A CA$^+$ Pair Adjacent to a Sheared GA or AA Pair Stabilizes Size-Symmetric RNA Internal Loops[†]

Gang Chen,[‡] Scott D. Kennedy,[§] and Douglas H. Turner*[,‡,‖]

[‡]*Department of Chemistry, University of Rochester, Rochester, New York 14627,* [§]*Department of Biochemistry and Biophysics and* [‖]*Department of Pediatrics, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642*

Received October 15, 2008; Revised Manuscript Received March 24, 2009

ABSTRACT: RNA internal loops are often important sites for folding and function. Residues in internal loops can have pK_a values shifted close to neutral pH because of the local structural environment. A series of RNA internal loops were studied at different pH by UV absorbance versus temperature melting experiments and imino proton nuclear magnetic resonance (NMR). A stabilizing CA pair forms at pH 7 in the $\frac{CG}{AA}$ and $\frac{CA}{AA}$ nearest neighbors when the CA pair is the first noncanonical pair (loop-terminal pair) in 3 × 3 nucleotide and larger size-symmetric internal loops. These $\frac{CG}{AA}$ and $\frac{CA}{AA}$ nearest neighbors, with CA adjacent to a closing Watson−Crick pair, are further stabilized when the pH is lowered from 7 to 5.5. The results are consistent with a significantly larger fraction (from ∼20% at pH 7 to ∼90% at pH 5.5) of adenines being protonated at the N1 position to form stabilizing wobble CA$^+$ pairs adjacent to a sheared GA or AA pair. The noncanonical pair adjacent to the GA pair in $\frac{CG}{AA}$ can either stabilize or destabilize the loop, consistent with the sequence-dependent thermodynamics of GA pairs. No significant pH-dependent stabilization is found for most of the other nearest neighbor combinations involving CA pairs (e.g., $\frac{CA}{AG}$ and $\frac{AG}{CA}$), which is consistent with the formation of various nonwobble pairs observed in different local sequence contexts in crystal and NMR structures. A revised free-energy model, including stabilization by wobble CA$^+$ pairs, is derived for predicting stabilities of medium-size RNA internal loops.

The N1 nitrogen of adenine and N3 nitrogen of cytosine normally have pK_a values of 3.5 and 4.2, respectively, but the pK_a values (*1−3*) and thermodynamic contributions (*4−7*) of noncanonical pairs involving A and C in folded DNA and RNA are sequence- and context-dependent.

General acid−base catalysis, involving protonation and deprotonation of nucleobases at physiological pH, has been found for ribozyme catalysis of cleavage and ligation of specific phosphodiester bonds (*2, 6*). The formation of wobble CA$^+$ (*cis* Watson−Crick/Watson−Crick) pairs (Figure 1b) causes local and global conformational changes in RNA (*8−13*). Understanding the sequence-dependent driving force of a pK_a shift of nucleobases within noncanonical pairs is needed to provide insight into RNA folding and catalytic mechanisms (*6, 7*). It may also facilitate better understanding of the pH-dependent assembly of RNA viruses (*14*).

The thermodynamics of CA pairs is also important for bioinformatic approaches that reveal structure−function relationships for RNA. For example, an approach for identifying which strand of complementary RNAs is most likely to rely on structure for function depends upon the different thermodynamic stabilities of CA and GU pairs (*15*).

Here, thermodynamic stabilities of a variety of RNA internal loops were measured in 1 M NaCl at pH 7 and 5.5. At pH 7, a nearest neighbor of $\frac{CG}{AA}$ or $\frac{CA}{AA}$, with the CA adjacent to a closing canonical pair, can stabilize 3 × 3 nucleotide and larger size-symmetric ($n1 = n2$)[1] internal loops on average by about 1 kcal/mol at 37 °C. Such nearest neighbors with the CA adjacent to a closing Watson−Crick pair are further stabilized on average by 1 kcal/mol at 37 °C when the pH is lowered from 7 to 5.5. Dependent upon the sequence, the noncanonical pair adjacent to

[1]Abbreviations: C_T, total concentration of all strands of oligonucleotides in solution; eu, entropy units in cal mol^{-1} K^{-1}; $n1 \times n2$, an internal loop with $n1$ nucleotides on one side and $n2$ nucleotides on the opposite side; P, purine riboside; RY, canonical pair of GC, AU, or GU, with R on the 5′ side and Y on the 3′ side of the internal loop; size-symmetric internal loops, a $n1 \times n2$ nucleotide internal loop with $n1 = n2$; T_M, melting temperature in kelvins; T_m, melting temperature in degrees Celsius; YR, canonical pair of CG, UA, or UG, with Y on the 5′ side and R on the 3′ side of the internal loop; $\Delta G°_{5'CR/3'AA\ bonus}$, a free-energy bonus derived to account for stabilization in the $\frac{CG}{AA}$ and/or $\frac{CA}{AA}$ nearest neighbors when the CA pair is the first noncanonical pair (loop-terminal pair) in 3 × 3 nucleotide and larger size-symmetric internal loops at pH 7, 1 M NaCl, and 37 °C; $\Delta G°_{5'CR/3'AA,\ pH\ bonus}$, the free-energy bonus derived to account for stabilization from pH 7 to 5.5 in the $\frac{CG}{AA}$ and/or $\frac{CA}{AA}$ nearest neighbors when the CA pair is adjacent to a closing Watson−Crick pair in 3 × 3 nucleotide and larger size-symmetric internal loops ($\Delta G°_{5'CR/3'AA,\ pH\ bonus}$ is also applied for loops with tandem CA pairs); $\Delta G°_{37,pH7,loop}$, the measured loop free energy at 37 °C and pH 7; $\Delta\Delta G°_{37,pH}$, the measured loop free-energy difference between pH 5.5 and 7, unless otherwise noted (see the footnotes of the tables).
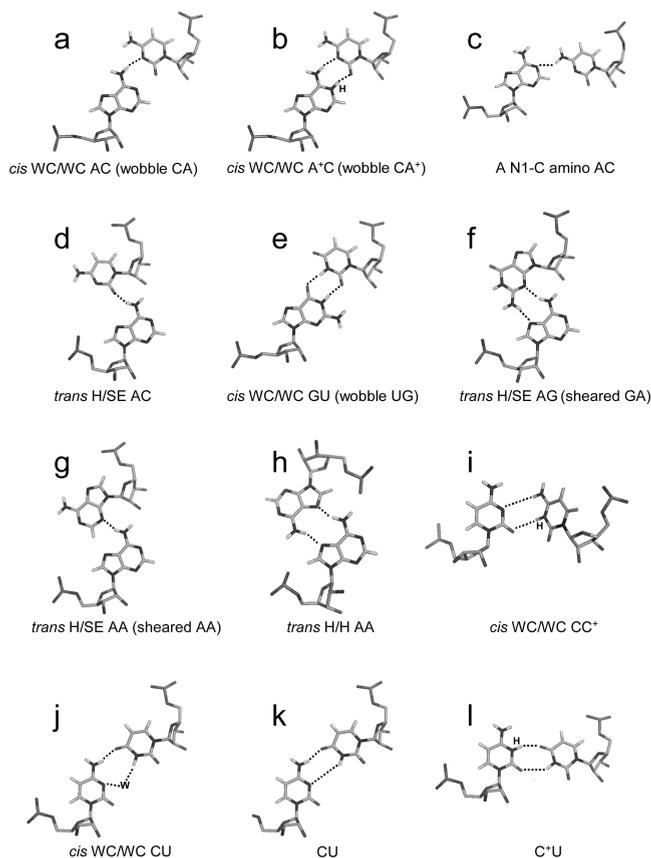
Published on Web 06/01/2009

FIGURE 1: Structures of several base pairs discussed in the paper. The hydrogen atoms in phosphate—sugar backbones are not shown. Only the base—base hydrogen bonds are shown. The proton from protonation and the bridging water are labeled with H and W, respectively. Most of the structures are taken from the BGSU Basepair Catalogue (http://rna.bgsu.edu/FR3D/basepair/). See ref *43* for notations.

the GA pair in $\frac{Y}{R}\frac{CG}{AA}$ or $\frac{R}{Y}\frac{CG}{AA}$ can either stabilize or destabilize the medium-size internal loops, consistent with the previous thermodynamic model (*16*). A better understanding of the protonation effects should help improve the prediction of the RNA internal loop structure and stability and provide a deeper insight into folding and function of large RNA.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Purification. Oligonucleotides were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer using the phosphoramidite method (*17*, *18*), deprotected, and purified, as described previously (*19*, *20*). Controlled pore glass (CPG) supports and phosphoramidites were purchased from Proligo, AZCO, Glen Research, or ChemGenes. The mass of all oligonucleotides was verified by electrospray ionization mass spectrometry (ESI–MS). Purities were checked by reverse-phase high-performance liquid chromatography (HPLC) or analytical thin-layer chromatography (TLC), and all were greater than 95% pure.

UV Absorbance Versus Temperature Melting Experiments and Thermodynamics. Concentrations of single-stranded oligonucleotides were approximated from the absorbance at 280 nm and 80 °C, and extinction coefficients were predicted from those of dinucleotide monophosphates and nucleosides (*21*, *22*) with RNAcalc (http://www.meltwin.com) (*23*). The extinction coefficients were estimated by replacing purine riboside with adenosine. Although extinction coefficients

differ upon functional group substitutions, individual nucleotides contribute only a small portion of the oligomer extinction and, thus, do not significantly affect thermodynamic measurements. UV melting buffer conditions were 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM sodium ethylenediaminetetraacetic acid (EDTA) at pH 7 and 5.5 or 1 M NaCl, 20 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid) (HEPPS), 0.5 mM sodium EDTA at pH 8. Cacodylate and HEPPS were used because their pK_a values are essentially temperature-independent. Curves of absorbance at 280 nm versus temperature were acquired using a heating rate of 1 °C/min with a Beckman Coulter DU640C spectrophotometer, having a Peltier temperature controller cooled with a water bath.

Melting curves were first fit to a two-state model with MeltWin (http://www.meltwin.com) (*23*), assuming linear sloping baselines and temperature-independent $\Delta H°$ and $\Delta S°$ (*23–25*). Presumably, the pK_a values do not change until the RNA duplex melts; i.e., pK_a values exhibit a two-state manner (with zero-sloping baselines) coupled with the melting of an RNA structure (*7*). This is a reasonable assumption because nucleobase protonation/deprotonation is linked with the two-state folding/ unfolding of the RNA duplex. The temperature at which half the strands are in duplex, T_M, at total strand concentration, C_T, was used to calculate thermodynamic parameters for duplex formation according to (*26*)

$$T_M{}^{-1} = (R/\Delta H°)\ln(C_T/a) + (\Delta S°/\Delta H°) \qquad (1)$$

Here, R is the gas constant, 1.987 cal mol^{-1} K^{-1}, and a is 1 for self-complementary duplexes and 4 for non-self-complementary duplexes. All of the $\Delta H°$ values from $T_M{}^{-1}$ versus $\ln(C_T/a)$ plots (eq 1) and from the average of the fits of melting curves to two-state transitions agree within 15%, suggesting that the two-state model is a reasonable approximation for these transitions. The equation $\Delta G°_{37} = \Delta H° - (310.15)\Delta S°$ was used to calculate the free-energy change at 37 °C (310.15 K).

Exchangeable Proton NMR Spectroscopy. All exchangeable proton spectra (*27*) were acquired on a Varian Inova 500 MHz (^{1}H) spectrometer. One-dimensional imino proton spectra were acquired with an S pulse sequence (*28*) at temperatures ranging from −5 to 40 °C in 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM sodium EDTA. SNOESY spectra (*28*) were recorded with an 150 ms mixing time from −5 to 10 °C. The Felix (2000) software package (Molecular Simulations, Inc.) was used to process 2D spectra. Proton spectra were referenced to H$_2$O or HDO at a known temperature-dependent chemical shift relative to 3-(trimethylsilyl)tetradeutero sodium propionate (TSP).

