J|A|C|S
ARTICLES
Published on Web 07/20/2004

Structural and Energetic Consequences of Expanding a Highly Cooperative Stable DNA Hairpin Loop

Ellen M. Moody and Philip C. Bevilacqua*

Contribution from the Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

Received March 22, 2004; E-mail: pcb@chem.psu.edu

Abstract: Many hairpin loops are expanded versions of smaller, stable ones. Herein we investigate the extent to which the energetics and structure of d(cGNAg) hairpin loops will tolerate sequence variation. Changing the closing base pair from CG to GC was found to completely eliminate loop−loop interactions; in contrast, expanding the loop at the 3′-end resulted in similar energetics and nonadditivity parameters as the parent loop, suggesting that loop−loop interactions remain intact and highly coupled upon expansion. Together, these data suggest that the CG closing base pair forms an essential platform upon which a stable d(GNA) hairpin loop can fold and that this loop can undergo 3′-expansion with little effect to its structure or energetics.

Introduction

Hairpins are the most common secondary structural elements in RNA, playing important roles in folding and interactions with proteins.[1,2] Double stranded DNA too can form hairpins, for example, in cruciforms.[3] Biological roles of these structures have been described and include regulating replication and transcription.[1,4−7]

Stable RNA triloop and tetraloop hairpin sequences have been reported,[8−11] and many structures are available.[12−16] As in RNA, the stability of DNA hairpins depends on both the sequence composition of the loop and the closing base pair (cpb).[8,17−21] Previously, we carried out experiments directed toward identifying unusually stable DNA hairpin loops using combinatorial selections and temperature gradient gel electrophoresis (TGGE).[21] One of the motifs identified was d(cGNABg) (where the closing base pair is in lower case, "N" is A, C, G, or T, and "B" is C, G, or T), and this motif appears to have biological relevance.[22] The d(cGNABg) motif is thought to be an expansion of the stable triloop motif, d(cGNAg), which contains a sheared GA base pair.[23,24] Indeed, studies on the d(cGNABg) motif revealed a large destabilization when the first and third positions of the loop were changed.[21,24]

It has also been observed in both DNA and RNA that for certain hairpin loop sequences a CG closing base pair provides much greater thermodynamic stability ($\Delta\Delta G_{37}^{\circ} = 2-3$ kcal/mol) than expected from standard Watson−Crick base pairing alone.[8,10,11,19,21] A large thermodynamic penalty is incurred for three-carbon spacer (C3) insertion before the 5′-end of both d(cGCAg) and d(cGCACg) hairpin loops, with much smaller penalties for insertion throughout the rest of the loop or in hairpins with other closing base pairs.[24] These data support a stabilizing interaction between G1 of the loop and the CG closing base pair as part of the basis for stability.

Recently, the cooperativity of d(cGCAg) loop folding was investigated by the use of double mutant cycles, and all interactions were found to be nonadditive and interdependent. These findings were consistent with loop−loop and loop−closing base pair interactions forming in a highly cooperative manner.[25] When the double mutant cycles were repeated in the absence of the other interaction, nonaddivity was significantly reduced,[25] consistent with indirect coupling and a concerted

(1) Varani, G. *Annu. Rev. Biophys. Biophys. Chem.* **1995**, *24*, 379−404.
(2) Tinoco, I., Jr.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271−281.
(3) Lilley, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6468−6472.
(4) Choi, K. H.; Choi, B. S. *Biochim. Biophys. Acta* **1994**, *1217*, 341−344.
(5) Glucksmann-Kuis, M. A.; Dai, X.; Markiewicz, P.; Rothman-Denes, L. B. *Cell* **1996**, *84*, 147−154.
(6) Bzymek, M.; Lovett, S. T. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8319−8325.
(7) Gellert, M. *Annu. Rev. Biochem.* **2002**, *71*, 101−132.
(8) Antao, V. P.; Lai, S. Y.; Tinoco, I., Jr. *Nucleic Acids Res.* **1991**, *19*, 5901−5905.
(9) Antao, V. P.; Tinoco, I., Jr. *Nucleic Acids Res.* **1992**, *20*, 819−824.
(10) Shu, Z.; Bevilacqua, P. C. *Biochemistry* **1999**, *38*, 15369−15379.
(11) Proctor, D. J.; Schaak, J. E.; Bevilacqua, J. M.; Falzone, C. J.; Bevilacqua, P. C. *Biochemistry* **2002**, *41*, 12062−12075.
(12) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680−682.
(13) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191−194.
(14) Allain, F. H. T.; Varani, G. *Nucleic Acids Res.* **1995**, *23*, 341−350.
(15) Chou, S. H.; Chin, K. H.; Wang, A. H. *Nucleic Acids Res.* **2003**, *31*, 2461−2474.
(16) Du, Z.; Yu, J.; Andino, R.; James, T. L. *Biochemistry* **2003**, *42*, 4373−4383.
(17) Senior, M. M.; Jones, R. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6242−6246.
(18) Paner, T. M.; Amaratunga, M.; Doktycz, M. J.; Benight, A. S. *Biopolymers* **1990**, *29*, 1715−1734.
(19) Sandusky, P.; Wooten, E. W.; Kurochkin, A. V.; Kavanaugh, T.; Mandecki, W.; Zuiderweg, E. R. *Nucleic Acids Res.* **1995**, *23*, 4717−4725.
(20) Vallone, P. M.; Paner, T. M.; Hilario, J.; Lane, M. J.; Faldasz, B. D.; Benight, A. S. *Biopolymers* **1999**, *50*, 425−442.
(21) Nakano, M.; Moody, E. M.; Liang, J.; Bevilacqua, P. C. *Biochemistry* **2002**, *41*, 14281−14292.
(22) Habig, J. W.; Loeb, D. D. *J. Virol.* **2002**, *76*, 980−989.
(23) Hirao, I.; Kawai, G.; Yoshizawa, S.; Nishimura, Y.; Ishido, Y.; Watanabe, K.; Miura, K. *Nucleic Acids Res.* **1994**, *22*, 576−582.
(24) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 2032−2033.
(25) Moody, E. M.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 16285−16293.

10.1021/ja048368+ CCC: $27.50

folding of the loop.[26] Since many DNA and RNA loops appear to be expansions of stable tri- and tetraloops,[21,27,28] an important question is whether expansion of the loop affects the energetics and cooperativity of folding. Herein the nature of loop−loop and loop−closing base pair interactions in the 3′-expanded d(cGNABg) loops is probed by examining the stability and NMR spectra of nucleotide analogue substituted oligonucleotides. We find that 3′-expanded loops, like their parent loops, fold in a highly cooperative manner.

Materials and Methods

Preparation of DNA. Synthesis, deblocking, desalting, and dialysis were as described.[25] DNA oligonucleotides were either from IDT, the Nucleic Acids Facility at the Pennsylvania State University, or the HHMI-Keck Facility at Yale University using reagents from Glen Research. Electrospray ionization mass spectrometry and gel electrophoresis was used to confirm the molecular weight and sizes of representative oligonucleotides. All DNA had the general sequence 5′-d(ggaX$L_1L_2L_3L_4$X′tcc), where X and X′ are complementary nucleotides forming the closing base pair and "L" indicates a loop nucleotide. Since all oligonucleotides have the same three beginning (5′gga) and ending (tcc3′) nucleotides, only the loop and closing base pair are provided in the text.

