene phosphate, and (E) RNA backbone torsion angles.

resolution of 2.4 Å. The assembled ribosome of *T. thermophilus*, with 1581 observable 16S rRNA residues, has a resolution of 2.8 Å. Tetraloops and 3′ dangling ends were identified by molecular interactions and coarse-grain analysis as described[17,36] and confirmed by visual inspection.

Three-Dimensional Structures of cGNRAg Tetraloops. Tetraloops found within the 23S rRNA from *H. marismortui* and the 16S rRNA from *T. thermophilus* were initially clustered exclusively by three-dimensional structure, as illustrated by the tetraloop family tree.[36] The tetraloops are further constrained here by a requirement that the first residue at the 5′ end of the loop is G. Tetraloops in which the first residue of the loop is U, which are reasonably common,[36] were excluded. In addition, structures with mismatches, insertions, or clips within the loop or stem, or stems shorter than three base-pairs, were excluded. This culling resulted in 20 tetraloops, with consensus sequence cGNRAg.

Three-Dimensional Structures of 3′ Dangling Ends. Dangling ends were identified and clustered as described.[17] A dangling end here requires a single-stranded region of RNA linked to a double-stranded region. Single-stranded regions are defined by at least three contiguous nucleotides that do not interact with the opposing strand through hydrogen-bonding interactions, although the two strands emerge from a common helical stem. The helical stem must contain at least three contiguous base-pairs. Perturbations such as bulges and mismatches (other than G-U) are not allowed in the stem. The single-strand/double-strand junctions obtained by these criteria reveal a strong preference for 3′ dangling ends over blunt ends or 5′ dangling ends.[17] Nine 3′ dangling ends with a single stacked nucleotide were identified in the 16S rRNA of *T. thermophilus* and 23S rRNA of *H. marismortui*.[17]

Conformation and Goodness of Fit. To quantitatively assess the degree of conservation and variation of structure between cGNRAg tetraloops and 3′ dangling ends, root mean square deviations (rmsd's) of atomic positions were calculated after superimposition. Among tetraloops, superimpositions used the backbone atoms of the GNRA loop plus those of the two adjacent base-pairs of the helix. The superimposition of 20 cGNRAg

tetraloops is shown in Figure 1B. Among 3′ dangling ends, superimpositions used the backbone atoms of the 3′ dangling residue plus those of the two adjacent base-pairs. The superimposition of nine 3′ dangling ends is shown in Figure 2B. To compare cGNRAg tetraloops with 3′ dangling ends, the backbone atoms of the 3′ dangling residue were superimposed on those of the G of the tetraloop. The corresponding backbone atoms of the two adjacent pairs were also included in the superimposition. A resulting superimposition of a cGNRAg tetraloop on a 3′ dangling end is shown in Figure 5A. The similarity of the two structures is readily apparent. The simultaneous superimposition of 20 cGNRAg tetraloops on nine 3′ dangling ends is shown in Figure 5B.

Sequence. Base identities were tabulated for tetraloops and 3′ dangling ends extracted from the three-dimensional database. The base frequencies (C, G, A, or U) at each position for the two motifs are represented as pie diagrams (Figure 6). Except for the G of the cGNRAg tetraloop, which is fixed, sequence was not used as a search criterion and so is a dependent variable.

Stacking. The extent of stacking in three-dimensional structures of tetraloops and 3′ dangling ends was calculated using the program 3DNA.[50] The "area of overlap", including exocyclic atoms, was used as a measure of the extent of base-stacking. Based on overlap values, the mode of stacking of G on the closing base-pair was classified as either predominantly same-strand or predominantly cross-strand.

Results

Here we determine salient structural and thermodynamic features of cGNRAg tetraloops, of 3′ dangling ends, and of structures intermediate between them. Oligonucleotides 1−6 (Table 1) each appear to form a stem-loop structure as indicated by (a) molecularity of folding, (b) agreement between observed and calculated thermodynamic parameters, and (c) the closure of experimental thermodynamic cycles.

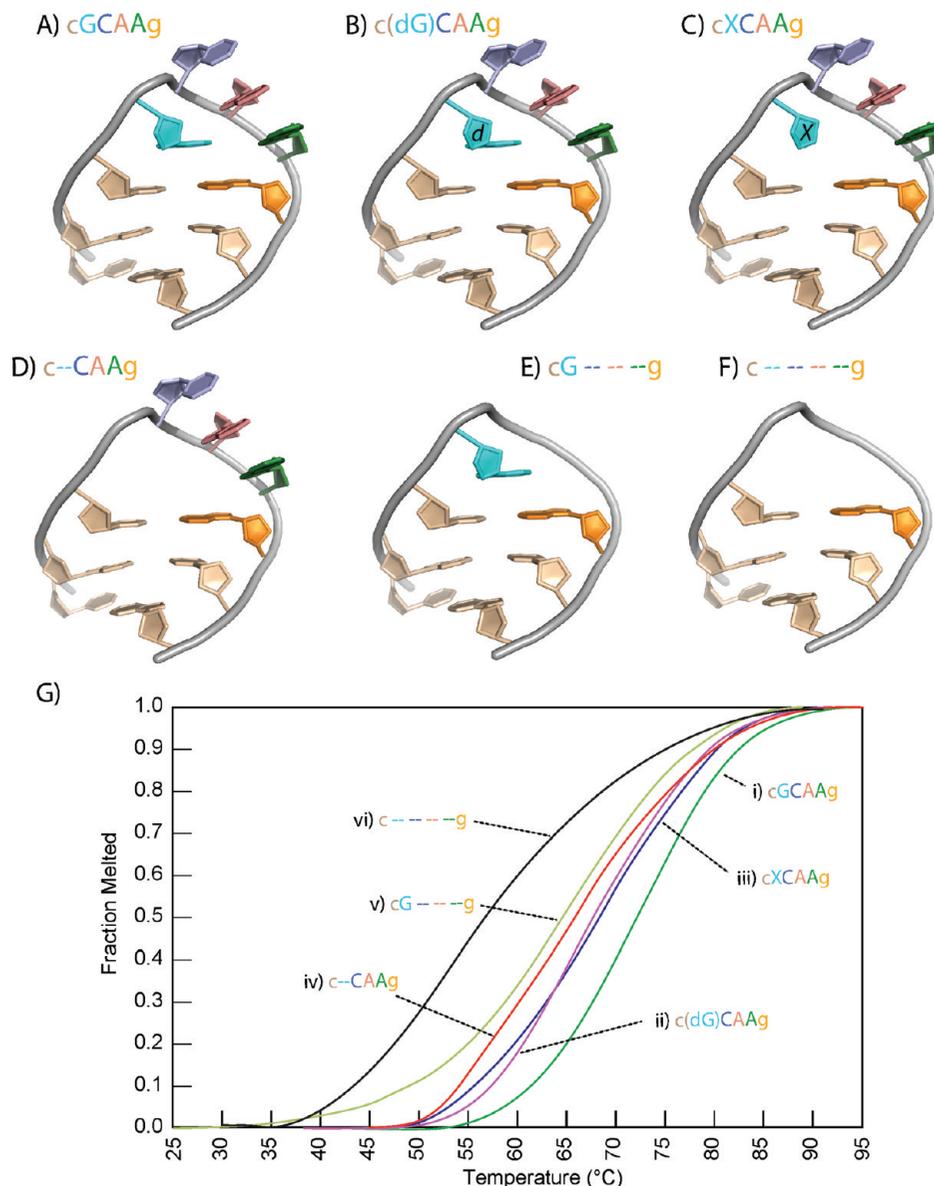(50) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108.