RESULTS

Thermodynamics at Different pH. An RNA secondary-structure prediction and analysis program, RNAstructure 4.2 (http://rna.urmc.rochester.edu/rnastructure.html) (*29*), was used to design sequences that form heteroduplexes without competing homoduplexes. Thermal melting studies of the individual single strands (*16*, *19*) (see Table S1 in the Supporting Information) confirmed the absence of competing homoduplexes. Measured thermodynamic parameters at 1 M NaCl for duplexes and internal loops (calculated by eq 3a shown below) are listed in Tables 1 and 2, respectively. For a given duplex or internal loop, the values from bottom to top are for pH values 5.5, 7, and 8, respectively. In Tables 1 and 2, most sequences are ordered from

Table 1: Measured Thermodynamic Parameters for RNA Duplex Formation in 1 M NaCl[a]

Sequences	Linear fit of T_M^{-1} vs $\ln(C_T/a)$ (eq 1)					Average of two-state melt curve fits			
	$\Delta\Delta G^\circ_{37,\,pH}$	$-\Delta H^\circ_{37}$	$-\Delta S^\circ$	$-\Delta G^\circ_{37}$	T_m	$-\Delta H^\circ$	$-\Delta S^\circ$	$-\Delta G^\circ_{37}$	T_m
	(kcal/mol)	(kcal/mol)	(eu)	(kcal/mol)	(°C)	(kcal/mol)	(eu)	(kcal/mol)	(°C)
	-0.42	55.6±1.2	154.1±3.8	7.78±0.02	44.1	47.9±3.8	129.7±12.4	7.68±0.04	44.6
GG**CGAA**GGCU[b]	-1.73	58.1±7.6	160.9±23.8	8.20±0.33	46.2	56.0±4.1	154.1±12.9	8.19±0.20	46.5
PCC**AAAG**CCG		73.6±1.4	205.3±4.4	9.93±0.06	52.0	67.5±5.1	186.2±16.0	9.72±0.19	52.4
GG**CGGA**GGCU	-1.72	79.1±3.1	222.3±9.7	10.13±0.14	51.8	79.8±6.4	224.6±19.7	10.16±0.29	51.8
PCC**AAAG**CCG		87.2±1.8	243.0±5.6	11.85±0.10	57.2	89.3±4.9	249.2±14.8	11.98±0.29	57.2
GC**CCGA**GCG[b,c]	-1.59	68.8±3.1	192.2±9.6	9.20±0.12	53.7	75.8±2.7	213.7±8.4	9.47±0.13	53.3
GCG**A**GCCCG		81.6±9.2	223.1±27.4	12.38±0.69	64.8	85.7±3.4	235.5±10.3	12.67±0.23	64.6
GAG**CGAAC**GAC	-1.56	77.1±4.3	235.2±14.3	4.14±0.16	27.7	72.5±7.5	219.9±24.7	4.28±0.25	27.6
CUC**AAGAA**CUG		75.1±4.5	223.8±14.7	5.70±0.10	33.6	72.0±5.9	213.8±19.0	5.73±0.10	33.6
GC**CGAA**GCCP[b,c]	-1.19	60.6±2.4	177.4±7.6	5.61±0.03	36.6	53.3±8.5	153.3±27.5	5.72±0.11	37.2
PCG**AAGC**CG		70.5±1.2	201.4±3.9	7.99±0.03	47.5	75.1±3.0	216.0±9.5	8.12±0.13	47.4
		(44.7±1.9)	(129.0±6.4)	(4.66±0.06)	(30.1)	(48.7±6.5)	(142.1±21.4)	(4.62±0.10)	(30.4)
GCG**AAA**CCGA[d]	-1.19	68.0±4.5	194.4±14.3	7.66±0.10	42.2	74.3±6.4	214.6±21.0	7.70±0.13	42.0
UCGC**AAC**GGC		71.5±3.3	202.1±10.5	8.85±0.10	47.4	75.4±6.6	214.2±20.9	8.96±0.18	47.3
GAGC**CGA**CGAC[e]	-1.14	83.1±3.2	235.9±10.0	9.99±0.12	50.5	82.7±4.0	234.5±12.6	9.97±0.15	50.5
CUCG**AAA**GCUG		93.1±7.5	264.4±23.2	11.13±0.34	53.1	92.3±5.9	261.9±18.8	11.08±0.14	53.1
	0.21	92.3±4.6	262.7±14.4	10.77±0.20	51.9	93.7±3.2	267.3±10.1	10.82±0.15	51.9
CGAC**CGA**GCAG[e,f]	-1.12	79.5±3.8	222.3±11.7	10.56±0.17	53.6	86.8±6.0	244.6±18.6	10.88±0.30	53.4
GCUG**AAG**CGUC		92.3±5.3	260.0±16.3	11.68±0.28	55.3	96.5±3.6	272.8±11.3	11.89±0.18	55.2
GC**CGAA**GCG[b,c]	-1.07	64.4±5.6	185.7±17.9	6.81±0.14	42.6	61.7±7.5	176.8±24.0	6.86±0.17	43.0
GCG**AAG**CCG		75.4±3.9	214.2±12.0	8.95±0.15	51.1	76.4±5.4	217.2±16.8	9.02±0.19	51.2
CGC**AAA**GGC[e]	-1.02	61.1±4.3	174.5±13.9	6.96±0.09	39.2	63.5±10.1	181.9±32.6	7.04±0.16	39.5
GCG**AAC**CCG		82.1±3.9	238.9±12.5	7.98±0.06	42.6	76.4±11.8	220.7±37.9	7.93±0.09	42.8
GAGC**CGA**CGAC[e]	-0.97	73.3±2.8	200.8±8.4	11.04±0.15	57.3	69.4±2.3	188.7±6.6	10.82±0.20	57.4
CUCG**AAG**GCUG		86.1±2.6	239.0±7.8	12.01±0.15	58.1	80.7±1.5	222.6±4.6	11.69±0.13	58.2
GAGCU**GC**CGAC	-0.92	95.0±4.3	270.4±13.1	11.12±0.19	52.7	88.6±5.1	250.6±15.8	10.83±0.22	52.8
CUCGA**GA**GCUG		101.2±3.3	287.5±10.2	12.04±0.17	54.8	97.4±4.0	276.0±12.0	11.84±0.25	54.9
GG**CGA_**GGCU	-0.89	80.3±2.7	229.3±8.4	9.12±0.07	47.4	74.4±4.0	210.8±12.6	8.97±0.15	47.5
PCC**AAAG**CCG		79.2±2.6	223.2±7.9	10.01±0.10	51.2	76.3±4.4	214.2±13.3	9.87±0.24	51.2
CGC**GAA**GGC[e]	-0.84	80.7±5.4	232.3±17.0	8.65±0.13	45.4	71.6±5.4	203.5±17.3	8.52±0.09	45.8
GCG**AAC**CCG		87.8±6.1	252.4±19.2	9.49±0.20	47.8	78.8±6.2	224.3±19.7	9.25±0.13	48.1
CGC**AAA**GGC[e]	-0.82	46.7±2.7	131.7±9.0	5.87±0.08	32.7	48.4±10.7	137.0±35.2	5.88±0.26	32.9
GCG**ACC**CCG		61.8±4.0	177.7±12.9	6.69±0.07	37.8	60.2±9.6	172.4±31.2	6.78±0.12	38.3
GGU**GAA**GGCU	-0.80	86.2±2.4	243.6±7.4	10.68±0.10	52.7	88.2±1.7	249.6±5.3	10.78±0.11	52.7
PCCG**AAC**CCG		90.4±2.6	254.4±7.8	11.48±0.13	55.0	89.7±5.0	252.2±15.2	11.46±0.27	55.1
GAGC**CU**CGAC	-0.79	94.3±1.9	270.5±6.0	10.46±0.08	50.5	89.1±3.6	254.2±11.1	10.27±0.11	50.6
CUCG**CU**GCUG		94.8±3.1	269.3±9.5	11.25±0.14	53.2	92.3±4.1	261.7±12.9	11.15±0.17	53.3
GCA**AGAA**GGC[g]	-0.73	81.5±1.8	232.8±5.5	9.31±0.05	47.9	84.6±3.9	242.4±12.0	9.43±0.15	48.0
UCGU**CAGC**CCG		81.1±2.5	229.1±7.8	10.04±0.10	51.0	86.1±5.6	244.5±17.3	10.23±0.22	50.9
	-0.21	72.2±6.0	208.5±19.1	7.50±0.14	41.2	70.6±4.2	203.6±13.2	7.48±0.19	41.2
CGAC**CGAG**CCAG[h]	-0.73	69.0±6.5	197.6±20.6	7.71±0.18	42.4	69.1±3.2	197.9±10.5	7.70±0.19	42.4
GCUG**AG**_GGUC		77.3±11.0	221.9±34.6	8.44±0.44	44.8	77.4±5.9	222.5±19.0	8.42±0.26	44.8
GCA**AGAA**GGC[g]	-0.73	66.3±4.2	190.8±13.6	7.09±0.08	39.6	61.6±5.8	175.6±18.7	7.11±0.10	39.9
UCGU**CUG**CCCG		76.0±2.6	219.7±8.2	7.82±0.04	42.3	73.0±4.4	210.1±14.1	7.80±0.05	42.5

Table 1. Continued.