UV Melting Experiments. UV absorbance melting profiles were obtained in $P_{10}E_{0.1}$ [= 10 mM sodium phosphate and 0.1 mM Na_2-EDTA (EDTA, ethylenediaminetetraacetic acid) (pH 7.0)] at 260 and 280 nm and analyzed using nonlinear least-squares fitting with Kaleidagraph v3.5 (Synergy Software) as described.[21,29] The equations used in the fit assumed linear baselines and temperature-independent enthalpy and entropy. Direct outputs from the fits were ΔH and T_M (melting temperature), from which ΔS and ΔG could be calculated using standard thermodynamic relationships. Melts were found to be largely independent of strand concentration, consistent with the hairpin conformation. The only exception was d(cGCICg) (I, inosine), which has a self-complementary loop; in this case the parameters differ between 14 and 75 μM but were similar between 4 and 14 μM. Therefore the lowest strand concentration for which acceptable data could be obtained (4 μM) was used.

NMR Spectroscopy. DNA oligonucleotides for NMR spectroscopy were synthesized, deblocked, and desalted by the manufacturer (IDT). Oligonucleotides were dialyzed as described.[11] DNA concentrations ranged from 240 to 570 μM, and the DNA was renatured prior to the start of each experiment by heating to 90 °C for 3 min and cooling on benchtop for 10 min. NMR data were collected on Bruker AMX2-500 and DRX-400 spectrometers using a 5 mm broadband probe and on a DRX-600 spectrometer using a fixed-frequency triple-resonance (^{1}H, ^{13}C, and ^{15}N) probe as described.[11]

Analysis of Double Mutant Cycles. The additivity of ΔG_{37}° values for double mutant cycles was analyzed similar to previously described,[25,26,30,31] with the wild-type (M_{00}), two single mutants (M_{10} and M_{01}), and the double mutant (M_{11}) comprising the corners of a box (see Figure 5A as an example). The free energy changes associated with mutations A and B are ΔG_A and ΔG_B, respectively, and the change associated with both mutations is ΔG_{AB}. The free energy change associated with mutation A in the presence of B is $^B\Delta G_A$, and that associated with mutation B in the presence of A is $^A\Delta G_B$, which are given along the other two edges of the box. The magnitude of the nonadditive effect between mutations A and B, δ_{AB}, is a coupling free energy that is derived from the notion that $^B\Delta G_A$ is ΔG_A with a "correction" of δ_{AB}: $^B\Delta G_A = \Delta G_A + \delta_{AB}$. δ_{AB} was calculated according to the following two equivalent equations,

$$\delta_{AB} = \Delta G_{37}°(M_{00}) + \Delta G_{37}°(M_{11}) - [\Delta G_{37}°(M_{10}) + \Delta G_{37}°(M_{01})] \quad (1a)$$

$$\delta_{AB} = \Delta G_{AB} - [\Delta G_A + \Delta G_B] \quad (1b)$$

Equation 1b emphasizes the point that δ_{AB} is the nonadditivity of the two single mutations. A negative value for δ_{AB} reflects deletion of the first interaction weakening the second interaction and signifies positive coupling between the functional groups.[26] A positive value for δ_{AB}, on the other hand, reflects deletion of the first interaction strengthening the second interaction and signifies negative coupling. A δ_{AB} of 0 supports no coupling. A double mutant is considered "completely nonadditive" if δ_{AB} equals the smaller of $-\Delta G_A$ or $-\Delta G_B$, which causes $^B\Delta G_A$ or $^A\Delta G_B$ to approach zero; note that this definition is more general than that previously used,[25] since functional groups A and B may have different maximal contributions to stability if they affect different interactions.

Certain double mutant cycles were repeated in the background of a third change elsewhere in the loop to probe whether coupling is direct or indirect.[26] The two equations for this case are

$$^C\delta_{AB} = \Delta G_{37}°(M_{001}) + \Delta G_{37}°(M_{111}) - [\Delta G_{37}°(M_{101}) + \Delta G_{37}°(M_{011})] \quad (2a)$$

$$^C\delta_{AB} = {}^C\Delta G_{AB} - [{}^C\Delta G_A + {}^C\Delta G_B] \quad (2b)$$

where a superscript C denotes the presence of a mutational configuration at site C. If the coupling between two sites, A and B, is direct, then δ_{AB} should equal $^C\delta_{AB}$; otherwise the coupling is indirect and requires a more concerted change.[26] Errors were propagated as described.[25]

Results and Discussion

Thermodynamic Effects of Nucleotide Analogue Substitutions. Previously, we carried out selection studies that led to the identification of the DNA tetraloop hairpin motif, d(cGNABg), as being unusually stable.[21] This motif was compared to the stable triloop, d(cGNAg), for which an NMR structure showed a sheared GA between positions 1 and 3 of the loop.[23,32] We reported large thermodynamic destabilizations and changes in CD spectra for d(cGNABg) sequences when the first and third positions of the loop were changed and proposed an interaction between the G and A.[21] These results and the comparison to d(cGNAg) suggested that the d(cGNABg) hairpin loop might also have a sheared GA base pair. To test this possibility directly, the G and A of d(cGCACg) were probed by functional group substitution and by NMR spectroscopy.

The four most common GA pairings involving at least two hydrogen bonds are provided in Figure 1.[33] The GA imino pairing is between the Watson−Crick faces of the G and the A (Figure 1A); the GA N1−N7, carbonyl-amino pairing is between the Watson−Crick face of the G and the Hoogsteen face of the A (Figure 1B); the GA N3-amino, amino-N1 pairing is between the minor groove face of the G and the Watson−Crick face of the A (Figure 1C); and the GA sheared pairing is between the minor groove face of the G and the Hoogsteen face of the A (Figure 1D). To distinguish between these possibilities, the

(26) Di Cera, E. *Chem. Rev.* **1998**, *98*, 1563−1592.
(27) Abramovitz, D. L.; Pyle, A. M. *J. Mol. Biol.* **1997**, *266*, 493−506.
(28) Theimer, C. A.; Finger, L. D.; Feigon, J. *RNA* **2003**, *9*, 1446−1455.
(29) Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C. *J. Am. Chem. Soc.* **2003**, *125*, 2390−2391.
(30) Horovitz, A.; Fersht, A. R. *J. Mol. Biol.* **1990**, *214*, 613−617.
(31) Klostermeier, D.; Millar, D. P. *Biochemistry* **2002**, *41*, 14095−14102.
(32) Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K.; Hirao, I. *Biochemistry* **1997**, *36*, 4761−4767.
(33) Tinoco, I., Jr. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1993; pp 603−607.