Figure 4. Modified cGNRAg tetraloops and their thermal transitions. (A) Native ggcGNRAgcc tetraloop. (B) The G of the tetraloop is converted to a deoxyribose (dG). (C) The G is converted to abasic 2′-deoxyribose, indicated by X. (D) The G is converted to a trimethylene phosphate, which effectively removes both the base and the sugar. (E) N, R, and A are each converted to a trimethylene phosphate, removing three bases and three riboses. (F) The four nucleotides of the loop, GNRA, are converted to four trimethylene phosphates, removing all four loop bases and riboses. (G) Thermal melting of native and modified cGNRAg tetraloops. Each curve represents equilibrium unimolecular denaturation. The green line (i) is the melt of native ggcGCAAgcc, the pink line (ii) is ggc(dG)CAAgcc, the blue line (iii) is ggcXCAAgcc, the orange line (iv) is ggc--CAAgcc, the light green line (v) is ggcG -- -- -- gcc, and the black line (vi) is ggc -- -- -- -- gcc, a duplex linked by four trimethylene phosphates.

Loop Size, Molecularity, and Calculated and Observed Thermodynamic Parameters. The minimum number of nucleotides in the loop before enthalpic and free energy penalties appears to be four. RNA fragments with three nucleotide loops form multimolecular complexes. Oligonucleotides 1−6 form unimolecular stem-loop structures, as indicated by the concentration-independence of their melting curves (not shown) and by the agreement between calculated and observed enthalpies of melting. For oligo 6 for example, a helical stem joined by a four-residue flexible linker (Table 1) gives an experimental enthalpy of melting (Table 2) that is equivalent to that predicted by MFOLD[51] for the three-base-pair helical stem alone [(^{5}ggc^3′)(^{5}gcc^3′)]. Similarly, the 3′ G linked to a three-residue flexible linker (a four-residue loop consisting of G + three trimethylene phosphates, oligo 5, ggcG -- -- -- gcc) gives an experimental enthalpy of melting (Table 2) that is consistent

with that predicted by MFOLD for the three-base-pair helical stem plus a 3′ dangling G [(^{5}ggcG^3′)(^{5}gcc^3′)]. Another four-residue loop (ggc -- CAAgcc, oligo 4) similarly forms a unimolecular structure, although there is not an unambiguous method to calculate its enthalpy of melting for comparison with the experimentally determined value.

In sum, the base-pairing interactions within the helical stem do not appear to be perturbed by loops of length greater than three residues. By contrast, the concentration-dependence of the melting of oligonucleotides with shorter loops (ggcG -- -- gcc and ggcG -- gcc, oligos 7 and 8) indicates that the global stem-loop structure is not conserved for loops of three residues or less. These two oligonucleotides form multimolecular complexes, presumably because loops of less than four residues are sufficiently short that a significant penalty is incurred upon making the 180° turn required to link the two helical strands.

Table 1. Native and Modified RNA Oligomers

no.	RNA	comment
1	ggcGCAAgcc	Santa Lucia and Turner tetraloop
2	ggc(dG)CAAgcc	G is converted to 2′ deoxy G
3	ggcXCAAgcc	dG is converted to a abasic residue
4[a]	ggc -- CAAgcc	abasic residue is converted to a trimethylene phosphate
5	ggcG -- -- -- gcc	NRA is converted to (trimethylene phosphate)$_3$
6	ggc -- -- -- -- gcc	GNRA is converted to (trimethylene phosphate)$_4$
7[b]	ggcG -- -- gcc	NRA is converted to (trimethylene phosphate)$_2$
8[b]	ggcG -- gcc	NRA is converted to (trimethylene phosphate)$_1$