Sequences	Linear fit of T_M^{-1} vs $\ln(C_T/a)$ (eq 1)					Average of two-state melt curve fits			
	$\Delta\Delta G^\circ_{37,\,pH}$	$-\Delta H^\circ_{37}$	$-\Delta S^\circ$	$-\Delta G^\circ_{37}$	T_m	$-\Delta H^\circ$	$-\Delta S^\circ$	$-\Delta G^\circ_{37}$	T_m
	(kcal/mol)	(kcal/mol)	(eu)	(kcal/mol)	(°C)	(kcal/mol)	(eu)	(kcal/mol)	(°C)
CGC**GAA**GGC[e]	-0.69	58.9±4.9	165.8±15.5	7.52±0.14	42.3	56.5±9.4	157.7±30.0	7.58±0.10	42.9
GCG**ACC**CCG		79.5±4.7	229.9±15.0	8.21±0.10	43.7	68.9±9.1	196.1±29.3	8.10±0.11	44.2
CGC**AUA**GGC	-0.64	54.5±3.5	155.9±11.6	6.12±0.08	34.7	56.3±5.1	161.6±16.4	6.13±0.15	34.8
GCG**AAC**CCG		55.6±3.8	157.4±12.3	6.76±0.08	38.3	61.0±5.8	175.0±18.3	6.78±0.20	38.3
	-0.25	85.4±5.7	245.5±17.8	9.28±0.19	47.3	76.2±3.7	216.7±11.7	9.02±0.18	47.5
GGU**CAA**GGCU	-0.64	77.9±2.2	220.5±6.9	9.53±0.08	49.4	78.0±1.5	220.8±4.9	9.53±0.05	49.4
PCCA**AAG**CCG		80.2±2.1	225.9±6.5	10.17±0.09	51.7	81.7±3.8	230.3±11.5	10.24±0.20	51.8
CGAC**CGA**GCAG[e]	-0.63	85.2±1.9	242.9±5.9	9.91±0.07	49.8	88.9±3.6	254.1±11.2	10.05±0.15	49.8
GCUG**AAA**CGUC		87.5±3.9	248.2±12.0	10.54±0.17	51.9	94.3±5.3	269.2±16.3	10.81±0.26	51.7
GAGC**CGA**CGAC[e]	-0.31	69.8±2.2	195.6±7.0	9.16±0.07	49.2	61.1±3.1	168.3±9.7	8.87±0.19	49.3
CUCG**AGA**GCUG		62.5±6.4	170.9±19.8	9.47±0.30	52.3	66.9±5.2	184.6±15.9	9.63±0.30	52.1
	-0.35	83.9±1.5	241.9±4.7	8.88±0.04	45.9	76.7±3.5	219.1±11.3	8.72±0.08	46.1
GGU**GUA**GGCU	-0.30	78.4±2.3	223.0±7.2	9.23±0.07	48.1	73.9±4.1	209.1±12.7	9.10±0.17	48.2
PCCG**AAC**CCG		79.3±2.3	225.0±7.1	9.53±0.08	49.2	73.4±2.6	206.6±8.2	9.35±0.09	49.4
	-0.23	93.8±2.7	268.7±8.4	10.50±0.10	50.7	83.6±2.5	237.0±7.8	10.12±0.14	50.9
GGU**GGA**GGCU	-0.29	87.9±2.3	248.7±7.2	10.73±0.10	52.6	80.6±1.8	226.3±5.7	10.43±0.10	52.8
PCCG**AAC**CCG		89.4±3.6	252.7±11.2	11.02±0.16	53.4	84.9±5.9	239.0±18.2	10.82±0.27	53.5
GGU**AGA**GGCU	-0.25	78.5±0.9	224.1±2.8	8.98±0.03	47.0	68.2±2.5	192.0±7.5	8.66±0.16	47.0
PCCG**AAC**CCG		82.6±2.2	236.4±7.1	9.23±0.06	47.5	72.0±3.9	203.4±12.2	8.93±0.19	47.7
	-0.43	90.9±4.5	261.5±14.1	9.74±0.14	48.3	83.8±4.1	239.4±12.9	9.51±0.16	48.4
GGU**CAA**GGCU[b]	-0.19	87.1±2.0	248.0±6.1	10.17±0.07	50.5	78.9±3.3	222.6±10.1	9.85±0.16	50.6
PCCG**AAG**CCG		84.9±1.9	240.5±5.8	10.36±0.08	51.6	80.6±2.7	226.9±8.4	10.19±0.14	51.8
GGU**GA__**GGCU	-0.16	73.7±2.1	208.7±6.6	8.98±0.06	47.6	66.6±2.2	186.6±6.9	8.77±0.14	47.8
PCCG**AAC**CCG		79.6±3.1	227.1±9.7	9.14±0.08	47.5	70.1±6.5	197.3±20.3	8.85±0.27	47.6
GAGC**CGA**CGAC[e]	-0.09	79.7±4.9	227.2±15.3	9.25±0.16	48.0	71.3±3.2	200.9±10.4	9.02±0.07	48.2
CUCG**AUA**GCUG		76.3±2.7	215.9±8.4	9.34±0.09	48.9	66.8±4.3	186.1±13.5	9.06±0.12	49.2
CGC**AGA**GGC	-0.04	56.5±1.7	160.2±5.4	6.75±0.02	38.2	55.1±3.7	155.8±12.0	6.77±0.06	38.4
GCG**AAC**CCG		55.8±1.0	157.9±3.1	6.79±0.01	38.5	54.9±2.0	155.0±6.4	6.81±0.05	38.6
GCGA**GAC__**CCG[i]	0.00	69.1±1.4	193.8±4.5	8.97±0.04	48.4	72.3±4.6	203.8±14.4	9.03±0.17	48.1
CGCU**AGGA**GGC		70.8±2.9	199.4±9.2	8.97±0.09	48.1	67.8±5.1	190.0±15.8	8.86±0.21	48.0
GAGC**AAA**CGAC	0.01	78.2±3.1	224.9±9.7	8.41±0.06	44.7	76.8±5.7	220.5±18.1	8.40±0.13	44.7
CUCG**CAA**GCUG		79.3±3.3	228.6±10.3	8.40±0.06	44.5	77.5±5.6	223.0±17.8	8.36±0.17	44.5
CGAC**GCA**GCAG[e]	0.06	75.9±1.6	215.0±4.9	9.23±0.05	48.4	69.8±4.6	196.0±14.3	9.02±0.18	48.5
GCUG**AAG**CGUC		76.4±2.9	216.7±9.2	9.17±0.09	48.1	70.1±4.5	196.9±14.3	8.99±0.15	48.3
GAGC**AAG**CGAC	0.08	68.9±2.8	195.7±8.7	8.20±0.05	44.7	71.3±3.6	203.3±11.2	8.23±0.13	44.6
CUCG**CAA**GCUG		73.2±2.0	209.7±6.3	8.12±0.03	43.9	70.3±3.2	200.6±10.4	8.10±0.05	44.1
CGAC**GCA**GCAG[e]	0.09	76.1±3.2	218.2±10.0	8.46±0.07	45.0	76.1±7.2	218.3±22.6	8.42±0.20	44.9
GCUG**AAA**CGUC		79.4±3.7	229.0±11.8	8.37±0.08	44.4	76.4±6.2	219.7±19.4	8.27±0.19	44.2
GAGC**C**UCGAC	0.11	76.6±1.9	219.2±5.9	8.62±0.06	49.4	71.5±2.5	203.4±7.8	8.47±0.08	49.6
CAGCUC**C**GAG		69.2±1.9	195.8±5.9	8.51±0.06	50.2	73.9±3.5	210.5±11.0	8.64±0.13	50.0
GAGC**UGC**CGAC[e]	0.13	86.8±4.8	247.6±14.7	10.05±0.19	50.1	88.4±4.6	252.5±14.5	10.12±0.15	50.1
CUCG**UAA**GCUG		89.0±3.7	254.8±11.4	9.92±0.13	49.3	83.3±6.2	237.0±19.3	9.74±0.22	49.4
GAGC**CU**CGAC[j]	0.16	90.7±1.8	260.2±5.7	9.97±0.06	49.2	83.5±2.9	237.7±9.1	9.73±0.15	49.4
CUCG**UU**GCUG		85.7±2.0	244.6±6.4	9.81±0.07	49.4	80.7±2.5	229.2±7.6	9.64±0.13	49.4
GAGC**UGC**CGAC[e]	0.17	81.0±1.9	231.3±5.8	9.24±0.05	47.7	77.9±5.5	221.8±17.4	9.14±0.14	47.8
CUCG**UAU**GCUG		82.2±1.8	235.7±5.8	9.07±0.05	46.9	75.0±3.4	213.3±10.6	8.87±0.17	47.0

Table 1. Continued.

Sequences	Linear fit of T_M^{-1} vs $\ln(C_T/a)$ (eq 1)					Average of two-state melt curve fits			
	$\Delta\Delta G°_{37,\,pH}$	$-\Delta H°_{37}$	$-\Delta S°$	$-\Delta G°_{37}$	T_m	$-\Delta H°$	$-\Delta S°$	$-\Delta G°_{37}$	T_m
	(kcal/mol)	(kcal/mol)	(eu)	(kcal/mol)	(°C)	(kcal/mol)	(eu)	(kcal/mol)	(°C)
	-0.40	90.4±4.8	259.1±14.9	10.07±0.16	49.6	78.6±3.8	222.3±11.8	9.66±0.23	49.9
GCA**AGAA**GGC[b,g]	0.19	88.5±1.0	251.6±3.2	10.47±0.04	51.4	80.9±5.0	228.2±15.3	10.15±0.27	51.5
UCGU**CAGG**CCG		88.5±1.4	252.1±4.2	10.28±0.05	50.7	75.4±3.6	211.5±11.6	9.78±0.13	51.0
GAGC**UU**CGAC[k]	0.21	99.7±1.3	285.3±4.1	11.20±0.06	52.2	97.1±7.3	277.1±22.5	11.13±0.36	52.4
CUCG**CU**GCUG		97.4±3.5	278.6±10.8	10.99±0.15	51.9	93.8±4.3	267.4±13.3	10.84±0.17	51.9
GAGC**CGA**CGAC[e]	0.41	79.6±3.1	226.2±9.6	9.42±0.08	48.7	71.5±1.2	200.7±3.9	9.21±0.07	49.1
CUCG**UAA**GCUG		65.0±2.7	180.4±8.4	9.01±0.08	49.3	56.7±3.4	154.3±10.9	8.79±0.14	49.9
CGCA**U**AGGC	0.42	79.2±2.1	222.6±6.5	10.20±0.08	52.1	86.2±2.8	244.2±8.4	10.49±0.19	51.9
GCGU**C**UCCG		75.8±3.6	212.9±11.3	9.78±0.13	50.9	76.7±6.2	215.6±19.2	9.86±0.24	51.1
CGC**UUU**GGC[e]	0.57	55.3±3.3	153.0±10.5	7.80±0.06	44.3	62.6±4.3	176.5±13.6	7.88±0.11	43.8
GCG**UCU**CCG		65.5±4.8	187.9±15.6	7.23±0.10	40.4	60.1±5.8	170.3±19.0	7.26±0.12	40.8

[a] For each duplex, the values from bottom to top are measured at pH 5.5, 7, and 8, respectively. Sequences are ordered from the most negative to the most positive values of $\Delta\Delta G°_{37,pH} = \Delta G°_{37,pH5.5} - \Delta G°_{37,pH7}$, unless noted in footnote c. T_m values were calculated from eq 1 at $C_T = 0.1$ mM. Data in parentheses were measured in NMR buffer with 80 mM NaCl at pH 7. [b] Imino proton NMR spectra were measured (Figure 2). [c] $\Delta\Delta G°_{37,pH}$ is per CA pair. [d] Loop sequence from a J4/5 loop of a group I intron (36). [e] Data at pH 7 are from ref 19. [f] Loop sequence from the substrate loop of a VS ribozyme (8, 9). [g] Loop sequence derived from the loop A of hairpin ribozyme (3). [h] Loop sequence from a leadzyme (1, 65−67). [i] Loop sequence from the Alu domain of human SRP RNA (71). [j] The pH-independent thermodynamics is consistent with the NMR structure without the formation of the C^+U pair (61). [k] The pH-independent thermodynamics is consistent with the NMR structure without the formation of the UC^+ pair (62).

the most negative to the most positive values of $\Delta\Delta G°_{37,pH}$, which is defined as

$$\Delta\Delta G°_{37,\,pH} = \Delta G°_{37,\,pH5.5} - \Delta G°_{37,\,pH7} \qquad (2)$$

For several duplexes with two loop-terminal CA pairs, $\Delta\Delta G°_{37,pH}$ is half the value given by eq 2 (see the footnotes of Tables 1 and 2). Measured thermodynamic parameters for the formation of the internal loops (Table 2 and Table S2 in the Supporting Information) are calculated according to the following equation (30):

$$\Delta G°_{37,\,loop} = \Delta G°_{37(\text{duplex with loop})} - \Delta G°_{37(\text{duplex without loop})} +$$
$$\Delta G°_{37(\text{interrupted base stack})} \qquad (3a)$$

For example,

$$\Delta G°_{37}\frac{\text{G CGAA G}}{\text{C }\overline{\text{AAAG}}\text{ C}} = \Delta G°_{37}\frac{\text{GG CGAA GGCU}}{\text{PCC }\overline{\text{AAAG}}\text{ CCG}} -$$
$$\Delta G°_{37}\frac{\text{GGGGCU}}{\text{PCCCCG}} + \Delta G°_{37}\frac{\text{GG}}{\text{CC}} \qquad (3b)$$

Here, $\Delta G°_{37}\frac{\text{GG CGAA GGCU}}{\text{PCC AAAG CCG}}$ is the measured value of the duplex containing the internal loop (Table 1); $\Delta G°_{37}\frac{\text{GGGGCU}}{\text{PCCCCG}}$ is calculated from the measured value of the duplex $\frac{\text{GGUGGCU}}{\text{PCCGCCG}}$ (20) by a nearest neighbor model (25, 31) ($\Delta G°_{37}\frac{\text{GGGGCU}}{\text{PCCCCG}} = \Delta G°_{37}\frac{\text{GGUGGCU}}{\text{PCCGCCG}} - \Delta G°_{37}\frac{\text{GU}}{\text{CG}} - \Delta G°_{37}\frac{\text{UG}}{\text{GC}} + \Delta G°_{37}\frac{\text{GG}}{\text{CC}}$); and $\Delta G°_{37}\frac{\text{GG}}{\text{CC}}$ is the free-energy increment for the nearest neighbor base stack interaction interrupted by the internal loop. Values for $\Delta H°_{loop}$ and $\Delta S°_{loop}$ are calculated similarly. Whenever available, measured thermodynamic values of canonical stems are used for the calculation of measured thermodynamic parameters of loops. All of the thermodynamic parameters used in this calculation are derived from T_M^{-1} versus $\ln(C_T/a)$ plots (eq 1).