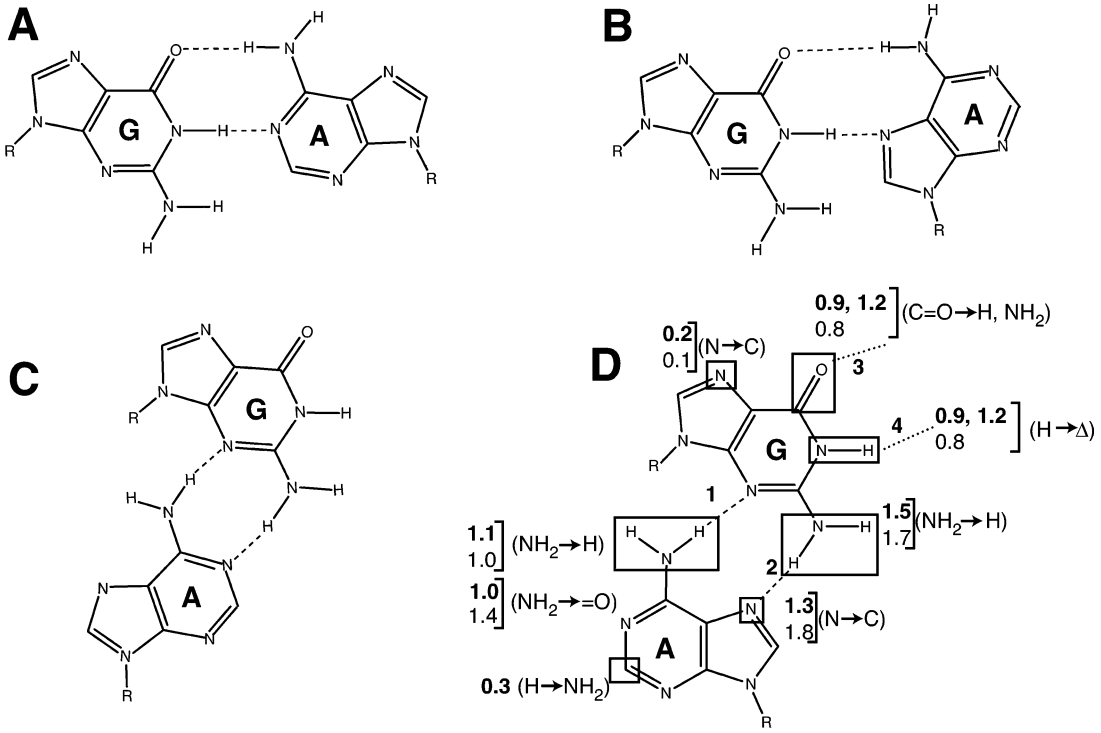

Figure 1. Possible GA base pairs.[33] (A) GA imino; (B) GA N1−N7, carbonyl-amino; (C) GA N3-amino, amino-N1; and (D) GA sheared. Also shown in panel D are $\Delta\Delta G°_{37}$ values for substitutions with purine derivatives for d(cGCACg) (bold font) and d(cGCAg) (normal font) (see Table 1); the functional group substitutions are also provided. Substitutions with pyrimidines were also performed. Hydrogen bonds 1 and 2 (dashed lines) are shown, as are two potential interactions, 3 and 4 (dotted lines), from the major groove face of the G to the CG closing base pair of the stem, which lies below based on the structure of d(cGCAg).[23] Dashed lines are not used for interactions 3 and 4 because it is not known if they involve hydrogen bonds.[24]

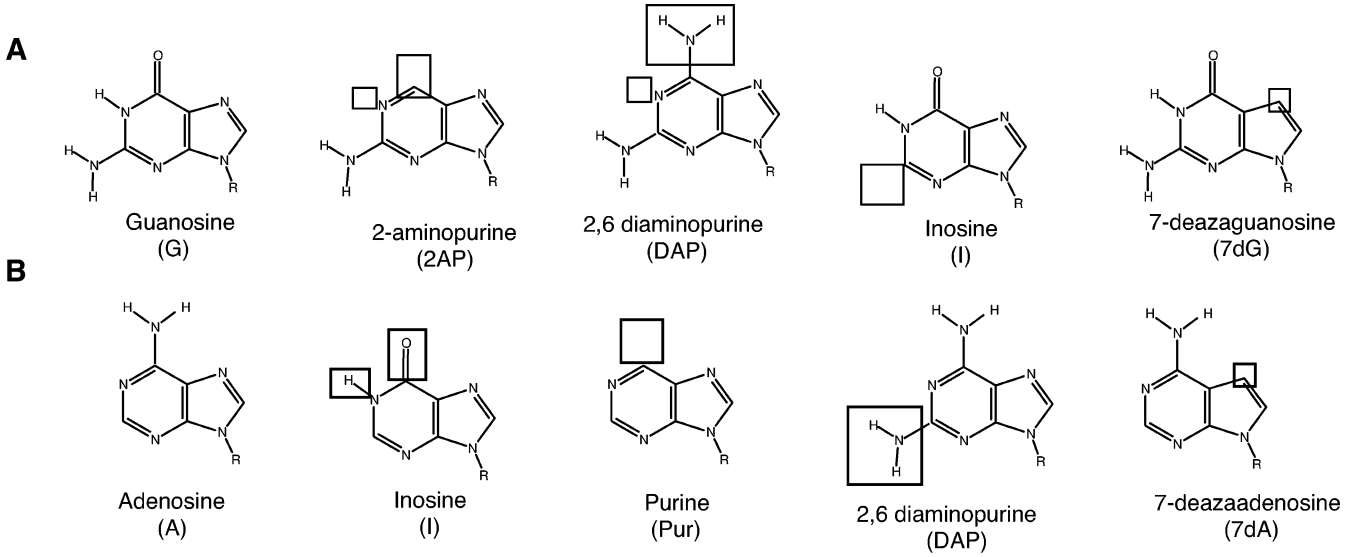

Figure 2. Nucleotide analogues used in this study. (A) Guanosine and analogues used to replace it and (B) adenosine and analogues used to replace it. Boxes represent areas of change between the natural base and the analogue.

functional groups were substituted with various nucleotide analogues and the effects on stability were determined.

The nucleotide analogues used and representative UV melts are provided in Figures 2 and 3. In some cases one functional group was exchanged for another, while in other cases it was possible to delete a functional group altogether, which lessens the possibility of creating new, compensating interactions. Substituting purine (Pur) at position 3 in the loop, d(cGC*Pur*Cg), destabilized the hairpin with a $\Delta\Delta G_{37}°$ of 1.09 kcal/mol and ΔT_M of −7.3 °C (Table 1). This substitution eliminates the 6-amino group of A and would disrupt hydrogen bonding in each of the four possible pairings (Figure 1 A−D), consistent with the destabilizing effect. The substitution of 7dG (7-deazaguanosine) at position 1 of the loop, d(c*7dG*CACg), did not give a significant thermodynamic penalty ($\Delta\Delta G_{37}° = 0.22$ kcal/mol and $\Delta T_M = -0.0$ °C) (Figure 3), which is also consistent with all four pairing possibilities and with values for 7dG substitution in the stem.[25,34] Addition of an amino group to C2 of A, d(cGC*DAP*Cg) (DAP, 2,6-diaminopurine) did not have a large effect either ($\Delta\Delta G_{37}° = 0.31$ kcal/mol and ΔT_M

(34) Burkard, M. E.; Turner, D. H. *Biochemistry* **2000**, *39*, 11748−11762.