[a] The trimethylene phosphate retains the C3′, C4′, and C5′ atoms and the phosphate group but omits the C1′, C2′, O2′, and O4′ along with the base (Figure 3D). [b] These modified tetraloops, with shortened loops, give concentration-dependent melting transitions and were not subject to thermodynamic analysis.
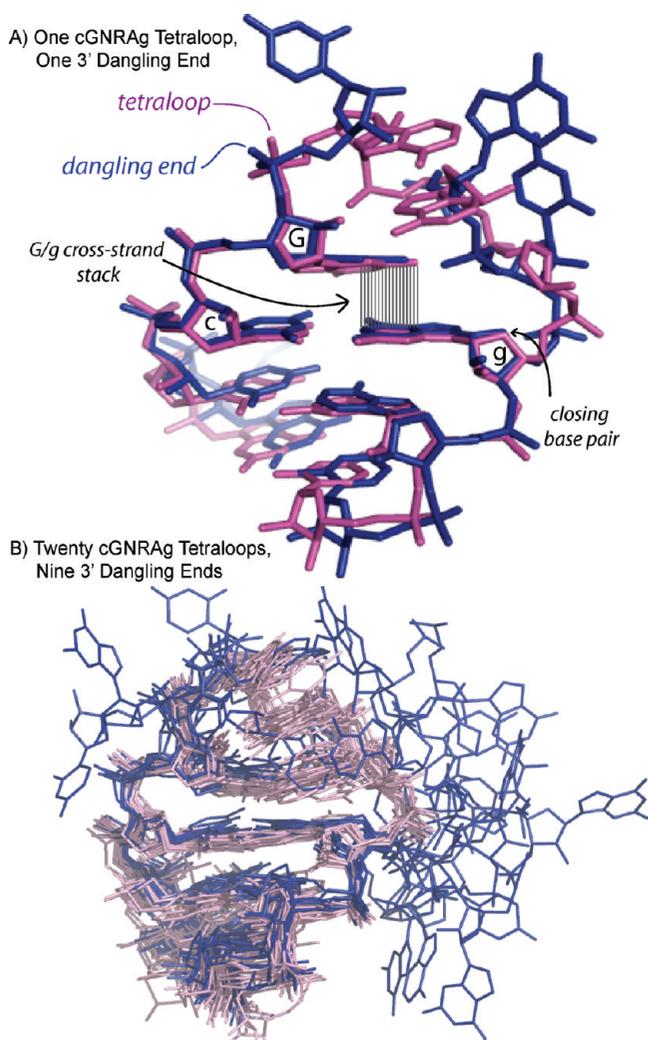

Figure 5. A shared architecture between cGNRAg tetraloops and 3′ dangling ends. (A) A cGNRAg tetraloop (pink, tetraloop 805, 23S rRNA of *H. marismortui*) is superimposed on a 3′ dangling end (blue, dangling end 617, 16S rRNA of *T. thermophilus*). (B) Superimposition of nine 3′ dangling ends (blue) and 20 GNRA tetraloops (pink). Gray vertical lines denote cross-strand stacking. The backbone atoms of the two terminal base-pairs and the stacked G were used for the superimpositions.

Tetraloop Thermodynamic Cycle 1. Both the enthalpy and entropy of folding oligonucleotides 1−6 can be combined to give closed, self-consistent thermodynamic cycles (Figure 7,
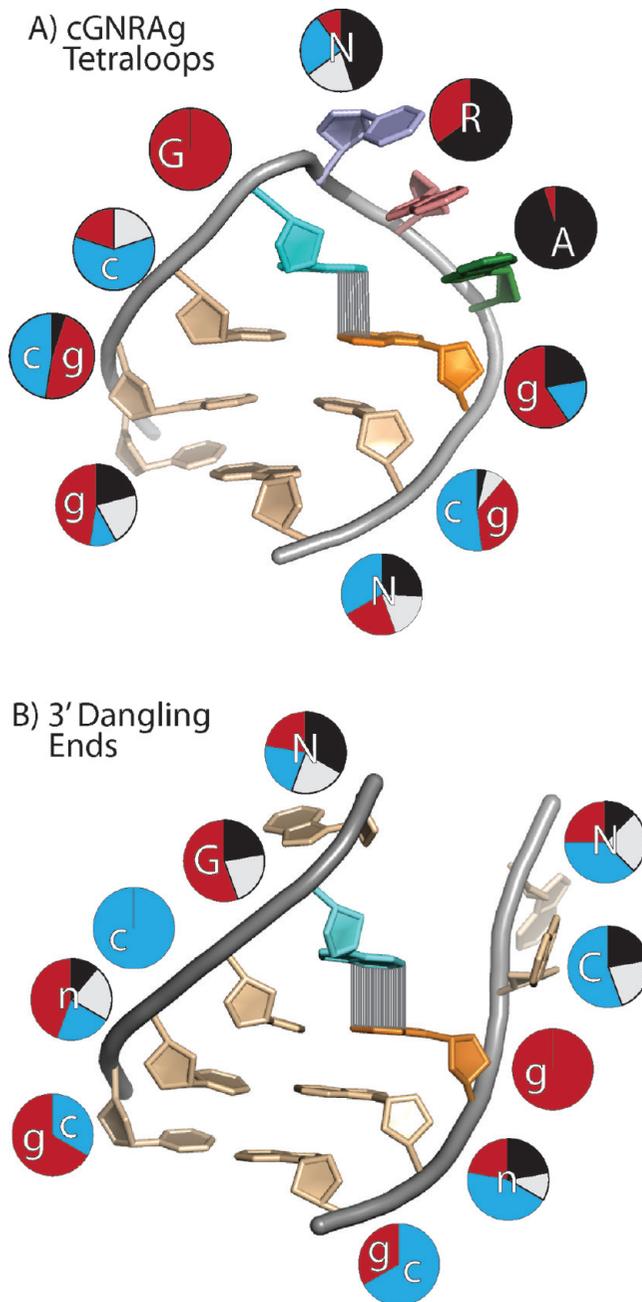

Figure 6. Sequence consensus for cGNRAg tetraloops and 3′ dangling ends. (a) Tetraloops show a preference for cGNRAg. (b) 3′ Dangling ends show a preference for a dangling G and a c·g closing base-pair. The colored wedges in the pie charts represent frequency: red, G; blue, C; black, A, and white, U. The uppercase letters in the pie charts indicate the consensus bases (most frequently observed). N indicates that no particular base is preferred. R indicates that purines are preferred.