The thermodynamics of canonical stems is calculated for pH 7 and assumed to be independent of the pH between 5.5 and 8, as shown for other stems (32, 33). This is a reasonable assumption because the N1 of adenine and N3 of cytosine normally have pK_a values of 3.5 and 4.2, respectively, and the pK_a values shift further down in forming Watson−Crick pairs in canonical stems (1−3, 7). In addition, most of the duplexes do not form wobble CA^+ or CC^+ pairs (panels b and i of Figure 1) and do not show a pH effect, consistent with the assumption of pH-independent thermodynamics in the absence of CA^+ or CC^+ pairs (Table 1 and Table S1 in the Supporting Information).

Thermodynamic Model Including Stabilization Effects of CA and CA^+ Pairs in Medium-Size RNA Internal Loops. Measured free energies of RNA internal loops with 6−10 nucleotides, $\Delta G°_{37,\,loop}$, reported here and previously (16, 19, 20) for 1 M NaCl at pH 7 and 37 °C were combined for linear regression to the equation

$$\Delta G°_{predicted} = \Delta G°_{\text{loop initiation}}(n) + m1\Delta G°_{\text{AU/GU penalty}}$$
$$+ |n1 - n2|\Delta G°_{\text{asym}} + m2\Delta G°_{\text{UU bonus}}$$
$$+ m3\Delta G°_{\text{5'YA/3'RG bonus}} + m4\Delta G°_{\text{GA bonus}}$$
$$+ \Delta G°_{\text{middle GA bonus (3×3 loop)}}$$
$$+ \Delta G°_{\text{5'GU/3'AN penalty (3×3 loop)}}$$
$$+ \Delta G°_{2\times(\text{5'GA/3'CG}) \text{ bonus (3×3 loop)}}$$
$$+ m5\Delta G°_{\text{2GA bonus}} + m6\Delta G°_{\text{3GA bonus}}$$
$$+ m7\Delta G°_{\text{5'UG/3'GA bonus}}$$
$$+ m8\Delta G°_{\text{5'CR/3'AA bonus}} \qquad (4)$$

Here, $n1$ and $n2$ are the number of nucleotides on each side of the loop; m1−m8 can be 0, 1, or 2; and the definitions of free-energy parameters are given in Table 3. Multiple linear regression on 168

Table 2: Measured and Predicted Thermodynamic Parameters for RNA Internal Loop Formation in 1 M NaCl[a]

Sequences	$\Delta\Delta G^{\circ}_{37,\,pH}$ (kcal/mol)	$\Delta G^{\circ}_{predicted}$ (kcal/mol)	$\Delta G^{\circ}_{37,loop}$ (kcal/mol)	ΔH°_{loop} (kcal/mol)	ΔS°_{loop} (cal/mol·K)
	-0.42		0.35±0.59	1.8±9.6	4.9±29.5
GG**CGAA**GGCU[b]	-1.73	0.13	-0.07±0.68	-0.7±12.2	-1.9±37.7
PCC**AAAG**CCG			-1.80±0.60	-16.2±9.6	-46.3±29.6
GG**CGGA**GGCU	-1.72	-2.23	-2.00±0.61	-21.7±10.0	-63.3±30.8
PCC**AAAG**CCG			-3.72±0.60	-29.8±9.7	-84.0±29.8
GC**CCGA**GCG[b]	-1.59	0.50	-0.56±0.47	-7.9±8.3	-23.6±25.0
GCG**AGC**CCG			-2.15±0.83	-14.3±12.0	-39.1±35.8
GAG**CGAAC**GAC	-1.56	0.72	0.79±0.25	-33.9±17.2	-112.1±15.2
CUC**AAGAAC**UG			-0.77±0.22	-31.9±17.3	-100.7±15.6
GC**CGAA**GCCP[b,c]	-1.19	-1.16	-1.76±0.20	-32.5±4.7	-99.3±14.5
PCC**GAAG**CCG			-4.14±0.20	-42.4±4.2	-123.3±12.9
GCG**AAA**CCGA[d]	-1.19	1.08	2.03±0.45	-13.3±8.4	-49.7±25.4
UCGC**AAC**GGC			0.84±0.45	-16.8±7.8	-57.4±23.5
GAGC**CGA**CGAC[e]	-1.14	0.19	0.69±0.58	-16.4±10.2	-54.6±30.8
CUCG**AAA**GCUG			-0.45±0.66	-26.4±12.2	-83.1±37.3
	0.21		-0.81±0.54	-32.4±9.8	-102.0±29.5
CGAC**CGA**GCAG[e,f]	-1.12	-1.02	-0.60±0.53	-19.6±9.4	-61.6±28.3
GCUG**AAG**CGUC			-1.72±0.57	-32.4±10.1	-99.3±30.5
GC**CGAA**GCG[b,c]	-1.07	-1.16	-1.16±0.24	-27.3±6.9	-84.6±21.7
GCG**AAG**CCG			-3.30±0.25	-38.3±5.6	-113.1±17.2
CGC**AAA**GGC[e]	-1.02	1.08	1.56±0.45	-12.1±8.5	-43.8±25.9
GCG**AAC**CCG			0.54±0.45	-33.1±8.3	-108.2±25.2
GAGC**CGA**CGAC[e]	-0.97	-1.02	-0.36±0.58	-6.6±10.1	-19.5±30.3
CUCG**AAG**GCUG			-1.33±0.58	-19.4±10.0	-57.7±30.2
GAGCU**GC**CGAC	-0.92	0.32	1.67±0.55	-15.8±10.4	-56.6±31.7
CUCGA**GA**GCUG			0.75±0.54	-22.0±10.0	-73.7±30.6
GG**CGA** GGCU	-0.89	-0.49	-0.99±0.60	-22.9±9.9	-70.3±30.4
PCC**AAAG**CCG			-1.88±0.60	-21.8±9.9	-64.2±30.3
CGC**GAA**GGC[e]	-0.84	0.14	-0.13±0.46	-31.7±9.1	-101.6±27.7
GCG**AAC**CCG			-0.97±0.49	-38.8±9.6	-121.7±29.1
CGC**AAA**GGC[e]	-0.82	2.15	2.65±0.45	2.4±7.8	-1.0±23.6
GCG**ACC**CCG			1.83±0.45	-12.8±8.4	-47.0±25.4
GGU**GAA**GGCU	-0.80	-0.10	-0.04±0.54	-16.2±9.4	-52.1±29.4
PCC**AAC**CCG			-0.84±0.55	-20.4±9.5	-62.9±29.5
GAGC**CU**CGAC	-0.79	0.57	0.22±0.57	-27.6±9.8	-89.2±29.7
CUCG**CU**GCUG			-0.57±0.58	-28.1±10.2	-88.0±30.6
GCA**AGAA**GGC[g]	-0.73	0.52	-0.09±0.37	-27.2±6.8	-87.6±20.4
UCGU**CAGC**CCG			-0.82±0.38	-26.8±7.0	-83.9±21.1
	-0.21		2.81±0.41	-7.0±9.9	-31.6±30.6
CGAC**CGAG**CCAG[h]	-0.73	3.07	2.60±0.42	-3.8±10.2	-20.6±31.6
GCUG**AG** GGUC			1.87±0.58	-12.1±13.5	-45.0±42.1
GCA**AGAA**GGC[g]	-0.73	2.42	2.13±0.38	-12.0±7.9	-45.7±23.9
UCGU**CUGC**CCG			1.40±0.37	-21.7±7.1	-74.5±21.3
CGC**GAA**GGC[e]	-0.69	1.21	1.00±0.47	-9.9±8.8	-35.1±26.8
GCG**ACC**CCG			0.31±0.46	-30.5±8.7	-99.2±26.5
	$\Delta\Delta G^{\circ}_{37,\,pH}$	$\Delta G^{\circ}_{predicted}$	$\Delta G^{\circ}_{37,loop}$	ΔH°_{loop}	ΔS°_{loop}

Table 2. Continued.

Sequences	$\Delta\Delta G^\circ_{37,\,pH}$	$\Delta G^\circ_{predicted}$	$\Delta G^\circ_{37,loop}$	ΔH°_{loop}	ΔS°_{loop}
	(kcal/mol)	(kcal/mol)	(kcal/mol)	(kcal/mol)	(cal/mol·K)
CGC**AUA**GGC	-0.64	2.15	2.40±0.46	-5.5±8.2	-25.2±24.7
GCG**AAC**CCG			1.76±0.46	-6.5±8.3	-26.7±25.0
	-0.25		1.09±0.55	-16.6±9.2	-57.0±29.5
GGU**CAA**GGCU	-0.64	0.75	0.84±0.52	-9.1±8.0	-32.0±24.5
PCCA**AAG**CCG			0.20±0.52	-11.4±8.0	-37.4±24.4
CGAC**CGA**GCAG[e]	-0.63	0.19	0.05±0.51	-25.3±8.8	-82.2±26.5
GCUG**AAA**CGUC			-0.58±0.53	-27.6±9.5	-87.5±28.4
GAGC**CGA**CGAC[e]	-0.31	2.15	1.52±0.57	-3.1±9.9	-14.3±30.0
CUCG**AGA**GCUG			1.21±0.64	4.2±11.6	10.4±35.3
	-0.35		1.76±0.53	-13.9±9.4	-50.4±28.8
GGU**GUA**GGCU	-0.30	1.71	1.41±0.54	-8.4±9.5	-31.5±29.3
PCCG**AAC**CCG			1.11±0.54	-9.3±9.5	-33.5±29.3
	-0.23		0.14±0.54	-23.8±9.6	-77.2±28.6
GGU**GGA**GGCU	-0.29	-0.19	-0.09±0.54	-17.9±9.4	-57.2±29.3
PCCG**AAC**CCG			-0.38±0.56	-19.4±9.8	-61.2±30.5
GGU**AGA**GGCU	-0.25	1.87	1.66±0.53	-8.5±9.3	-32.6±28.5
PCCG**AAC**CCG			1.41±0.53	-12.6±9.5	-44.9±29.3
	-0.43		0.90±0.55	-20.9±10.3	-70.0±31.7
GGU**CAA**GGCU[b]	-0.19	0.75	0.47±0.54	-17.1±9.4	-56.5±29.1
PCCG**AAG**CCG			0.28±0.54	-14.9±9.4	-49.0±29.0
GGU**GA**_GGCU	-0.16	1.92	1.66±0.53	-3.7±9.5	-17.2±29.2
PCCG**AAC**CCG			1.50±0.54	-9.6±9.7	-35.6±30.0
GAGC**CGA**CGAC[e]	-0.09	2.15	1.43±0.58	-13.0±10.9	-45.9±32.9
CUCG**AUA**GCUG			1.34±0.57	-9.6±10.1	-34.6±30.3
CGC**AGA**GGC	-0.04	1.26	1.77±0.45	-7.5±7.6	-29.5±22.5
GCG**AAC**CCG			1.73±0.45	-6.8±7.4	-27.2±22.0
GCGA**GAC**_CCG[i]	0.00	1.19	0.69±0.39	-10.7±7.1	-36.8±21.7
CGCU**AGGA**GGC			0.69±0.38	-12.4±7.5	-42.4±23.1
GAGC**AAA**CGAC	0.01	2.15	2.27±0.56	-11.5±10.2	-43.6±30.7
CUCG**CAA**GCUG			2.28±0.56	-12.6±10.2	-47.3±30.9
CGAC**GCA**GCAG[e]	0.06	0.27	0.73±0.50	-16.0±8.8	-54.3±26.3
GCUG**AAG**CGUC			0.79±0.51	-16.5±9.1	-56.0±27.4
GAGC**AAG**CGAC	0.08	2.15	2.48±0.56	-2.2±10.1	-14.4±30.4
CUCG**CAA**GCUG			2.56±0.56	-6.5±9.9	-28.4±29.8
CGAC**GCA**GCAG[e]	0.09	1.21	1.50±0.51	-16.2±9.2	-57.5±27.7
GCUG**AAA**CGUC			1.59±0.51	-19.5±9.4	-68.3±28.4
GAGC**C**CUCGAC	0.11	0.50	0.55±0.33	-12.2±7.4	-40.7±23.1
CAGCUC**C**GAG			0.66±0.33	-4.8±7.4	-17.3±23.1
GAGC**UGC**CGAC[e]	0.13	0.65	0.63±0.59	-20.1±10.8	-66.3±32.6
CUCG**UAA**GCUG			0.76±0.58	-22.3±10.4	-73.5±31.3
GAGC**CU**CGAC[j]	0.16	0.42	0.71±0.56	-24.0±9.9	-78.9±29.7
CUCG**UU**GCUG			0.87±0.56	-19.0±9.9	-63.3±29.8
GAGC**UGC**CGAC[e]	0.17	1.54	1.44±0.56	-14.3±9.9	-50.0±29.7
CUCG**UAU**GCUG			1.61±0.56	-15.5±9.9	-54.4±29.7
	-0.40		-0.85±0.40	-36.1±8.1	-113.9±24.7
GCA**AGAA**GGC[b,g]	0.19	-0.55	-1.25±0.37	-34.2±6.6	-106.4±19.9
UCGU**CAGG**CCG			-1.06±0.37	-34.2±6.7	-106.9±20.1

Table 2. Continued.