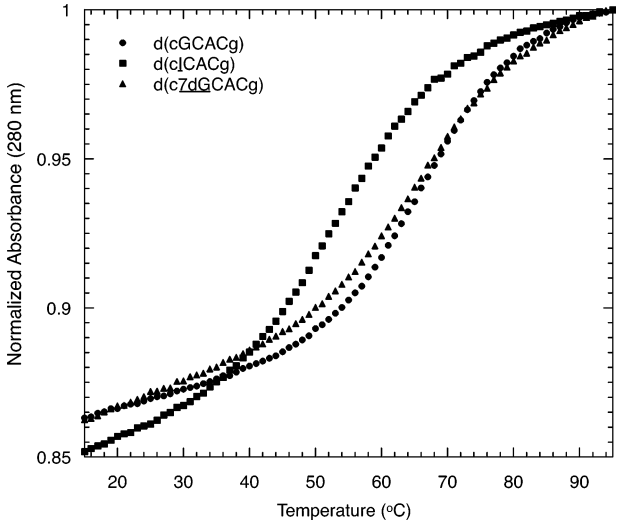

Figure 3. Representative UV melting curves. Melts were carried out as described in the Materials and Methods section. d(cGCACg) (●) was the reference sequence used in these studies. d(c*I*CACg) (■) was destabilized, with a $\Delta\Delta G_{37}°$ of 1.47 kcal/mol and ΔT_M of −11.7 °C. d(c*7dG*CACg) (▲) was essentially unchanged in stability, with a $\Delta\Delta G_{37}°$ of 0.22 kcal/mol and ΔT_M of −0.0 °C. Absorbance values were normalized by dividing each trace by its maximum absorbance value.

= −1.3 °C), consistent with all four pairings except perhaps the GA imino pairing where the amino of G might present a steric clash (Figure 1A); additional data below are also inconsistent with the imino pairing.

Next, we describe substitutions that provide discrimination among the GA pairing possibilities. Substitution of the loop G with I, d(c*I*CACg) gave a large energetic penalty with a $\Delta\Delta G_{37}°$ of 1.47 kcal/mol and ΔT_M of −11.7 °C (Figure 3, Table 1). Effects of this substitution, which eliminates the 2-amino group of G, support pairings C and D but not pairings A and B (Figure 1). To distinguish between pairings C and D, 7-deazaadenosine (7dA) was substituted for the loop A, d(cGC*7dA*Cg). This change had a large destabilizing effect ($\Delta\Delta G_{37}°$ of 1.33 kcal/mol and ΔT_M of −8.5 °C), supporting pairing D with a sheared GA but not pairing C. For comparison, substitution of 7dA for an A in the stem gives a much smaller destabilization, with a $\Delta\Delta G_{37}°$ of 0.50 kcal/mol and a ΔT_M of −2.6 °C.[25]

G1 of the loop was further probed by substitution with DAP and 2AP (2-aminopurine), which were destabilizing with $\Delta\Delta G_{37}°$ values of 1.16 and 0.94 kcal/mol, respectively, and ΔT_M values of −8.4 °C. For comparison, similar values were found for the d(cGCAg) triloop (see below), and the imino proton and carbonyl of the G of a sheared GA base pair provide similar energetic contributions in an RNA hairpin loop.[35]

Similar functional group substitutions have been made throughout the d(cGCAg) triloop (Figure 1D). In general, the thermodynamic consequences of functional group substitutions

Table 1. Thermodynamic Parameters for Folding of Single, Double, and Triple Mutants in the d(cGCACg) and d(cGCAg) Hairpins[a]

	$\Delta H°$[b] (kcal mol^{-1})	$\Delta S°$[b] (cal mol^{-1} K^{-1})	$\Delta G°_{37}$[b,c] (kcal mol^{-1})	T_M (°C)	ΔG_A[b−d] (kcal mol^{-1})
			GCAC Loop Modifications		
GCAC[e,f]	−33.9 ± 0.6	−99.6 ± 2.2	−2.97 ± 0.09	67.0 ± 1.6	
*I*CAC[f]	−26.9 ± 0.9	−82.0 ± 2.8	−1.50 ± 0.05	55.3 ± 0.5	1.47 ± 0.10
*DAP*CAC	−27.9 ± 0.8	−84.1 ± 2.5	−1.81 ± 0.06	58.6 ± 0.3	1.16 ± 0.11
*2AP*CAC	−31.2 ± 0.6	−94.3 ± 1.8	−2.03 ± 0.04	58.6 ± 0.4	0.94 ± 0.10
*7dG*CAC	−31.2 ± 1.1	−91.6 ± 3.4	−2.75 ± 0.06	67.0 ± 0.5	0.22 ± 0.11
GC*7dA*C	−25.3 ± 1.9	−76.3 ± 6.0	−1.64 ± 0.11	58.5 ± 1.0	1.33 ± 0.14
GC*Pur*C	−27.5 ± 1.2	−82.7 ± 3.8	−1.88 ± 0.07	59.7 ± 1.6	1.09 ± 0.11
GC*DAP*C	−31.4 ± 1.6	−92.6 ± 4.6	−2.66 ± 0.13	65.7 ± 0.2	0.31 ± 0.16
GC*I*C	−30.5 ± 1.8	−92.1 ± 5.6	−1.94 ± 0.13	58.1 ± 1.1	1.03 ± 0.16
*I7dA*C	−24.2 ± 2.9	−72.9 ± 8.8	−1.60 ± 0.17	59.0 ± 1.4	1.37 ± 0.19
*ICI*C	−30.1 ± 0.9	−91.9 ± 2.6	−1.60 ± 0.13	54.4 ± 1.3	1.37 ± 0.16
			GCA Loop Modifications		
GCA[e,f]	−31.7 ± 1.4	−90.7 ± 4.1	−3.60 ± 0.14	76.8 ± 0.5	
*I*CA[f]	−28.8 ± 1.3	−86.6 ± 4.0	−1.95 ± 0.05	59.6 ± 0.9	1.65 ± 0.15
*2AP*CA[f]	−32.6 ± 1.6	−95.9 ± 4.7	−2.81 ± 0.15	66.3 ± 0.8	0.79 ± 0.21
*7dG*CA[f]	−27.0 ± 3.3	−75.6 ± 9.6	−3.55 ± 0.32	84.2 ± 3.3	0.05 ± 0.32
GC*7dA*[f]	−26.5 ± 2.6	−79.7 ± 7.9	−1.78 ± 0.13	59.4 ± 0.9	1.82 ± 0.19
GC*I*[f]	−28.8 ± 0.9	−85.5 ± 2.9	−2.25 ± 0.11	63.3 ± 1.3	1.35 ± 0.18
GCA*C3*[f]	−34.7 ± 1.1	−99.6 ± 3.2	−3.81 ± 0.09	75.3 ± 0.5	−0.21 ± 0.17
GCA*C*[f]	−33.9 ± 0.6	−99.6 ± 2.2	−2.97 ± 0.09	67.0 ± 1.6	0.63 ± 0.17
*g*GCA*c*[f]	−14.4 ± 2.4	−44.5 ± 7.6	−0.63 ± 0.38	51.3 ± 8.9	2.97 ± 0.40
ICI[f]	−26.2 ± 0.6	−79.4 ± 1.8	−1.53 ± 0.05	56.3 ± 0.9	2.07 ± 0.15
ICAC3	−33.5 ± 3.1	−100.7 ± 9.8	−2.31 ± 0.07	60.1 ± 1.9	1.29 ± 0.16
GC*I C3*[f]	−31.9 ± 1.5	−95.7 ± 4.6	−2.24 ± 0.10	60.4 ± 1.2	1.36 ± 0.17
GC*I C*	−30.5 ± 1.8	−92.1 ± 5.6	−1.94 ± 0.13	58.1 ± 1.1	1.66 ± 0.19
ICAC[f]	−26.9 ± 0.9	−82.0 ± 2.8	−1.50 ± 0.05	55.3 ± 0.5	2.10 ± 0.15
*g*GC*I c*	−24.1 ± 5.4	−75.7 ± 17.6	−0.61 ± 0.09	45.5 ± 2.5	2.99 ± 0.17
g ICAc	−27.6 ± 5.5	−87.3 ± 17.7	−0.56 ± 0.5	43.5 ± 1.1	3.04 ± 0.15
ICI C	−30.1 ± 0.9	−91.9 ± 2.6	−1.60 ± 0.13	54.4 ± 1.3	2.00 ± 0.19
ICI C3	−30.4 ± 1.6	−91.7 ± 5.0	−1.98 ± 0.07	58.6 ± 0.9	1.62 ± 0.16
g ICI c	−24.0 ± 7.0	−76.7 ± 22.3	−0.18 ± 0.18	39.2 ± 1.8	3.42 ± 0.23