Table 2). The data suggest that the ribose causes a decrease in favorable molecular interactions of the folded cGCAAg tetraloop. Removal of the 2′ OH of G of cGCAAg to give c**dG**CAAg (where dG is 2′ deoxyG) costs 1.7 kcal/mol in terms of the enthalpy of folding. Removal of the base from 2′ deoxyG to give c**X** CAAg (where X = 2′ deoxy abasic residue) costs 3.8 kcal/mol. Removal of the deoxyribose ring from the 2′ deoxyX yields c--CAAg (where -- is a trimethylene phosphate, Figure 3D) *adds* 1.3 kcal/mol in enthalpy. The enthalpic thermodynamic cycle 1 is formed by taking the sum of the base, the 2′ OH, and the deoxyribose terms (−4.2 kcal/mol) and subtracting

Table 2. Folding of Tetraloops and 3′ Dangling Ends

RNA	$-\Delta H°$ (kcal mol^{-1})	$-\Delta S°$ (cal mol^{-1}K^{-1})	$\Delta G°_{37}$ (kcal mol^{-1})
(1) gccGCAAgcc[a]	41.6 ± 0.6	121.3 ± 1.8	−4.0 ± 0.2
(2) ggc(dG)CAAgcc[a]	39.9 ± 0.9	117.7 ± 1.8	−3.5 ± 0.2
(3) ggcXCAAgcc[a]	36.1 ± 0.5	106.3 ± 1.3	−3.2 ± 0.2
(4) gcc -- CAAgcc[a]	37.4 ± 0.4	111.0 ± 1.4	2.9 ± 0.1
(5) ggcG -- -- -- gcc[a]	35.0 ± 0.5	103.9 ± 1.9	−2.8 ± 0.03
(6) ggc -- -- -- -- gcc[a]	26.3 ± 1.0	81.2 ± 4.4	−1.1 ± 0.4

contribution of	$-\Delta\Delta H°$	$-\Delta\Delta S°$	$\Delta\Delta G°_{37}$
2′ OH of G[b]	1.7 ± 1.1	3.6 ± 2.5	−0.5 ± 0.2
base of dG[c]	3.8 ± 0.7	11.4 ± 2.2	−0.3 ± 0.2
deoxyribose of G[d]	−1.3 ± 0.7	−4.7 ± 2.2	−0.3 ± 0.2
base plus ribose of G[e]	4.2 ± 1.1	10.3 ± 4.8	−1.1 ± 0.7
CAA[f]	11.1 ± 0.4	29.8 ± 4.6	−1.8 ± 0.5
G[g]	8.7 ± 1.2	22.7 ± 4.8	−1.7 ± 0.1
GCAA loop[h]	15.3 ± 1.2	40.1 ± 4.8	−2.9 ± 0.5
3′ dangling end (G)[i]	1.7 ± 1.1	3.6 ± 2.6	−0.5 ± 0.2

[a] Table 1 and Figures 3 and 4 give oligonucleotide sequences and modifications. In the top panel, the uncertainty estimates are 95% confidence limits from three successive measurements. In the bottom panel, the uncertainty estimates are root of the sum of the squares of the uncertainties of the values used for the difference.[64] [b] Difference between rows 1 and 2. [c] Difference between rows 2 and 3. [d] Difference between rows 3 and 4. [e] Difference between rows 1 and 4. [f] Difference between rows 6 and 4. [g] Difference between rows 6 and 5. [h] Difference between rows 6 and 1. [i] These values are derived from MFOLD. They correlate well with experimental values reported by Turner.[54]

the combined contribution of the base and the ribose of G (−4.2 kcal/mol), which is separately determined by taking the difference between the enthalpies of folding of native cGCAAg and c--CAAg.

One can construct a similar thermodynamic cycle with the entropy of folding (Figure 7). Removal of the 2′ OH of G in cGCAAg adds 3.6 cal/(mol K) to the entropy of folding. Removal of the base from the 2′ deoxyG adds 11.5 cal/(mol K). Removal of the ribose ring from 2′ deoxyX costs 4.8 cal/(mol K). The sum of these three terms is −10.3 cal/(mol K), which is equivalent to the combined contribution of the base and the ribose.

These thermodynamic cycles, which are consistent with a model in which the global structure of the stem loop is generally conserved over the modifications, allow parsing of the molecular interactions. Thermodynamic cycle 1 suggests that removing the sugar relieves torsional restraints and allows better molecular interactions. The simplest way to explain this observation is that a cGNRAg tetraloop is under a sort of tension, where torsional restraints imposed by the ribose backbone prevent optimum stacking and/or hydrogen bonding within the loop region. Three-dimensional data-mining results (below) support this model.

Thermodynamic cycle 1 further suggests that the cross-strand G/g stack, involving the first nucleotide of the cGCAAg tetraloop and the opposing "g" of the closing base-pair (Figure 1), makes the largest single contribution to the enthalpy of folding (Figure 7). This inference is consistent with results of Bevilacqua,[52] who demonstrated the importance of the closing base-pair to cGCAAg folding, and the results of Santa Lucia and Turner,[47] who concluded that the hydrogen-bonding interactions of G make secondary contributions.