Sequences	$\Delta\Delta G°_{37, pH}$ (kcal/mol)	$\Delta G°_{predicted}$ (kcal/mol)	$\Delta G°_{37,loop}$ (kcal/mol)	$\Delta H°_{loop}$ (kcal/mol)	$\Delta S°_{loop}$ (cal/mol·K)
GAGC**UU**CGAC[k] CUCG**CU**GCUG	0.21	0.42	-0.52±0.56 -0.31±0.58	-33.0±9.8 -30.7±10.3	-104.0±29.4 -97.3±31.1
GAGC**CGA**CGAC[e] CUCG**UAA**GCUG	0.41	1.26	1.26±0.57 1.67±0.57	-12.9±10.2 1.7±10.1	-44.9±30.7 0.9±30.3
CGCA**U**AGGC GCGU**C**UCCG	0.42	1.90	2.36±0.51 2.78±0.52	-8.1±8.8 -4.7±9.3	-33.9±26.3 -24.2±27.9
CGC**UUU**GGC[e] GCG**UCU**CCG	0.57	0.93	0.72±0.45 1.29±0.46	-6.3±8.1 -16.5±8.8	-22.3±24.2 -57.2±26.8

[a] Calculated from eq 3a and data in Table 1 unless noted otherwise. Experimental errors for $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ for the canonical stems are estimated as 4, 12, and 13.5%, respectively, according to ref *25*. These errors were propagated to estimate errors in loop thermodynamics. For each duplex, the values from the bottom to the top are measured at pH 5.5, 7, and 8, respectively. Sequences are ordered from the most negative to the most positive values of $\Delta\Delta G°_{37,pH} = \Delta G°_{37,pH5.5} - \Delta G°_{37,pH7}$, except for $(GCCCGAGCG)_2$ and those noted in footnote c, where $\Delta\Delta G°_{37,pH}$ is divided by 2. $\Delta G°_{predicted}$ values are calculated according to eq 4. Loops smaller than 3×3 nucleotides are predicted according to refs (*16, 29*, and *31*). [b] Imino proton NMR spectra were measured (Figure 2). [c] $\Delta G°_{5'CR/3'AA bonus}$ is applied twice to predict the free energy for loop formation. [d] Loop sequence from a J4/5 loop of a group I intron (*36*). [e] Data at pH 7 are from ref *19*. [f] Loop sequence from the substrate loop of a VS ribozyme (*8, 9*). [g] Loop sequence derived from the loop A of hairpin ribozyme (*8*). [h] Loop sequence from a leadzyme (*1, 65−67*). [i] Loop sequence from the Alu domain of human SRP RNA (*71*). [j] The pH-independent thermodynamics is consistent with the NMR structure without the formation of the C^+U pair (*61*). [k] The pH-independent thermodynamics is consistent with the NMR structure without the formation of the UC^+ pair (*62*).

loop free energies (Table S2 in the Supporting Information) gives the free-energy parameters listed in Table 3, with an $R^2 = 0.87$ and standard deviation of 0.55 kcal/mol, which averages less than 0.07 kcal/mol for each nucleotide contributing to $\Delta G°_{predicted}$ at 37 °C. The last term ($\Delta G°_{5'CR/3'AA bonus} = -1.07$ kcal/mol) in eq 4 represents the only difference with the equation derived previously (*16*). Without the last term, $R^2 = 0.82$ and the standard deviation is 0.65 kcal/mol. Aside from the last term, the parameters in Table 3 are essentially the same as previously derived (*16*). Note that the bonus and penalty parameters have negative and positive values, respectively.

Size-symmetric internal loops with 5′CR/3′AA nearest neighbors with the CA adjacent to a closing Watson−Crick pair, are further stabilized on average by 1.03 ± 0.32 kcal/mol when the pH is lowered from 7 to 5.5 (see Table S2 in the Supporting Information). Thus, a bonus, $\Delta G°_{5'CR/3'AA, pH bonus} = -1.03 \pm 0.32$ kcal/mol, is used to account for the pH stabilization at pH 5.5 compared to pH 7 (Table 3). At this stage, we do not apply $\Delta G°_{5'CR/3'AA, pH bonus}$ for the size-symmetric internal loops with 5′CR/3′AA nearest neighbors with the CA adjacent to a closing UG or GU pair. Loops with tandem CA pairs are also further stabilized when the pH is lowered from 7 to 5.5 (see Table S2 in the Supporting Information).

Dependent upon the sequence, the noncanonical pair adjacent to the GA pair in $_R^Y {CG \atop AA}$ or $_Y^R {CG \atop AA}$ can either stabilize or destabilize the medium-size internal loops, consistent with the previous thermodynamic model (e.g., $\Delta G°_{2GA bonus}$ and $\Delta G°_{5'GU/3'AN penalty (3 \times 3 loop)}$) (*16*). No significant stabilization at pH 7 and 5.5 is found for most of the other nearest neighbor combinations involving CA pairs, which is consistent with wobble CA^+ pairs (Figure 1b) not forming in different local sequence contexts in crystal and NMR structures ((*3, 34−40*)). Thermodynamics of several duplexes were measured at pH 8, and no significant differences were observed compared to those at pH 7.

Exchangeable Proton NMR Spectra at Different pH. Figure 2 shows 1D imino proton NMR spectra for selected sequences. The resonances observed are consistent with the expected canonical and sheared GA base pairs. Figure 3 shows 2D SNOESY spectra for $_{UCGU}^{GCA} {AGAA \atop CAGG} {GGC \atop CCG}$ and $_{PCCG}^{GC} {CGAA \atop AAGC} {GCCP \atop CG}$. The spectra contain the typical cross-peak patterns expected for the imino protons in the duplexes, although in some cases, definitive assignment is not made. In Figure 3a, four of the five imino protons between 12 and 14 ppm exhibit cross-peak patterns typical of a Watson−Crick GC pair (two strong cross-peaks to resonances that show a very strong cross-peak to each other and to a likely H5 resonance, as expected for the C amino protons of a GC pair). The fifth imino proton shows a strong cross-peak to a narrow resonance, as expected for a U imino proton close to the AH2 in a Watson−Crick AU pair. There is a very weak cross-peak between the imino protons of two of the GC pairs, which are assigned to G1 and G19. Three other resonances between 9.5 and 11 ppm have chemical shifts and cross-peaks typical of G imino protons in sheared GA pairs, including those observed in a duplex with the same sequence of three GA pairs (*20*). In Figure 3b, the two imino proton resonances between 13.0 and 13.5 ppm show typical GC pair characteristics. A cross-peak between the equivalent imino protons in the similar sequence, $_{GCG}^{GC} {CGAA \atop AAGC} {GCG \atop CG}$, confirms that these protons are in adjacent pairs (see Figure S2 in the Supporting Information).

The 1D imino proton spectra of several duplexes in Figure 2 reveal a similar peak near ~10.6 ppm that increases in intensity at lower pH. These peaks are likely due to adenine amino protons in CA^+ pairs, as observed in other cases of CA^+ pairs (*12*). The broad peak in Figure 3b at ~10.6 ppm assigned to the A6 amino group has a strong cross-peak to the other amino proton and a weak cross-peak to the G7 imino proton.

DISCUSSION

The pK_a of N1 nitrogen of adenine is about 3.5 and shifted by less than 0.3 pK unit when incorporated into unpaired single

Table 3: Free-Energy Parameters (kcal/mol) at 37 °C for Predicting 3 × 3 Nucleotide and Larger RNA Internal Loops[a]

$\Delta G^{\circ}_{\text{loop initiation}}(6)$	2.15 ± 0.10	Applied for loops with 6 nucleotides (n1 + n2 = 6).
$\Delta G^{\circ}_{\text{loop initiation}}(7)$	2.22 ± 0.19	Applied for loops with 7 nucleotides (n1 + n2 = 7).
$\Delta G^{\circ}_{\text{loop initiation}}(8)$	2.14 ± 0.16	Applied for loops with 8 nucleotides (n1 + n2 = 8).
$\Delta G^{\circ}_{\text{loop initiation}}(9)$	2.35 ± 0.26	Applied for loops with 9 nucleotides (n1 + n2 = 9).
$\Delta G^{\circ}_{\text{loop initiation}}(10)$	2.95 ± 0.28	Applied for loops with 10 nucleotides (n1 + n2 = 10).
$\Delta G^{\circ}_{\text{AU/GU penalty}}$	0.61 ± 0.10	Applied for each AU, UA, GU, or UG pair closing the loop.
$\Delta G^{\circ}_{\text{asym}}$	0.46 ± 0.07	Applied when the loop is size asymmetric, n1 ≠ n2.
$\Delta G^{\circ}_{\text{UU bonus}}$	-0.61 ± 0.12	Applied for loops with a UU loop-terminal pair.
$\Delta G^{\circ}_{\text{5'YA/3'RG bonus}}$	-0.72 ± 0.28	Applied for an AG loop-terminal pair adjacent to a YR canonical pair.
$\Delta G^{\circ}_{\text{GA bonus}}$	-0.94 ± 0.07	Applied for loops with a GA loop-terminal pair.
$\Delta G^{\circ}_{\text{middle GA bonus (3 × 3 loop)}}$	-0.89 ± 0.19	Applied for 3 × 3 loops with a middle pair of GA and at least one non-pyrimidine-pyrimidine loop-terminal pair unless a $\Delta G^{\circ}_{\text{2GA bonus}}$ or $\Delta G^{\circ}_{\text{3GA}}$ bonus has been applied.
$\Delta G^{\circ}_{\text{5'GU/3'AN penalty (3 × 3 loop)}}$	0.74 ± 0.20	Applied for 3 × 3 loops with a single loop-terminal GA pair that has a U 3' to the G of the GA pair. This penalty is also applied to 4 × 4 loops with a single GA pair adjacent to a loop-terminal CA pair with the motif 5'CGU/3'AAN.
$\Delta G^{\circ}_{\text{2×(5'GA/3'CG) bonus (3 × 3 loop)}}$	-1.11 ± 0.40	Applied for loops with two motifs of 5'GA/3'CG in 3 × 3 loops.
$\Delta G^{\circ}_{\text{2GA bonus}}$	-1.16 ± 0.14	Applied for loops with the motif 5'YGA/3'RAG, 5'RGA/3'YAG, 5'YGG/3'RAA, or 5'RGG/3'YAA (i.e., loops with the closing canonical pair 3' to the A of a GA pair) unless the motif is represented by a 3GA bonus or has asymmetry | n1 - n2 | > 1. This bonus is also applied for loops with the motif of 5'RGGA/3'YAAG or 5'GGAY/3'AAGR (i.e. $^{5'}_{3'}\frac{\text{GGA3}'}{\text{AAG5}'}$, not closed at least on one side with a YR canonical pair). This bonus is also applied for the motif 5'YCGA/3'RAAG, 5'RCGA/3'YAAG, 5'YCGG/3'RAAA, or 5'RCGG/3'YAAA in size-symmetric loops.
$\Delta G^{\circ}_{\text{3GA bonus}}$	-2.36 ± 0.14	Applied for loops with the motif of 5'YGGA/3'RAAG or 5'GGAR/3'AAGY (i.e. $^{5'}_{3'}\frac{\text{GGA3}'}{\text{AAG5}'}$, closed at least on one side with a YR canonical pair).
$\Delta G^{\circ}_{\text{5'UG/3'GA bonus}}$	-0.85 ± 0.14	Applied for 3 × 3 and larger loops with the motif of 5'UG/3'GA.
$\Delta G^{\circ}_{\text{5'CR/3'AA bonus}}$	-1.07 ± 0.13	Applied for 3 × 3 and larger size-symmetric internal loops with nearest neighbors of $\frac{\text{CG}}{\text{AA}}$ and/or $\frac{\text{CA}}{\text{AA}}$ when the CA pair is the loop-terminal pair.
$\Delta G^{\circ}_{\text{5'CR/3'AA, pH bonus}}$	-1.03 ± 0.32	Applied for stabilization from pH 7 to pH 5.5 in nearest neighbors of $\frac{\text{CG}}{\text{AA}}$ and/or $\frac{\text{CA}}{\text{AA}}$ when the CA pair is adjacent to a closing Watson-Crick pair in 3 × 3 and larger size-symmetric internal loops. This pH dependent bonus is also used to predict the free energy of loops with tandem CA pairs, 5'CA/3'AC and 5'CC/3'AA.