[a] All sequences are DNA and are for loops that have a CG closing base pair unless otherwise indicated. These hairpin loops conform to the d(cGNABg) or d(cGNAg) motifs.[21,32] [b] Errors are the standard deviations from three or more measurements and were propagated by standard methods. [c] An extra significant figure is provided to avoid round-off error in subsequent calculations. [d] ΔG_A is the free energy change associated with mutation A. [e] Sequences in bold type are the reference for the sequences below. Some oligonucleotides appear more than once in the table, since they have multiple reference sequences. Sequences are listed in order of position then most penalizing change and grouped by single, double, and triple mutations. [f] Thermodynamic parameters from previous studies,[21,24,25] but are provided here to facilitate comparisons and analysis of thermodynamic cubes.

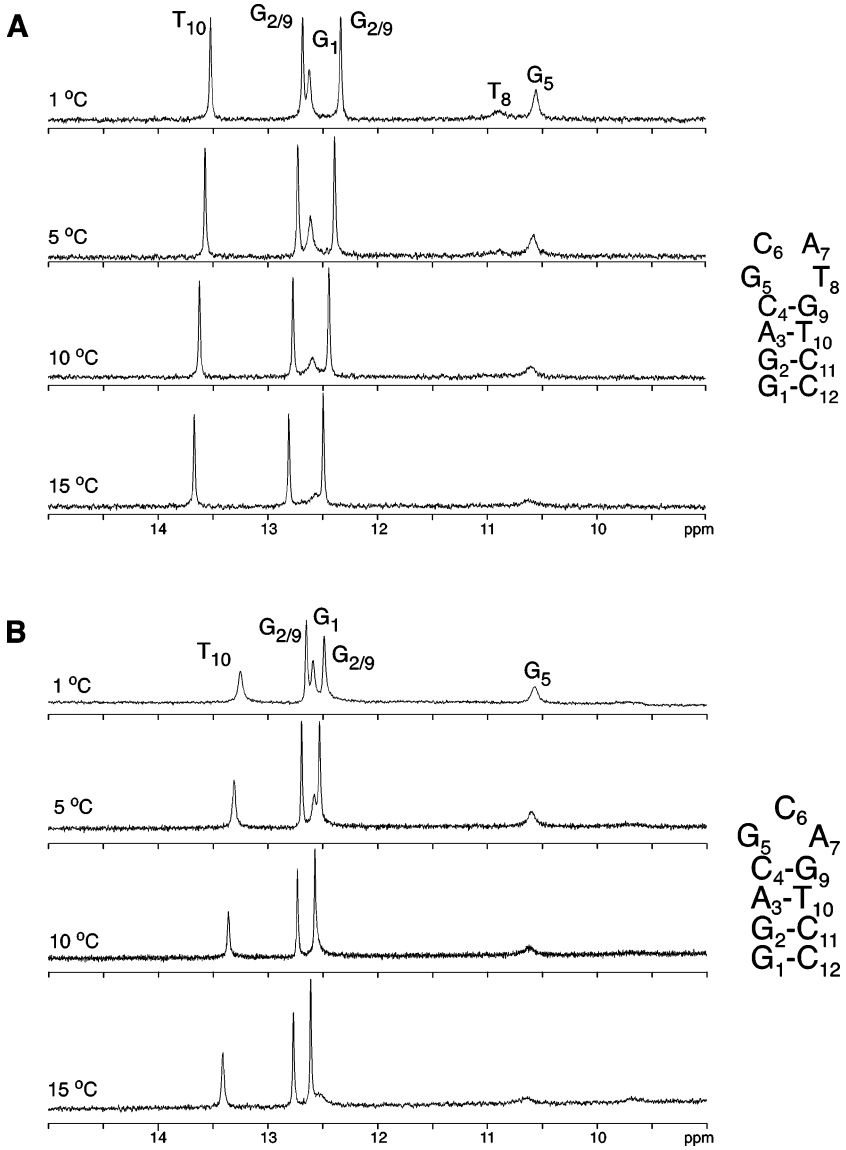

Figure 4. Exchangeable proton NMR spectra (9−15 ppm) in 10 mM phosphate and 0.1 mM Na_2EDTA buffer, pH 7. Sequence and numbering for the full-length hairpins are provided. Numbering of the stem in the triloop hairpin is based on the tetraloop hairpin. (A) Spectra of d(cGCATg) at 240 μM at 1, 5, 10, and 15 °C. (B) Spectra of d(cGCAg) at 280 μM at 1, 5, 10, and 15 °C. Assignments were made as described in the text. The assignments of G2 and G9 were ambiguous and so are indicated with "$G_{2/9}$".

in the triloop and tetraloop hairpins are quite similar. For example, the G to 7dG substitution gave a small destabilization of $\Delta G_{37}^\circ = 0.05$ kcal/mol, similar to the effect of 0.22 kcal/mol in the tetraloop. The G to I substitution was very destabilizing with a $\Delta\Delta G_{37}^\circ$ of 1.65 kcal/mol, similar to 1.47 kcal/mol. The A to 7dA substitution was also significantly destabilizing with a $\Delta\Delta G_{37}^\circ$ of 1.82 kcal/mol and a ΔT_M of −17.4 °C, compared to 1.33 kcal/mol and −8.5 °C in the tetraloop. Also, the 7dA effects are in agreement with values reported for d(cGAAg).[23] 2AP substitution at the first position of the loop, d(c2*AP*CAg), was destabilizing by 0.79 kcal/mol, similar to the value of 0.94 kcal/mol in the tetraloop. In summary, the patterns of $\Delta\Delta G_{37}^\circ$ effects for the triloop and tetraloop hairpins are highly similar supporting sheared GA pairings in both loops.