(51) Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406.
(52) Blose, J. M.; Proctor, D. J.; Veeraraghavan, N.; Misra, V. K.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2009**, *131*, 8474.
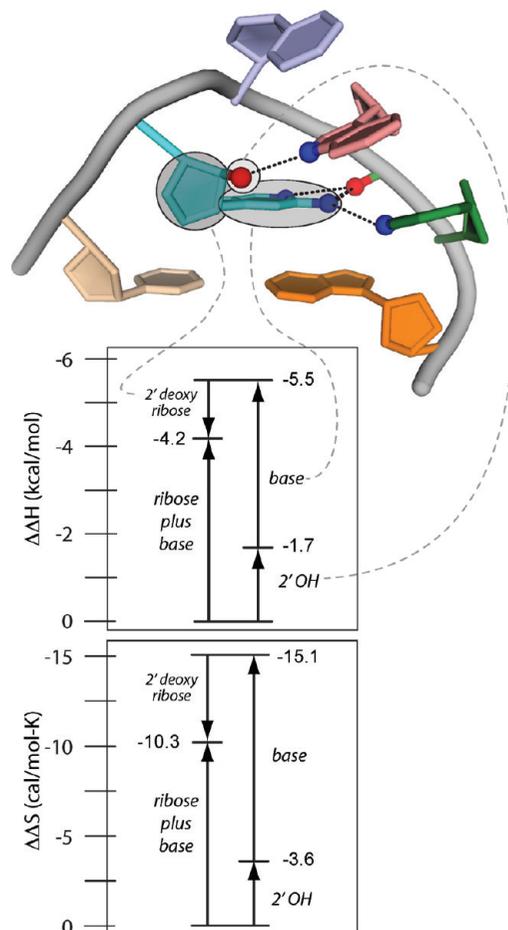
Figure 7. Thermodynamic cycle 1: stepwise excision of G from a cGNRAg tetraloop. These cycles show the effects of removal of various parts of G on the standard enthalpy and entropy of folding. The modifications form a thermodynamic cycle where the sum of the effects of removing the base, then the 2′ hydroxyl, then the sugar, are equivalent to the effect of performing all three steps simultaneously.

Enthalpy−entropy compensation is observed at each step of both cycle 1 and cycle 2 (below). When $\Delta\Delta H_{fold}$ is negative, $\Delta\Delta S_{fold}$ is also negative. For example, removal of the deoxyribose of G decreases $\Delta\Delta H_{fold}$ and $\Delta\Delta S_{fold}$.

Tetraloop Thermodynamic Cycle 2. A series of modified cGCAAg hairpins was constructed in which riboses and bases of the loop residues were removed by conversion of nucleotides to flexible trimethylene phosphates (Figure 3D). The tetraloop ggcG -- -- -- gcc, a 3′ dangling end joined by a flexible linker, was used to estimate the contribution of the bases and sugars of the NRA trinucleotide. The tetraloop ggc -- CAAgcc was used to estimate the contribution of the bases and sugars of G. The tetraloop ggc -- -- -- -- gcc, a blunt helix joined by a flexible linker, was used to estimate the contribution of the bases and sugars of the four loop nucleosides (Figure 8). Removal of the CAA trinucleotide from cGCAAg to form cG -- -- -- g costs 11.0 kcal/mol in the enthalpy of folding. Removal of the G from cGCAAg to form c -- CAAg costs 8.7 kcal/mol. By contrast, removal of all four riboses and bases of the loop nucleotides costs 15.3 kcal. The resulting c -- -- -- -- g structure is a blunt end helix, linked by a flexible tether. As noted above, the tether does not appear to perturb the structure, as indicated by the observed enthalpy of folding, which is equivalent to that predicted by MFOLD for the unlinked helix.
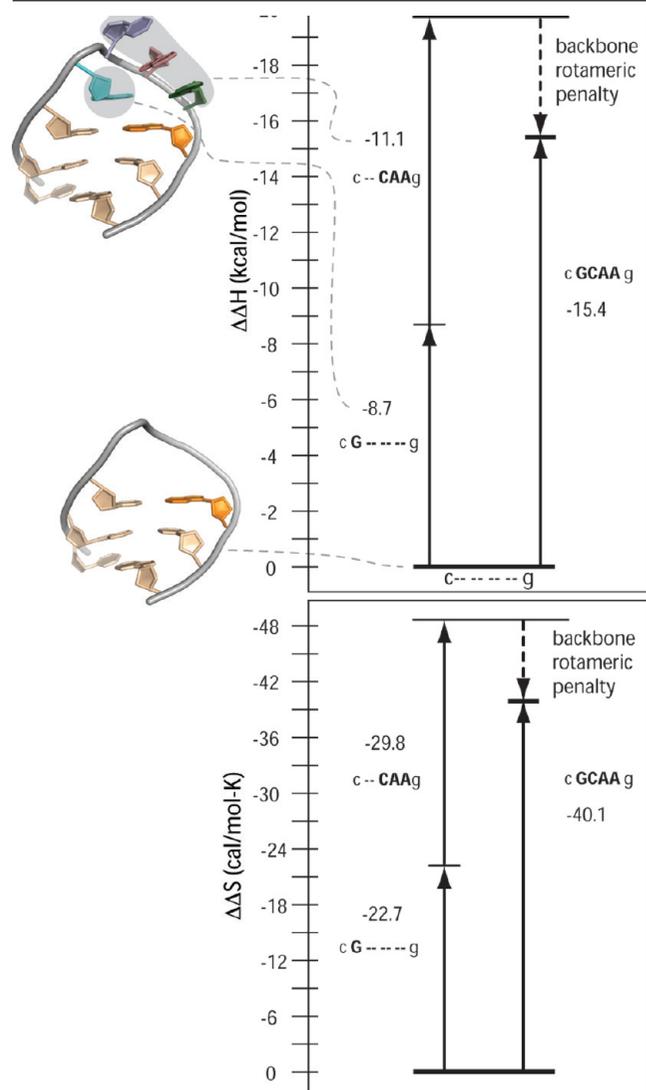
Figure 8. Thermodynamic cycle 2: building a cGNRAg tetraloop. These cycles show the effects of modification of nucleotides within the loop on the standard enthalpy and entropy of folding. A blunt helical terminus, linked by a flexible tether, is converted, in a stepwise fashion, into a tetraloop. The sum of the effects of converting trimethyl phosphates to G and NRA is not equivalent to simultaneously converting all four trimethyl phosphates to GNRA. Including the effect of torsional stress closes the thermodynamic cycle.