[a] These parameters are used to predict the free energy of the 3 × 3 nucleotide and larger internal loops in 1 M NaCl according to eq 4. Except for the new parameters, $\Delta G^{\circ}_{\text{5'CR/3'AA bonus}}$ and $\Delta G^{\circ}_{\text{5'CR/3'AA, pH bonus}}$, the parameters derived here are similar to those in ref *16*. YR is a canonical pair of CG, UA, or UG, with the pyrimidine Y on the 5′ side of the internal loop. In general, Y and R are defined, respectively, as U or C and A or G in the UG, UA, or CG pair.

strands (*7*). Small pK_a shifts were also observed for other nucleobases when incorporated in unpaired single strands (*7, 41*). When incorporated into double helices, however, the pK_a of A shifts down in Watson−Crick pairs but up by as much as 3 pK units in some noncanonical pairs (*1, 7*). For example, the pK_a of the A in a $^{\text{GAC}}_{\text{CUG}}$ sequence is ≤3.1, whereas the two A's in $^{\text{C}}_{\text{G}}\frac{\text{CGAG}}{\text{AG}}^{\text{C}}_{\text{G}}$ (loop sequence of a leadzyme) have pK_a values of 6.5 (shown in bold) and 4.3, respectively (*1*).

In addition to local context effects, pK_a values may also be shifted by global context. For example, the local dielectric constant in the middle of large structures, such as the ribosome and viral RNA encapsidated in virion, may differ from that in bulk water. Thus, it is important to know the possible effects of protonation on thermodynamic stability of RNA structures.

Dependent upon the sequence context and pH, a CA^+ pair can form with A protonated at the N1 position (Figure 1b). The CA^+ pair can form two hydrogen bonds and easily fit into an A-form helix. Thus, it has the potential to stabilize a helix. Protonation will also affect base stacking and other interactions, however, so that effects of protonation will be sequence-dependent.

The thermodynamic studies of short oligonucleotides at pH 7 and 5.5 provide insight into the sequence- and context-dependent stabilization effects of CA pairs. Many of the sequences studied were chosen because three-dimensional structures are available to allow stability−structure correlations (*1, 3, 8, 9, 12, 34−40, 42*).

Single CA^+ Pairs Stabilize Watson−Crick Stems. The CA^+ wobble pair is isosteric with a UG wobble pair (panels b and e of Figure 1) and can fit in an A- or B-form structure without large backbone distortion (see Figure S1 in the Supporting Information) (*12, 42, 43*). Consistent with formation of a CA^+ wobble pair, the measured loop free energy of $^{\text{GC}}_{\text{GCG}}\frac{\text{C}}{\text{A}}\frac{\text{CG}}{\text{GC}}\frac{\text{A}}{\text{C}}\frac{\text{GCG}}{\text{CG}}$ ($\Delta G^{\circ}_{\text{37,pH7,loop}}$ = −0.56 kcal/mol for each CA pair) is about 1 kcal/mol more stable than that predicted by a previous thermodynamic model (*29, 44*), without considering a stabilization effect for the CA pair (Table 2). In addition, a stabilization of $\Delta\Delta G^{\circ}_{\text{37,pH}}$ = −1.59 kcal/mol was found per $^{\text{C}}_{\text{G}}\frac{\text{C}}{\text{A}}\frac{\text{C}}{\text{G}}$ nearest neighbor combination at pH 5.5 compared to that at pH 7 (Table 2). The resonance at ~10.6 ppm in $^{\text{GC}}_{\text{GCG}}\frac{\text{C}}{\text{A}}\frac{\text{CG}}{\text{GC}}\frac{\text{A}}{\text{C}}\frac{\text{GCG}}{\text{CG}}$ (Figure 2a) is consistent with a previous assignment to A amino protons in a CA^+ pair (*12*). Thus, both UV thermal melting and
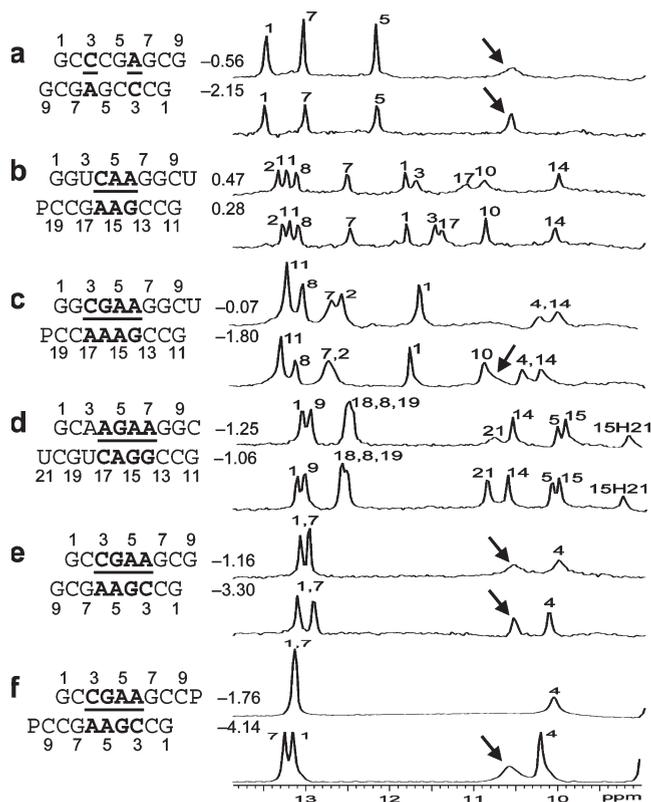
FIGURE 2: One-dimensional imino proton NMR spectra in 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM sodium EDTA at 0 °C unless otherwise noted at different pH values, with the top spectrum of each RNA sequence acquired at near pH 7 and the bottom spectrum at lower pH. Assignments are preliminary and largely based on assignments for similar sequences. Values between sequence and spectra are $\Delta G°_{37,loop}$ in kcal/mol measured in 1 M NaCl at pH 5.5 (bottom) and pH 7 (top). Resonances labeled with arrows are consistent with a previous assignment to the adenine amino protons in a CA$^+$ pair (*12*). No resonances were observed between 14 and 16 ppm. (a) $C_T =$ 0.5 mM, pH 6.9 and 5.4; (b) $C_T = 0.3$ mM, pH 6.9 and 5.0; (c) $C_T =$ 1.8 mM, pH 6.8 and 5.1; (d) $C_T = 0.5$ mM, pH 6.9 and 5.3 (see Figure 3a for 2D spectrum); (e) $C_T = 0.5$ mM, pH 6.9 and 5.9; and (f) $C_T =$ 1.5 mM, pH 6.6 and 5.1 (5 °C, see Figure 3b for the 2D spectrum).

NMR results are consistent with the formation of the hydrogen bonds in a wobble CA$^+$ pair (Figure 1b).

A similar pH effect on thermodynamics was found for single CA mismatches in DNA (*4, 7*). The A$^+$ imino proton was not observed by NMR (*4*), probably because of broadening by solvent exchange. The pK_a of the N1 of adenine in the DNA nearest neighbor combination, $^C_G {}^C_A {}^C_G$, is about 6.6, as measured with a pH profile of the chemical shifts of the N1 nitrogen (*45*).

Detailed understanding of the stabilization effect of CA or CA$^+$ wobble pairs within different Watson–Crick stems will provide insight into RNA structure and function. For example, a single CA mismatch has been shown to be preferred for efficient A to I editing by adenosine deaminases acting on RNA (ADAR) (*46*). Understanding the sequence-dependent thermodynamics of CA (*44*) and CI mismatches and the pH effect might facilitate better understanding of the editing specificity and mechanism (*46*).

$\frac{CR}{AA}$ *Nearest Neighbor with CA Adjacent to a Closing Canonical Pair Stabilizes 3 × 3 and Larger Size-Symmetric Internal Loops at pH 7.* When the CA is the first noncanonical (loop-terminal) pair, most of the size-symmetric internal loops with nearest neighbors of $\frac{CG}{AA}$ and/or $\frac{CA}{AA}$ are more stable than predicted by a recently proposed thermodynamic

model (*16, 19*). A bonus parameter, $\Delta G°_{5'CR/3'AA\,bonus} = -1.07 \pm 0.13$ kcal/mol at pH 7, is derived here for such nearest neighbor combinations with a loop-terminal CA pair followed by a GA or AA pair (Table 3). These nearest neighbor combinations occur in several internal loops within catalytic ribozymes, e.g., the VS ribozyme substrate loop (*8, 9*), $^C_G \frac{CGA}{AAG} {}^G_C$ ($\Delta G°_{37,pH7,loop} = -0.60$ kcal/mol), the loop A of hairpin ribozyme (*3*), $^A_U \frac{AGA}{CUG} {}^G_C$ ($\Delta G°_{37,pH7,loop} = 2.13$ kcal/mol), and the J4/5 loop of a group I intron (*36*), $^G_C \frac{AAA}{AAC} {}^C_G$ ($\Delta G°_{37,pH7,loop} = 2.03$ kcal/mol).

The thermodynamic stabilization is consistent with the geometric compatibility of $\frac{CG}{AA}$ and $\frac{CA}{AA}$ nearest neighbors if the CA pair is protonated and the purine–purine pair is sheared (panels f and g of Figure 1) (*3, 8, 9, 36*). Solution NMR reveals a protonated wobble CA$^+$ pair adjacent to a sheared GA pair ($^C_G \frac{CG}{AA}$, sequence in a hairpin ribozyme and VS ribozyme) (see Figure 4 and Figure S1a in the Supporting Information), and the pK_a of the A (in bold) is about 6.3, according to the pH profile of the chemical shifts of the C2 carbon in adenine (*3, 8, 9*). Consistently, the amino protons of A$^+$ (shown in bold) for the symmetric loop $^C_G \frac{CGAA}{AAGC} {}^G_C$ resonate at 10.6 ppm at neutral and lower pH (panels e and f of Figure 2 and Figure 3b). In addition, a wobble CA pair forms adjacent to a sheared AA pair (shown in bold) within the J4/5 loop, $^G_C \frac{CAA}{AAA} {}^C_G$, in the crystal structure of a group I intron (see Figure S1c in the Supporting Information) (*36*).

The noncanonical pair adjacent to a loop-terminal GA pair was previously found to either stabilize (e.g., $\Delta G°_{2GA\,bonus}$) or destabilize (e.g., $\Delta G°_{5'GU/3'AN\,penalty\,(3\times3\,loop)}$) the loop (*16, 19*). Here, the noncanonical pair adjacent to the GA pair in the nearest neighbor combinations $^Y_R \frac{CG}{AA}$ and $^R_Y \frac{CG}{AA}$ was also found to be stabilizing or destabilizing, although the CA but not GA pair is a loop-terminal pair. Thus, when the parameters in Table 3 were derived, the CA pair in the nearest neighbor combinations $^Y_R \frac{CG}{AA}$ and $^R_Y \frac{CG}{AA}$ was treated in a way similar to a canonical wobble UG pair; i.e., the thermodynamic effect of the GA pair was modeled as a first noncanonical (loop-terminal) pair. For example, a penalty of $\Delta G°_{5'GU/3'AN\,penalty\,(3\times3\,loop)} = 0.74$ kcal/mol was applied for $\frac{GCA}{UCGU} \frac{AGAA}{CUGC} \frac{GGC}{CCG}$ ($\Delta G°_{37,pH7,loop} = 2.13$ kcal/mol), although this parameter was proposed only for 3 × 3 nucleotide internal loops (*16, 19*). This is suggested by NMR data for this loop, which shows the formation of a stabilizing CA$^+$ wobble pair, isosteric to a canonical wobble UG pair and adjacent to a sheared GA pair, even at nearly neutral pH (*3*). Consistent with the penalty of $\Delta G°_{5'GU/3'AN\,penalty\,(3\times3\,loop)}$, the U is flipped out in an NMR structure of the $^A_U \frac{AGAA}{CUGC} {}^G_C$ loop, which is from a hairpin ribozyme (*3*).