NMR Characterization of the Hairpin Loop. NMR structures of several d(cGNAg) hairpins have been determined and shown to contain a sheared GA base pair.[23,32] Data on the thermodynamic consequences of functional group substitutions strongly supported a sheared GA pair in the d(cGNABg) loops as well (previous section). To provide further support for the sheared GA pair in the tetraloops, NMR experiments of several d(cGNABg) loops were conducted. Initially, ^{31}P NMR spectra were collected on d(cGCATg) and d(cGCAg) hairpins (data not shown). After accounting for the resonance from phosphate buffer, the expected 11 resonances for d(cGCATg) and 10 resonances for d(cGCAg) were observed. These resonances were sharp and lacked the complication of additional resonances that might come from a population of duplex conformation, consistent with these sequences exclusively adopting the hairpin conformation.

Next, ^1H NMR was performed on d(cGCATg), d(cGCACg), and d(cGCAg) loops to observe exchangeable protons. The number of peaks and their chemical shifts were virtually identical between d(cGCATg) and d(cGCAg) (Figure 4), consistent with the formation of hairpins with a sheared GA

(35) SantaLucia, J., Jr.; Kierzek, R.; Turner, D. H. *Science* **1992**, *256*, 217−219.

Table 2. Free Energy Parameters and δ Values for Double Mutant Cycles in the d(cGCACg) Hairpin[a]

	interaction probed[b]	ΔG_A[c] (kcal/mol)	ΔG_B[d] (kcal/mol)	ΔG_{AB}[e] (kcal/mol)	ΔG_{AB} (if additive)[f] (kcal/mol)	δ_{ab}[g]	δ_{ab} (kcal/mol)
			Probing Sheared GA in d(GCAC)				
IC7dAC	2	1.47 ± 0.10	1.33 ± 0.14	1.37 ± 0.19	2.80 ± 0.15	red.	−1.4 ± 0.2
ICIC	2, 1	1.47 ± 0.10	1.03 ± 0.16	1.37 ± 0.16	2.50 ± 0.17	δ_{12}	−1.1 ± 0.2
			Thermodynamic Box A				
GC*I C*	1, C exp	1.35 ± 0.18	0.63 ± 0.17	1.66 ± 0.19	1.98 ± 0.20	δ_{1C}	−0.3 ± 0.2
ICAC	2, C exp	1.65 ± 0.15	0.63 ± 0.17	2.10 ± 0.15	2.28 ± 0.17	δ_{2C}	−0.2 ± 0.2
ICI	2,1	1.65 ± 0.15	1.35 ± 0.18	2.07 ± 0.15	3.00 ± 0.18	δ_{12}	−0.9 ± 0.2
			Thermodynamic Box B				
GC*I C3*	1, C3 exp	1.35 ± 0.18	−0.21 ± 0.17	1.36 ± 0.17	1.14 ± 0.20	δ_{1C3}	0.2 ± 0.2
ICAC3	2, C3 exp	1.65 ± 0.15	−0.21 ± 0.17	1.29 ± 0.16	1.44 ± 0.17	δ_{2C3}	−0.2 ± 0.2
			Thermodynamic Box C				
*g*GC*I c*	1, 3	1.35 ± 0.18	2.97 ± 0.40	2.99 ± 0.17	4.32 ± 0.42	δ_{1cbp}	−1.3 ± 0.4
g ICAc	2, 3	1.65 ± 0.15	2.97 ± 0.40	3.04 ± 0.15	4.62 ± 0.41	δ_{2cbp}	−1.6 ± 0.4

	interaction probed	$^c\Delta G_A$[h] (kcal/mol)	$^c\Delta G_B$[h] (kcal/mol)	$^c\Delta G_{AB}$[h] (kcal/mol)	$^c\Delta G_{AB}$[h] (if additive) (kcal/mol)	$^c\delta_{ab}$[h]	$^c\delta_{ab}$[h] (kcal/mol)
ICI C	1, C exp in 2 bkg	0.42 ± 0.07	0.45 ± 0.07	0.35 ± 0.14	0.87 ± 0.09	$^2\delta_{1C}$	−0.5 ± 0.2
	2, C exp in 1 bkg	0.72 ± 0.12	0.31 ± 0.17	0.65 ± 0.17	1.03 ± 0.18	$^1\delta_{2C}$	−0.4 ± 0.2
	2,1 in C exp bkg	1.47 ± 0.10	1.03 ± 0.16	1.37 ± 0.16	2.50 ± 0.17	$^C\delta_{12}$	−1.1 ± 0.2
ICI C3	1, C3 exp in 2 bkg	0.42 ± 0.07	−0.36 ± 0.09	−0.03 ± 0.09	0.06 ± 0.10	$^2\delta_{1C3}$	−0.1 ± 0.1
	2, C3 exp in 1 bkg	0.72 ± 0.12	0.01 ± 0.15	0.27 ± 0.13	0.73 ± 0.16	$^1\delta_{2C3}$	−0.5 ± 0.2
	2, 1 in C3 exp bkg	1.50 ± 0.11	1.57 ± 0.13	1.83 ± 0.11	3.07 ± 0.15	$^{C3}\delta_{12}$	−1.2 ± 0.2
g ICI c	1, 3 in 2 bkg	0.42 ± 0.07	1.39 ± 0.07	1.77 ± 0.19	1.81 ± 0.09	$^2\delta_{1cbp}$	−0.1 ± 0.2
	2, 3 in 1 bkg	0.72 ± 0.12	1.64 ± 0.14	2.07 ± 0.21	2.36 ± 0.15	$^1\delta_{2cbp}$	−0.3 ± 0.2
	2, 1 in 3 bkg	0.07 ± 0.38	0.02 ± 0.39	0.45 ± 0.42	0.09 ± 0.39	$^{cbp}\delta_{12}$	0.4 ± 0.4

[a] All sequences are DNA and are for loops that have a CG closing base pair unless otherwise specified. [b] "Interaction probed" refers to the interactions shown in Figure 1D for the sheared GA conformation. "C exp" and "C3 exp" refer to expansion of the loop at the 3′-end by cytidine and a 3-carbon spacer, respectively. Interactions 3 and 4 are represented by 3 for simplification. [c] ΔG_A values are the free energy changes associated with the single modification that breaks the first of the "interactions probed" listed. [d] ΔG_B values are the free energy changes associated with the single modification that breaks the second of the "interactions probed" listed. [e] ΔG_{AB} values are the free energy changes associated with both modifications in a single oligonucleotide. [f] Values are the sum of ΔG_A and ΔG_B. [g] δ values were calculated as the difference betweens columns 5 and 6, and errors were propagated from eq 1a. "Red." refers to redundant modifications that affect the same interaction. In all other cases, the two interactions, and any background interaction, are given for δ. [h] Superscript "c" refers to quantities measured in the background of a third change.

base pair. Four resonances were located in the Watson−Crick hydrogen bonded imino region (12−14 ppm), with one to two imino resonances shifted upfield to between 10 and 11 ppm. The T10 resonance was near 13.5 ppm in both hairpins, as expected for a Watson−Crick AT base pair[36] and was confirmed by NOEs to both G2 and G9 (data not shown). The cluster of three resonances between 12 and 13 ppm is consistent with the expected chemical shifts for the three Gs in Watson−Crick GC base pairs.[36] The center peak in this cluster of Gs broadened with temperature between 1 and 15 °C (Figure 4), consistent with fraying and assignment as the terminal base pair.[37,38] The two other Gs were confirmed as G2 and G9 by both having an NOE to T10 but not to each other (data not shown). The imino protons of G1 and G5 were not visible in the NOE experiment prohibiting further assignment of G2 and G9; however this was not crucial to this study.