Closing cycle 2 requires 4.4 kcal/mol in enthalpy (Figure 8). The "real" cGCAAg tetraloop is 4.4 kcal/mol less favorable in enthalpy than the "theoretical" tetraloop made by independent addition of G and CAA to c -- -- -- g. Thus it appears that the net effect of the molecular interactions achieved by the independent addition of components is not achieved in a real tetraloop. This observation is consistent with the results of cycle 1 (above), where addition of a ribose to the loop was similarly inferred to decrease the extent of molecular interactions. The simplest model that is consistent with these observations is one in which torsional restraints imposed by the riboses of the loop oppose optimum stacking and hydrogen bonding. In fact, analysis of three-dimensional structures (below) suggests that neither bond rotamers nor molecular interactions are optimized in cGNRAg tetraloops.

cGNRAg Tetraloop Structure. Using an objective set of structural criteria[36] and restricting the first nucleotide of the loop to G, we have identified 20 unique cGNRAg tetraloops in the ribosomal structural database. Tetraloops with bulges or mismatches in their stems, except g•u wobbles, are excluded. The conformation, molecular interactions, and sequences of these cGNRAg tetraloops are consistent with previous GNRA tetraloops observed in a group I ribozyme[28] and other RNA structures.[53] Our 20 cGNRAg tetraloops overlap substantially in membership, sequence, conformation, and molecular interactions with a set of cGNRAg tetraloops defined by Correll and co-workers.[37]

Superposition of the 20 GNRA tetraloops reveals conserved backbone conformation (Figure 1B). The backbone atoms of the GNRA loops plus the three base-paired nucleotides of the stems superimpose with an average rmsd of atomic positions of 0.8 Å. If the three NRA nucleotides of the loops are omitted (to facilitate comparison to 3′ dangling ends, see below), the average rmsd of atomic positions decreases slightly to 0.7 Å.

cGNRAg Tetraloop Molecular Interactions. Molecular interactions within these 20 cGNRAg tetraloops are highly conserved.[36,37]

Hydrogen Bonds. The 2′-hydroxyl group of G forms a hydrogen bond with N7 of R. The 2-amino group (and/or N1) of G forms a hydrogen bond with the O2P backbone atom of R. The N7 of A forms a hydrogen bond with the 2-amino group of G. This collection of hydrogen-bonding interactions is nearly identical to those identified by Quigley and Rich in the U-turn in yeast[Phe] tRNA.[39]

Stacking. In cGNRAg tetraloops, G stacks on a closing c•g base-pair (Figures 1A and 5A). Stacking is primarily cross-stranded, between G and g. The average area of overlap of G on g is $2.3 \pm 0.8_{95\%CL}$ Å^2. Same-strand stacking of G on c is less pronounced, 1.3 ± 0.8 Å^2. The other three bases of the loop, namely NRA, generally form a continuous, single-stranded stack, although in some cases N is unstacked from R.

The A is unstacked on the 3′ side and does not interact significantly with the adjacent closing c•g base-pair. The average area of overlap of A on g is only 0.3 ± 0.3 Å^2. In fact, only one cGNRAg tetraloop (1013 of the 16S rRNA, with an anomalous u•g closing base-pair) shows significant A/g overlap (3.3 Å^2). Elimination of this anomalous tetraloop reduces the average A/g overlap area to only 0.2 ± 0.1 Å^2. A is poised to form a sheared base-pair with the opposing G. The potential hydrogen bonds are from G(N2) to A(N7) and from G(N3) to A(N6). However, as noted previously,[36,37] these two bases do not pair. A lone hydrogen bond links the A and G [G(N2) to A(N7)]. The average (G)N3-to-(A)N6 distance is 4.7 Å, which is well beyond the cutoff for a reasonable hydrogen bond.

cGNRAg Tetraloop Sequence. Except for the G at the first position of the loop, sequence was not used as a criterion and is an independent variable here. The closing base-pair of cGNRAg tetraloops shows some variability in base identity, with g > a > c on the 3′ side and c > g, u on the 5′ side of the tetraloop. The three-dimensional structures indicate that the tetraloop fold tolerates this variability; the cross-strand stack is maintained whether the closing base on the 3′ side is g, a, or c.

3′ Dangling End Structure, Molecular Interactions, and Sequence. A 3′ dangling end is an unpaired nucleotide stacked on, and attached to, a helical terminus (Figure 2). We have

(53) Batey, R. T.; Sagar, M. B.; Doudna, J. A. *J. Mol. Biol.* **2001**, *307*, 229.

(54) Serra, M. J.; Turner, D. H. In *Energetics of Biological Macromolecules*; Abelson, J. N., Simon, M. I., Johnson, M. L., Ackers, G. K., Eds.; Methods in Enzymology *259*; Acedemic Press: San Diego, CA, 1995; p 242.

Table 3. cGNRAg Tetraloop and 3′ Dangling End Consensus Sequences and Molecular Interactions

	3′ dangling ends[a]	cGNRAg tetraloops[b]	dangling ends and tetraloops[c]
consensus sequence[d]	(^{5}ggcG$^{3'}$)(^{5}gcc$^{3'}$)	^{5}ggcGNRAggc$^{3'}$	(^{5}ggcG$^{3'}$)(^{5}gcc$^{3'}$)
overlap area[e]	1.7 ± 1.1 Å^2	2.3 ± 0.8 Å^2	2.1 ± 0.9 Å^2
backbone rmsd[f]	0.8 Å	0.7 Å	0.9 Å

[a] Nine 3′ dangling ends, each with a one-nucleotide 3′ stack. [b] Twenty cGNRAg tetraloops. [c] Twenty-nine structures (nine 3′ dangling ends plus 20 cGNRAg tetraloops). [d] The sequence is most conserved at the closing base-pair and 3′ unpaired nucleotide. [e] The overlap areas were calculated using the closing base-pair and the 3′ unpaired nucleotide. [f] The rmsd of atomic positions was calculated after superimposition using the backbone atoms of the 3′ unpaired nucleotide and the two adjacent base-pairs.

extracted structures of nine 3′ dangling ends from the three-dimensional database. For simplicity, we have excluded dangling ends with multiple stacked bases.[17,45] The dangling ends show conserved conformation (Table 2), with an average rmsd of backbone atomic positions, after superimposition of the last two base-pairs of the helix and the dangling 3′ nucleotide, of 0.8 Å.