Similarly, a bonus of $\Delta G°_{2GA\,bonus} = -1.16$ kcal/mol (Table 3) was applied for $\frac{GCA}{UCGU} \frac{AGAA}{CAGC} \frac{GGC}{CCG}$ ($\Delta G°_{37,pH7,loop} = -0.09$ kcal/mol), $\frac{GC}{PCCG} \frac{CGAA}{AAGC} \frac{GCCP}{CG}$ ($\Delta G°_{37,pH7,loop} = -1.76$ kcal/mol), and $\frac{GC}{GCG} \frac{CGAA}{AAGC} \frac{GCG}{CG}$ ($\Delta G°_{37,pH7,loop} = -1.16$ kcal/mol), although the two consecutive GA pairs are not adjacent to a canonical pair on either side. Note that, for the latter two sequences, the $\Delta G°_{5'CR/3'AA\,bonus}$ was applied twice.

CA pairs are not treated exactly as canonical UG closing pairs, however. Only one thermodynamic parameter, $\Delta G°_{5'CR/3'AA\,bonus}$, is applied for the nearest neighbor combinations of $^Y_R \frac{CG}{AA}$ or $^R_Y \frac{CG}{AA}$, but for $^{YU}_{RG} {}^G_A$ or $^{RU}_{YG} {}^G_A$, three parameters of $\Delta G°_{RG}^{YU}$ (or $\Delta G°_{YG}^{RU}$), $\Delta G°_{5'UG/3'GA\,bonus}$, and $\Delta G°_{GA}$ are applied (*16*).

$\frac{CR}{AA}$ *Nearest Neighbors with the CA Adjacent to a Watson–Crick Pair in Size-Symmetric Internal Loops Are More Stabilizing When the pH Is Lowered from 7 to 5.5.* If a wobble CA$^+$ pair is responsible for the extra stabilities
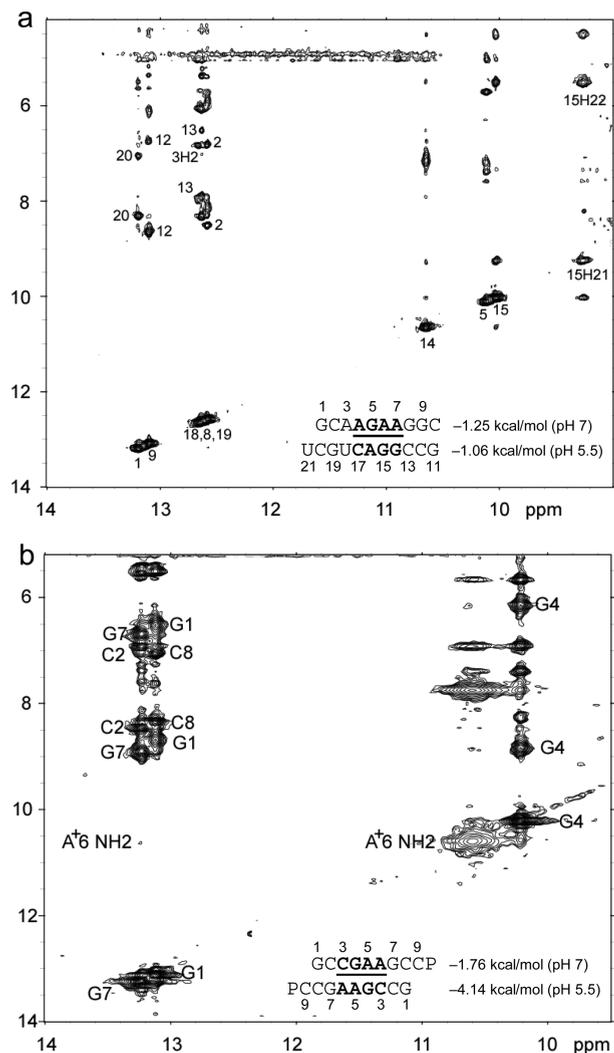
FIGURE 3: Two-dimensional exchangeable proton SNOESY spectra (150 ms mixing time in 80 mM NaCl, 10 mM sodium phosphate, and 0.5 mM sodium EDTA). The NOE cross-peaks of G imino protons to C amino and G amino protons are labeled with corresponding residues. Values beside the sequence are $\Delta G°_{37,loop}$ in kcal/mol measured in 1 M NaCl at pH 5.5 (bottom) and pH 7 (top). (a) $_{\text{UCGU}}^{\text{GCA}} {\frac{\text{AGAA}}{\text{CAGG}}} {_{\text{CCG}}^{\text{GGC}}}$ (C_T = 0.5 mM, pH 5.3, 0 °C, see Figure 2d for 1D spectrum). There is a very weak cross-peak of G1H1−G19H1 (not shown). The imino protons of G5, G14, and G15 have chemical shifts and cross-peaks typical of consecutive sheared GA pairs (*16, 20, 72*). The G15 amino protons resonate at 9.2 and 5.5 ppm, respectively, suggesting the formation of sheared GA pairs with G5 and G15 in the C2′-endo sugar pucker (*73, 74*). There is no indication of the formation of A^+C pair in this loop. (b) $_{\text{PCCG}}^{\text{GC}} {\frac{\text{CGAA}}{\text{AAGC}}} {_{\text{CG}}^{\text{GCCP}}}$ (C_T = 1.5 mM, pH 5.1, −5 °C, see Figure 2f for 1D spectrum). The cross-peak of G1H1−G7H1 is unresolved because of overlap but is observed in $_{\text{GCG}}^{\text{GC}} {\frac{\text{CGAA}}{\text{AAGC}}} {_{\text{CG}}^{\text{GCG}}}$ (see Figure S2 in the Supporting Information and Figure 2e for 1D spectrum). The broad peak at ~10.6 ppm is likely due to the amino protons of A^+6, which shows a strong cross-peak to the other amino proton and a weak cross-peak to the G7 imino proton. Adenine amino protons with similar chemical shift have been observed in other cases of CA^+ pairs (*12*). The G4 amino protons resonate at 8.8 and 6.2 ppm, respectively, suggesting the formation of sheared GA pairs with G4 in the C2′-endo sugar pucker (*73, 74*).

observed for $\frac{\text{CR}}{\text{AA}}$ nearest neighbors, then lowering the pH should further enhance stability because a larger fraction of A is protonated for the formation of CA^+ pairs. About 89 and 17% of adenine N1 residues are protonated at pH 5.5 and 7.0, respectively, with a pK_a of 6.3, as shown for the A (in bold) in $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$ (*3, 8, 9*). We observed an enhanced stabilization of 1.03 ± 0.32 kcal/mol on average per nearest neighbor $\frac{\text{CG}}{\text{AA}}$ or $\frac{\text{CA}}{\text{AA}}$ with the CA

adjacent to a Watson−Crick pair when lowering pH from 7 to 5.5, e.g., the VS ribozyme substrate loop (*8, 9*), $_{\text{G}}^{\text{C}} {\frac{\text{CGA}}{\text{AAG}}} {_{\text{C}}^{\text{G}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −1.12 kcal/mol), and the J4/5 loop of a group I intron (*36*), $_{\text{C}}^{\text{G}} {\frac{\text{AAA}}{\text{AAC}}} {_{\text{G}}^{\text{C}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −1.19 kcal/mol). Note that two CA^+ pairs can form in the symmetric loop $_{\text{G}}^{\text{C}} {\frac{\text{CGAA}}{\text{AAGC}}} {_{\text{C}}^{\text{G}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −1.19 and −1.07 kcal/mol per CA pair for the two duplexes measured).

Only one pH-dependent bonus parameter, $\Delta G°_{5'\text{CR}/3'\text{AA, pH bonus}}$ = −1.03 ± 0.32 kcal/mol, is derived here for $\frac{\text{CG}}{\text{AA}}$ or $\frac{\text{CA}}{\text{AA}}$ with CA adjacent to a Watson−Crick pair (Table 3). The sequence dependence is likely more complicated, however. For example, the thermodynamic stabilities of $_{\text{C}}^{\text{G}} {\frac{\text{CG}}{\text{AA}}}$ and $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$ may be significantly different. A stabilization effect of $\Delta\Delta G°_{37,\text{pH}}$ = −1.67 ± 0.10 kcal/mol per $_{\text{C}}^{\text{G}} {\frac{\text{CG}}{\text{AA}}}$ nearest neighbor combination was observed when lowering pH from 7 to 5.5, e.g., $_{\text{PCC}}^{\text{GG}} {\frac{\text{CGAA}}{\text{AAAG}}} {_{\text{CCG}}^{\text{GGCU}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −1.73 kcal/mol) and $_{\text{PCC}}^{\text{GG}} {\frac{\text{CGGA}}{\text{AAAG}}} {_{\text{CCG}}^{\text{GGCU}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −1.72 kcal/mol). This contrasts with the average of $\Delta\Delta G°_{37,\text{pH}}$ = −0.98 ± 0.21 kcal/mol for loops with a $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$ combination (see Table 2 and Table S2 in the Supporting Information). The pK_a of A N1 in the CA pair of $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$ (sequence found in a hairpin ribozyme and VS ribozyme) is about 6.3 with a wobble CA pair adjacent to a sheared GA pair (*3, 8, 9*). Presumably, the same noncanonical base pairs form in $_{\text{C}}^{\text{G}} {\frac{\text{CG}}{\text{AA}}}$, although the pK_a of A N1 in the CA pair is not known. Further detailed experimental (e.g., measurement of pK_a) and computational studies (*47, 48*) are needed to understand the different pH effect on the thermodynamics of $_{\text{C}}^{\text{G}} {\frac{\text{CG}}{\text{AA}}}$ and $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$.

No Significant Stabilizing Effect Was Observed for CA Pairs within Other Sequence Contexts in Size-Symmetric Internal Loops. No significant thermodynamic stabilization (i.e., free-energy stabilization of 1 kcal/mol or more) at either pH 7 or 5.5 was found for size-symmetric loops with the A of a potential AC pair 3′ to the adjacent Watson−Crick pair, e.g., $_{\text{CUCG}}^{\text{GAGC}} {\frac{\text{AAA}}{\text{CAA}}} {_{\text{GCUG}}^{\text{CGAC}}}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 2.27 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = 0.01 kcal/mol), $_{\text{CUCG}}^{\text{GAGC}} {\frac{\text{AAG}}{\text{CAA}}} {_{\text{GCUG}}^{\text{CGAC}}}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 2.48 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = 0.08 kcal/mol), $_{\text{CUCG}}^{\text{GAGC}} {\frac{\text{UGC}}{\text{UAA}}} {_{\text{GCUG}}^{\text{CGAC}}}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 0.63 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = 0.13 kcal/mol), and $_{\text{UCGU}}^{\text{GCA}} {\frac{\text{AGAA}}{\text{CAGG}}} {_{\text{CCG}}^{\text{GGC}}}$ ($\Delta G°_{37,\text{pH7,loop}}$ = −1.25 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = 0.19 kcal/mol). The pH stabilization for $_{\text{CUC}}^{\text{GAG}} {\frac{\text{CGAAC}}{\text{AAGAA}}} {_{\text{CUG}}^{\text{GAC}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −1.56 kcal/mol) and $_{\text{UCGU}}^{\text{GCA}} {\frac{\text{AGAA}}{\text{CAGC}}} {_{\text{CCG}}^{\text{GGC}}}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −0.73 kcal/mol) can be attributed to the $_{\text{C}}^{\text{G}} {\frac{\text{CG}}{\text{AA}}}$ and $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$ segments (see the discussion above for different pH stabilization observed for $_{\text{C}}^{\text{G}} {\frac{\text{CG}}{\text{AA}}}$ and $_{\text{G}}^{\text{C}} {\frac{\text{CG}}{\text{AA}}}$ when lowering pH from 7 to 5.5), respectively, with no contribution from the $_{\text{G}}^{\text{C}} {\frac{\text{AA}}{\text{CA}}}$ and $_{\text{G}}^{\text{A}} {\frac{\text{AG}}{\text{CA}}}$ segments.