The resonance at ≈10.6 ppm in both the triloop and tetraloops broadened further with temperature between 1 and 15 °C, consistent with absence of protection from chemical exchange.[37] Similar behavior has been reported for sheared GA base pairs in RNA duplexes.[39] Moreover, the imino proton in a sheared GA in a DNA duplex resonates at ∼10.5 ppm.[40,41] Together, these data led us to assign the 10.6 ppm resonance to the sheared GA base pair. It should be noted that the behavior of this resonance is inconsistent with findings for the imino proton of a GA imino base pair (Figure 1A), which is sharp and typically resonates near 12 ppm,[38] as well as expectations for a GA N1−N7, carbonyl-amino base pair (Figure 1B). The additional resonance at ∼10.9 ppm in the tetraloop spectrum was identified as T8 (Figure 4A) due to its absence from the triloop spectrum. The T8 resonance in d(cGCATg) was very broad and broadened further with temperature until it was absent by 15 °C. This observation is also consistent with the fourth base of the d(cGNABg) tetraloop being accessible to exchange with solvent[37] and supports the notion that the d(cGNABg) loop is a d(cGNAg) triloop with the extra base extruded into solution.[21,24] Proton NMR on d(cGCACg) was also carried out and revealed an imino spectrum identical to that of d(cGNAg), including absence of the 10.9 ppm resonance (data not shown). These data support d(cGCACg), which is the reference d(cGNABg) sequence in the thermodynamic cycles described below, having the same fold as d(cGCATg), as well as assignment of T8.

In summary, thermodynamic analyses of functional group substituted oligonucleotides and NMR data together support the d(cGNABg) hairpin loops having a similar structure to d(cGNAg) with a sheared GA base pair and the "B" position extruded

(36) Schroeder, S. J.; Turner, D. H. *Biochemistry* **2000**, *39*, 9257−9274.
(37) Gueron, M.; Kochoyan, M.; Leroy, J. L. *Nature* **1987**, *328*, 89−92.
(38) Wu, M.; SantaLucia, J., Jr.; Turner, D. H. *Biochemistry* **1997**, *36*, 4449−4460.
(39) SantaLucia, J., Jr.; Turner, D. H. *Biochemistry* **1993**, *32*, 12612−12623.
(40) Chou, S. H.; Cheng, J. W.; Reid, B. R. *J. Mol. Biol.* **1992**, *228*, 138−155.
(41) Chou, S. H.; Cheng, J. W.; Fedoroff, O. Y.; Chuprina, V. P.; Reid, B. R. *J. Am. Chem. Soc.* **1992**, *114*, 3114−3115.

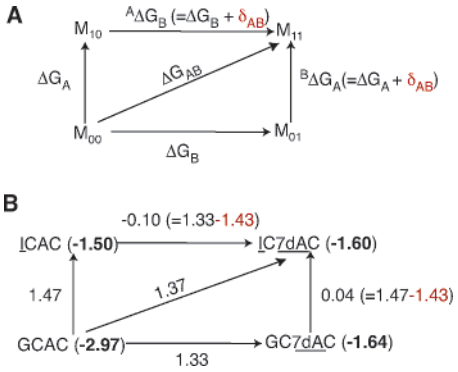

Figure 5. Thermodynamic boxes for double mutant cycles. (A) Cycle showing thermodynamic relationship between mutants A and B. M_{00} is the wild-type, M_{10} is the single mutant A, M_{01} is the single mutant B, and M_{11} is the double mutant. ΔG_A is the free energy change associated with mutation A, and ${}^B\Delta G_A$ is the free energy change for mutation A in the background of mutation B. δ_{AB} (in red) represents the nonadditive free energy of combining the two mutations. δ_{AB} was calculated according to eq 1b. Note also that ${}^B\Delta G_A = \Delta G_A + \delta_{AB}$. (B) Example for the double mutant d(c*IC7dA*Cg). Experimentally measured $\Delta G_{37}°$ values are at the corners of the box (Table 1). δ_{AB} (red) for this cycle is -1.43 ± 0.20 kcal/mol (Table 2).

into solution. Double mutant cycles on d(cGNABg) loops (described below) further strengthen this conclusion.

Effects of Loop Expansion on Energetics and Cooperativity. Comparison between d(cGCACg) and d(cGCAg) reveals that the 3′-expanded loop is less stable by a $\Delta\Delta G_{37}°$ of 0.6 kcal/mol (Table 1). Examination of the enthalpy and entropy changes for these two sequences suggests that the expanded sequence is less stable for entropic reasons ($-\Delta T\Delta S$ of 2.8 kcal/mol); in fact, the ΔH for d(cGCACg) is slightly more favorable than that for d(cGCAg) ($\Delta\Delta H$ of -2.2 kcal/mol). Apparently the extra nucleotide, which is extruded into solution, incurs a significant loss in entropy without a complete enthalpic compensation.

To test for cooperativity of the interactions within the loop, double mutants were constructed and analyzed. The first double mutant, d(c*IC7dA*Cg), tests the presence of hydrogen bond 2 in the sheared GA of a d(cGNABg) loop (Figure 1D). The double substitution had a destabilizing effect of $\Delta\Delta G_{37}° = 1.37$ kcal/mol (Table 2), similar to the effect of each of the single modifications, which had $\Delta\Delta G_{37}°$ values of 1.47 (G to I) and 1.33 (A to 7dA) kcal/mol. These data result in a δ of -1.4 ± 0.2 kcal/mol and complete nonadditivity (Figure 5B). Complete nonadditivity strongly supports these changes affecting the same hydrogen bond and, therefore, further supports the sheared GA pairing in a d(cGNABg) loop (Figure 1D). The same double analogue substitution in the triloop had a very similar δ value of -1.8 ± 0.2 kcal/mol.[25]

Double mutant cycles were repeated in the background of a change at a third site, either expansion of the 3′ end of the loop or modification of the closing base pair. These experiments can be represented on a thermodynamic cube (Figure 6) with the wild-type, three single mutants, three double mutants, and the triple mutant at the vertexes. The first triple mutant cube