Stacking of dangling ends at helical junctions of RNA has previously been geometrically defined and quantified.[17,40,46] Dangling ends show cross-strand stacking between the dangling nucleotide and the cross-strand base (e.g., G/g stacking, Figure 2), with an average area of overlap of 1.7 ± 1.1 Å^2. Although sequence was not used as a search criterion, all nine 3′ dangling ends close with c·g base-pairs (Figure 6B). The identity of the dangling base is G > U = A.

Comparison of Tetraloops and 3′ Dangling Ends. 3′ Dangling ends are analogues of tetraloops in that both contain a helical terminus and a stacked unpaired base, followed by a break in the single-strand stack (between G and N, Figures 1A and 2A), although in cGNRAg tetraloops the G forms a single hydrogen bond with the opposing A.[36,37] Therefore, it appears that tetraloops and 3′ dangling ends share a common substructure. The backbone atoms of a 3′ dangling end superimpose well on the corresponding atoms of a cGNRAg tetraloop (Figure 5A, Table 3). For 20 cGNRAg tetraloops and nine 3′ dangling ends, the backbone atoms of the two base-paired nucleotides and the unpaired G superimpose with an average rmsd of 0.86 Å (Figure 5B). In addition, the consensus sequences for cGNRAg tetraloops and 3′ dangling ends are the same (Figure 6, Table 3). The consensus closing base-pair for both cGNRAg tetraloops and 3′ dangling ends is c·g. The consensus 3′ stacked base is G for both tetraloops (by definition) and 3′ dangling ends.

Cross-strand stacking of G on g is a conserved feature of cGNRAg tetraloops (Figure 1A). An analogous interaction is conserved in 3′ dangling ends (Figure 2A). Structural data-mining indicates that the average area of cross-strand overlap is 2.3 ± 0.8 Å^2 for cGNRAg tetraloops and 1.7 ± 1.1 Å^2 for 3′ dangling ends.

The stacking of the G on the c·g base-pair appears to be an important contributor to stability of cGNRAg tetraloops. The analogous interaction contributes even more to the stability of 3′ dangling ends. G contributes 8.7 ± 0.5 kcal/mol to the enthalpy of folding of both the ggcG -- -- -- gcc modified tetraloop (determined here by comparison to ggc -- -- -- gcc) and to a 3′ dangling G on a c·g base-pair, as determined by Serra and Turner.[54] Stacking is the apparent origin of this favorable enthalpy. The similarity in the behavior of these bimolecular (3′ dangling end) and unimolecular (modified

tetraloop) structures suggests that ggc -- -- -- -- gcc is a useful unimolecular model for double-stranded RNA. By contrast, the contribution of G to the enthalpy of folding of an unmodified cGNRAg tetraloop is significantly less than to that of a 3′ dangling end or ggcG -- -- -- gcc. Addition of G to ggc -- CAAgcc (to form ggcGCAAgcc) adds only 4.2 kcal to the enthalpy of folding. If one subtracts the unfavorable deoxyribose contribution, the favorable enthalpy contributed by the base plus the 2′OH is only 5.5 kcal/mol. The molecular basis for this attenuation of the enthalpy of interaction of the G of a cGNRAg tetraloop merits further investigation.

Discussion

Santa Lucia and Turner used chemical modifications to determine the thermodynamic contributions of hydrogen bonding to cGNRAg tetraloop folding.[47] Bevilacqua extended this approach, constructing thermodynamic cycles for folding of both DNA[55] and RNA[56] hairpins, demonstrating, for example, that folding of at least some RNA hairpins is less cooperative than that of the corresponding DNA hairpins.

Here we use chemical modifications (Figure 3) to determine the contributions of stacking and backbone restraints to cGNRAg folding. Starting with the cGNRAg tetraloop, we omitted first the 2′-hydroxyl group (Figure 3B), then the base (Figure 3C), then the ribose (Figure 3D) from the G. The ribose was excised by substituting a trimethylene phosphate (denoted --, also see Bevilacqua). This substitution conserves the length and number of bonds of the backbone (Figure 3D) but relieves rotameric restraints imposed by the ribose. Similarly, we excised other bases and riboses of the loop of cGNRAg, forming a series of structures that are intermediate between tetraloops and 3′ dangling ends. Both the enthalpy and entropy of folding of oligonucleotides 1−6 can be combined to give closed, self-consistent thermodynamic cycles (Figures 7 and 8, Table 2). These cycles allow us to parse the contributions of various functional groups and nucleotides to folding thermodynamics.

In addition, we have performed sequence and conformational analyses of tetraloops and 3′ dangling ends obtained from the three-dimensional database. The combined thermodynamic and structural results suggest that a cross-strand stack (from G to g, Figure 1) is a primary stabilizing element of cGNRAg tetraloops. An analogous stacking interaction is a primary stabilizing element of 3′ dangling ends (from G to g, Figure 2). We observe that the helical stem plus the cross-strand G/g stack of a cGNRAg tetraloop averaged over many structures is superimposable on the analogous segment of a 3′ dangling end (Figure 5, Table 1). The combined results suggest cGNRAg tetraloops and 3′ dangling ends share a common stacked G/g substructure. The fundamental strategy for terminating a helix is the same for cGNRAg tetraloops and 3′ dangling ends.

Tension between Bond Rotamers and Molecular Interactions. Here we propose that, within a cGNRAg tetraloop, tension between backbone conformation and molecular interactions prevents the folded tetraloop from achieving a global thermodynamic minimum.

The energetic landscape of RNA torsion angles (Figure 3E) restricts the available conformational space.[57,58] RNA backbone torsion angles, except ζ, fall into well-defined frequency envelopes,[58] reflecting the rotameric free energy landscape.