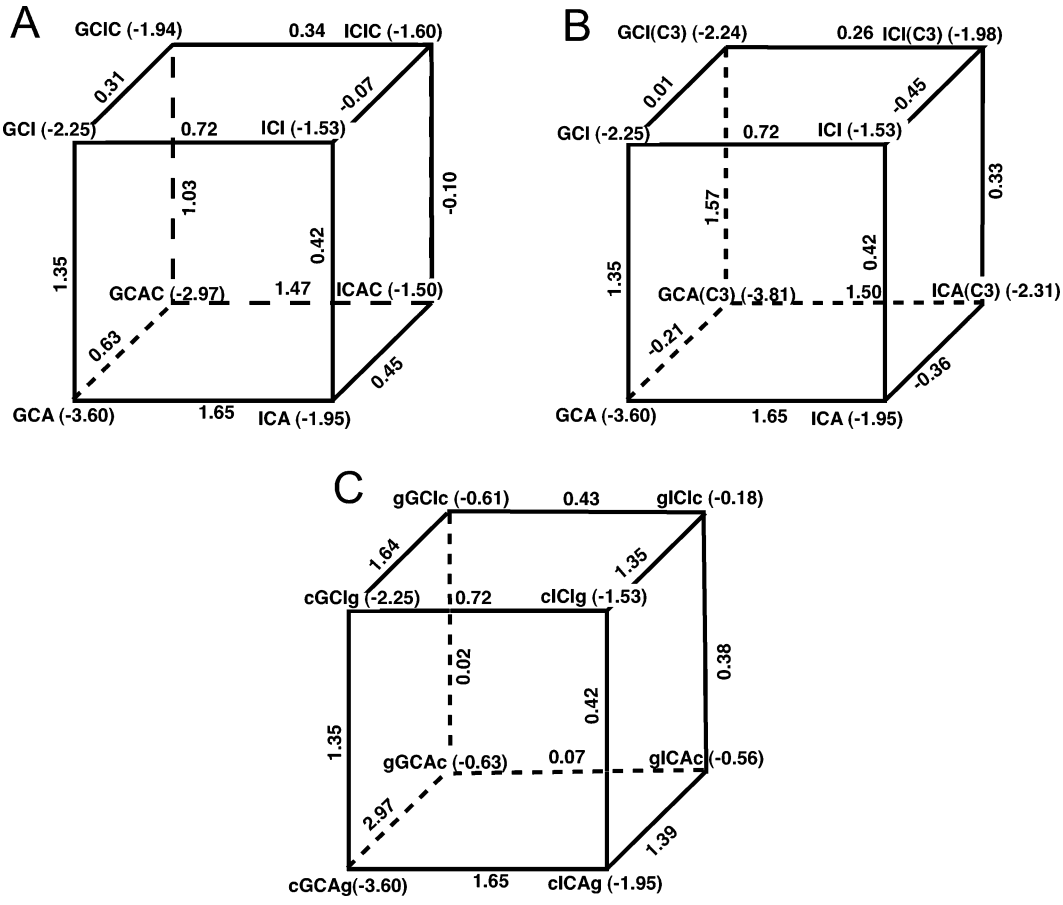

Figure 6. Thermodynamic cubes for triple mutant cycles. (A) Cycle for d(c*ICIC*g), (B) cycle for d(c*ICI C3*g), and (C) cycle for d(*gICIc*). Experimentally measured $\Delta G_{37}°$ values are at the vertexes of the cube, and the free energy change associated with a mutation is given along an edge. δ values are provided in Table 2. Closing base pair is shown explicitly for panel C only.

explores the consequences of expanding the d(cGCAg) loop by adding a cytidine at the 3′ end, d(cGCACg). The δ_{1C} and δ_{2C} values are -0.3 ± 0.2 and -0.2 ± 0.2, respectively, where "C" denotes 3′-expansion of the loop with a cytidine (Table 2). These values are small indicating that the energetics of hydrogen bonds 1 and 2 are not significantly affected by expansion of the loop and that the loop is not significantly coupled to 3′-expansion. Probing δ_{12} in the background of the d(cGNABg) loop, ${}^{C}\delta_{12}$, gives a large coupling free energy of -1.1 ± 0.2, which is the same within error as the δ_{12} in the wild-type background, -0.9 ± 0.2. These comparisons support the expanded d(cGNABg) loop motif having similar energetics and cooperativity as the parent d(cGNAg) loop.

Previous studies have shown that adding one C3 spacer at the 3′ end of the triloop has little effect on stability.[24] The effects of C3 insertion on the cooperativity of loop−loop interactions were therefore studied in d(cGCA *C3* g). Similar to observations for d(cGCACg), values of δ_{1C3} and δ_{2C3} for d(cGCA *C3* g) were small at 0.2 ± 0.2 and -0.2 ± 0.2, respectively. This indicates that expansion of the loop with either a natural base or a three carbon spacer has little thermodynamic consequence on loop stability. This is also evidenced by the δ_{12} term in the background of the C3 expansion, ${}^{C3}\delta_{12} = -1.2 \pm 0.2$, which exhibits a large degree of nonadditivity, essentially identical to the δ_{12} values for the cytidine-expanded and parent triloops.

Last, it can be noted that all of the coupling free energy terms between expansion and a loop interaction remain close to zero in the background of a change in the configuration at the other loop interaction (Table 2). This observation is also consistent with the notion that 3′-expansion of the loop has no appreciable effect on loop energetics.

Effects of Closing Base Pair on Energetics and Cooperativity. The third thermodynamic cube provides the coupling free energy of loop−loop interactions in the presence of a GC closing base pair (Figure 6C). As revealed by δ_{1cbp} and δ_{2cbp} of -1.3 ± 0.5 and -1.6 ± 0.4, respectively, there is a very large coupling between the loop and the closing base pair. A similar effect was found when the loop−closing base pair interaction was disrupted with C3 spacers or 2AP.[24] Since δ_{1cbp} and δ_{2cbp} are approximately equal to $-\Delta G_1$ and $-\Delta G_2$, the closing base pair and loop interaction 1 (and 2) are completely nonadditive; this is also seen in that ${}^{cbp}\Delta G_1 \approx {}^{cbp}\Delta G_2 \approx 0$. Likewise, loop−loop coupling is close to zero in the background of the GC closing base pair, ${}^{cbp}\delta_{12} \approx 0.4 \pm 0.4$, which differs significantly from the δ_{12} of -0.9 ± 0.2. Together, these data indicate that loop−loop coupling is indirect and requires a CG base pair to be optimal.

Conclusions

The d(cGNAg) motif is stabilized by two loop−loop interactions and a loop−closing base pair interaction.[23,32] Functional group substitution and NMR experiments supported the 3′-expanded d(cGNABg) loop having essentially the same structure, with the "B" position extruded into solution. Double and triple mutant cycles showed no significant effect of loop expansion on loop−loop cooperativity. In contrast, closing base pair changes revealed that a CG closing base pair is critical to loop−loop cooperativity. Together, these data support the CG closing base pair forming an essential platform upon which a stable and expandable loop can be assembled. Many RNA and DNA hairpin loops found in nature appear to be expansions of smaller, unusually stable loops.[21,27,28] The results of the present study suggest that such loops may retain not only structural characteristics of the parent loop but also energetic features such as nonadditivity. Similar energetics may occur because local interactions, in particular stacking, are exceptionally strong.

Acknowledgment. Supported by National Science Foundation CAREER Grant MCB-9984129 and a Camille Dreyfus Teacher-Scholar Award and Sloan Fellowship to P.C.B. and a Paul Berg Award to E.M.M.. We thank Chris Falzone and David Proctor for assistance with NMR experiments.

JA048368+