(55) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2004**, *126*, 9570.
(56) Moody, E. M.; Feerrar, J. C.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 7992.

Ribose prefers the 3′ endo conformation,[59,60] although the 2′ endo is allowed at low frequency. In a cGNRAg tetraloop, several torsion angles within the loop and stem are pushed to minor frequency envelopes or outside of frequency envelopes altogether, reflecting a switch to less stable bond rotamers.

Torsion angle β of residue N is invariably shifted from the primary gauche(+) envelope to gauche(−). A simple Boltzmann calculation suggests a free energy penalty for this conversion of around 1.5 kcal/mol. Other rotameric switches are variable but seem to focus on ζ of A and α, β, and γ of g. The less favorable bond rotamers, except angle β of residue N, are focused primarily on regions of the cGNRAg tetraloop where stacking is disrupted (on the 5′ side of N and on the 3′ side of A). These unfavorable backbone rotamers appear to reflect significant energetic penalties.

The bases in RNA structures have a well-known tendency to stack and pair. For example, in tRNA the overwhelming majority of bases are both stacked and paired.[61] Thus it is striking that, in a cGNRAg tetraloop, A is neither paired with the opposing G nor stacked on the 3′ side (Figure 1A),[36,37] and the adjacent bond torsions are suboptimal. A model in which favorable bond torsions are opposed by favorable molecular interactions (stacking and pairing) is consistent with the observation that release of torsional restraints upon conversion of one or more loop riboses to flexible trimethylene phosphate(s) (Figure 3D) contributes favorably to the enthalpy of folding. This effect presumably results from improved molecular interactions upon release of torsional restraints. The most obvious possibility for increasing molecular interactions is a repositioning of A.

Stacking Interactions: cGNRAg Tetraloops and 3′ Dangling Ends. Both cGNRAg tetraloops and 3′ dangling ends contain cross-strand purine/purine stacks (Figures 1 and 2, Table 3) in which G (the first loop nucleotide of cGNRAg and the unpaired nucleotide of a 3′ dangling end) stacks predominantly on the opposing member of a closing base-pair. It is known that, for 3′ dangling ends, this cross-strand stack is an important component of stability.[46,62,63] It was previously suggested that insertion of trimethylene phosphate residues between c and G of cGNRAg (not a substitution as described here) has a destabilizing effect,[56] which was attributed to the disruption of base stacking. The combined evidence suggests that cGNRAg

tetraloops and 3′ dangling ends share a simple stacking substructure with conserved geometry and sequence.

The preferred sequence is similar for cGNRAg tetraloops and 3′ dangling ends in the region of the 3′ cross-strand stack (Figure 6). The consensus closing base-pair is c·g for both 3′ dangling ends and cGNRAg tetraloops. The consensus is G for both the first nucleotide of GNRA (by definition) and the 3′ dangling nucleotide of the dangling end. In addition, the conformation of the RNA is similar for cGNRAg tetraloops and 3′ dangling ends (Figure 5).

Bevilacqua's results indicate the importance of stacking on the closing base-pair in hairpins in general. Our results here support that work and further suggest that a simple stacking substructure with conserved geometry and sequence, and a specific thermodynamic fingerprint, forms a basic scaffold for both cGNRAg tetraloops and 3′ dangling ends. For reasons that are unclear and are not evident from the extent or geometry of stacking, the enthalpic contribution of this stacking interaction to stability is less for cGNRAg tetraloops than for 3′ dangling ends.

Hydrogen-Bonding Interactions. The importance of stacking to stability of 3′ dangling ends is unambiguous because the dangling base has no hydrogen-bonding partner. The importance of stacking to cGNRAg tetraloop stability is underscored by considering the results here, along with the work of Santa Lucia and Turner, who determined that, individually, the loop hydrogen bonds contribute less to stability than Watson−Crick-type hydrogen bonds contribute to duplex stability.[47] They used modified bases such as inosine to determine the $\Delta G/\Delta H/\Delta S$ increments for each H-bond in a GNRA tetraloop.[47] Turner, in this classical work on context-dependent effects of hydrogen bonding, states that "ΔG increments for hydrogen bonds within the GCAA hairpin contribute relatively little to thermodynamic stability".

cGNRAg Consensus Structure. In a cGNRAg tetraloop, the A is unstacked on the 3′ side and is not paired with the opposing G. This destacking and unpairing of A was originally noted by Correll and co-workers[37] and is a tightly conserved feature of X-ray structures of cGNRAg tetraloops.[36] This unstacking and unpairing is now seen to arise from rotameric restraints of the loop's backbone. The results presented here indicate that, when the rotameric restraints of the ribose are lifted, the enthalpic contribution to stability increases and is greater than predicted from a thermodynamic cycle (Figure 8, Table 2). The individual thermodynamic contributions of the G and the NRA stack are not additive and do not represent the net thermodynamic contribution of those elements to the folded cGNRAg tetraloop. The observed noncyclic thermodynamic effect suggests that folded cGNRAg tetraloops are destabilized by rotameric restraints and by unstacking of A.

JA104387K

(57) Murray, L. J.; Arendall, W. B., III; Richardson, D. C.; Richardson, J. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13904.
(58) Hershkovitz, E.; Tannenbaum, E.; Howerton, S. B.; Sheth, A.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2003**, *31*, 6249.
(59) Rich, A. *Nat. Struct. Mol. Biol.* **2003**, *10*, 247.
(60) Sundaralingam, M. *Biopolymers* **1969**, *7*, 821.
(61) Kim, S. H.; Sussman, J. L.; Suddath, F. L.; Quigley, G. J.; McPherson, A.; Wang, A. H.; Seeman, N. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 4970.
(62) Freier, S. M.; Burger, B. J.; Alkema, D.; Neilson, T.; Turner, D. H. *Biochemistry* **1983**, *22*, 6198.
(63) Freier, S. M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H. *Biochemistry* **1986**, *25*, 3214.
(64) Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences*; McGraw-Hill: New York, 1969.