The lack of extra stability when the A of an AC pair is 3′ of a Watson−Crick pair is probably general. For example, on the basis of NMR spectra of a 7 × 9 nucleotide loop B domain of a hairpin ribozyme, the apparent pK_a of the N1 position of the bold A in a $_{\text{G}}^{\text{C}} {\frac{\text{AG}}{\text{CA}}}$ segment is 5.4, and, at pH 6.8, the AC has a single hydrogen-bond, A N1−C amino pair (Figure 1c). The GA is a sheared pair (*37*). Here, the single hydrogen-bond (A N1−C amino) AC pair has A and C shifted to major and minor grooves, respectively, which is opposite to a wobble AC pair. A sheared GA pair has G and A shifted to major and minor grooves, respectively, which favors base stacking between the single hydrogen-bond (A N1−C amino) AC and sheared GA pairs (see Figure 4d and Figure S1b in the Supporting Information).

The enhanced stability of a CA pair with the C on the 3′ side of a Watson−Crick pair relative to one with the A on the 3′ side of a Watson−Crick pair may be related to stacking on the adjacent
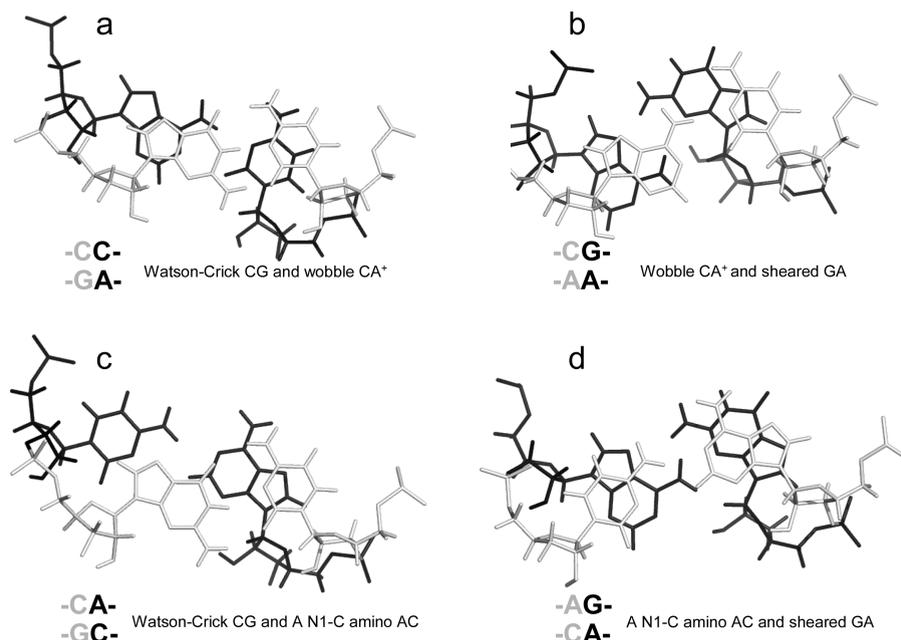
FIGURE 4: Base stacking and base pairing involving CA. Base pairs shown in gray lines are closer to the viewer. The Watson−Crick CG, wobble CA^+ (the proton from protonation is not shown), and sheared GA pairs shown in a and b are taken from the $_G^C \underline{\frac{CG}{AA}}$ segment of the NMR structure of the substrate loop of VS ribozyme (9). The Watson−Crick CG, A N1−C amino single hydrogen-bond AC pair, and sheared GA shown in c and d are taken from $_G^C \underline{\frac{AG}{CA}}$ segment of the NMR structure of loop B of a hairpin ribozyme (37). The stacking figures are generated by the 3DNA program (75).

helix. As with the G of a UG pair (49), the A of a CA^+ pair stacks to its 3′ side by shifting to the minor groove. Thus, having the Watson−Crick pair 3′ of the A provides more favorable stacking by increasing the base overlap (see Figure 4a and Figure S1 in the Supporting Information). Interestingly, the U in $_C^G \underline{\frac{AU}{C}} _C^G$ (loop sequence in a U6 RNA intramolecular stem loop), which is stacked within the helix at pH 7.0, is flipped out at pH 5.7 to favor a stacking interaction between wobble A^+C and Watson−Crick GC pairs flanking the U bulge (10). Evidently, the stabilization effect of CA and/or CA^+ pairs and the pK_a of A in a CA pair is sequence-context-dependent.

No significant stabilization was observed for $_G^C \underline{\frac{CA}{AG}}$ at either pH 7 or 5.5 in $_{PCCG}^{GGU} \underline{\frac{GGA}{AAC}} _{CCG}^{GGCU}$ ($\Delta G°_{37,\text{pH7,loop}}$ = −0.09 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.29 kcal/mol), $_{PCCG}^{GGU} \underline{\frac{AGA}{AAC}} _{CCG}^{GGCU}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 1.66 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.25 kcal/mol), $_{PCCG}^{GGU} \underline{\frac{GA}{AAC}} _{CCG}^{GGCU}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 1.66 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.16 kcal/mol), and $_{GCG}^{CGC} \underline{\frac{AGA}{AAC}} _{CCG}^{GGC}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 1.77 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.04 kcal/mol). This is consistent with sheared-type CA and GA pairs (trans Hoogsteen/sugar-edge AC and AG) (panels d and f of Figure 1) forming in the loop $_G^C \underline{\frac{CA}{AG}} _G^C$ in helix 41a of the crystal structure of Thermus thermophilus 16S rRNA (39). It is possible, however, that a $\frac{CA}{AG}$ nearest neighbor may provide enhanced stability in other contexts. A wobble CA pair adjacent to a sheared GA pair was observed by NMR for the internal loop $_C^G \underline{\frac{CA}{AG}} _U^G$ (sequence of a VS ribozyme active site loop), where the pK_a of N1 of the bold A is 6.2 at 30 °C (50). Evidently, the formation of a stabilizing wobble AC or CA pair adjacent to a GA, AG, or AA pair is sequence-context-dependent.

Adjacent CA Pairs Provide No Significant Stabilization at pH 7 But Are Stabilized by 0.8 kcal/mol on Average at pH 5.5. The 3 × 3 loops in duplexes, $_{GCG}^{CGC} \underline{\frac{AAA}{ACC}} _{CCG}^{GGC}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 2.65 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.82 kcal/mol) and $_{GCG}^{CGC} \underline{\frac{GAA}{ACC}} _{CCG}^{GGC}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 1.00 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.69 kcal/mol), are predicted well without a bonus parameter at pH 7 but have enhanced stabilities at pH 5.5. A similar

pH-dependent effect was observed for the 2 × 2 loops in duplexes, $_{GCG}^{CGC} \underline{\frac{CA}{AC}} _{CGC}^{GCG}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −0.87 kcal/mol) and $_{CCG}^{GGC} \underline{\frac{AC}{CA}} _{CGG}^{GCC}$ ($\Delta\Delta G°_{37,\text{pH}}$ = −0.77 kcal/mol) (5). Perhaps adjacent protonated pairs are not electrostatically favorable and, thus, result in a lower pK_a and stabilized only when pH is as low as 5.5. Tandem wobble CA pairs were observed in $_{GCG}^{CGC} \underline{\frac{CA}{AC}} _{CGC}^{GCG}$ by X-ray crystallography at pH 5.5 (see Figure S1e in the Supporting Information) (51). An average $\Delta G°_{37,\text{pH bonus}}$ = −1.03 ± 0.32 kcal/mol (Table 3) is used to predict pH stabilization for both tandem CA pairs and single CA pairs in appropriate contexts, as described above.

No Significant pH-Dependent Thermodynamic Effect Was Found for Nearest Neighbors with CA Adjacent to UG. For the three duplexes, $_{UCGU}^{GCA} \underline{\frac{AGAA}{CUGC}} _{CCG}^{GGC}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 2.13 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.73 kcal/mol, but the pH effect is presumably only due to $_G^C \underline{\frac{CG}{AA}}$ because no noncanonical pairs form in $_U^A \underline{\frac{AG}{CU}}$ (3)), $_{CUCG}^{GAGC} \underline{\frac{CGA}{AUA}} _{GCUG}^{CGAC}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 1.43 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.09 kcal/mol), and $_{PCCG}^{GGU} \underline{\frac{CAA}{AAG}} _{CCG}^{GGCU}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 0.47 kcal/mol, $\Delta\Delta G°_{37,\text{pH}}$ = −0.19 kcal/mol), no significant pH effect could be attributed to a CA pair adjacent to a UG pair. With the exception of $_{CUCG}^{GAGC} \underline{\frac{CGA}{AUA}} _{GCUG}^{CGAC}$ ($\Delta G°_{37,\text{pH7,loop}}$ = 1.43 kcal/mol versus $\Delta G°_{\text{predicted}}$ = 2.15 kcal/mol at pH 7, $\Delta\Delta G°_{37,\text{pH}}$ = −0.09 kcal/mol), all of the loop free energies are well-predicted at pH 7 for the loops with a CA adjacent to a UG pair. Thus, it is unlikely that in these loops a wobble CA^+ pair is formed adjacent to a wobble UG pair, with the pK_a significantly above 7 for the adenine N1.

Note that there is also no significant thermodynamic difference between pH 8 and 7 for the loop $_{PCCG}^{GGU} \underline{\frac{CAA}{AAG}} _{CCG}^{GGCU}$ ($\Delta G°_{37,\text{pH8,loop}}$ = 0.90 kcal/mol). We applied the bonus parameter of $\Delta G°_{5'CR/3'AA \text{ bonus}}$ for $_{PCCG}^{GGU} \underline{\frac{CAA}{AAG}} _{CCG}^{GGCU}$ at pH 7, although there is no further stabilization at pH 5.5. The pH-dependent shifting of the imino proton resonances from the UG pair suggests a pH-dependent conformational change within the loop, however (Figure 2b). This may be another example of the idiosyncratic behavior of UG pairs. For example, thermodynamic and

NMR studies suggest that adjacent UG pairs do not always form canonical wobble pairs (*52, 53*).

Context-Dependent pH Effect of CC Pairs. CC can form a *cis* Watson−Crick/Watson−Crick CC^+ pair (Figure 1i). A pH-dependent stabilization was observed in 2×2 loops $\frac{CGC}{GCG}\frac{CC}{CC}\frac{GCG}{CGC}$ ($\Delta\Delta G°_{37,pH} = -2.48$ kcal/mol for two CC pairs) (*5*) and in $\frac{GAGC}{CUCG}\frac{CU}{CU}\frac{CGAC}{GCUG}$ ($\Delta\Delta G°_{37,pH} = -0.79$ kcal/mol) (Table 2). In contrast, the stability of a single CC mismatch is essentially pH-independent and well-predicted, $\frac{GAG}{CAGCUC}\frac{C}{C}\frac{CUCGAC}{GAG}$ ($\Delta G°_{37,pH7,loop} = 0.55$ kcal/mol, $\Delta\Delta G°_{37,pH} = 0.11$ kcal/mol). The thermodynamic effect of GAC 3′ dangling ends are assumed to be the same as GA 3′ dangling ends (*54, 55*) to calculate the measured thermodynamic parameters of the 1×1 loop with a single CC mismatch.

UC Pairs Are Not More Stable at Lower pH. No significant pH effect was found for $\frac{CGC}{GCG}\frac{UC}{CU}\frac{GCG}{CGC}$ ($\Delta\Delta G°_{37,pH} = -0.10$ kcal/mol) and $\frac{CGC}{GCG}\frac{CU}{UC}\frac{GCG}{CGC}$ ($\Delta\Delta G°_{37,pH} = -0.06$ kcal/mol) (*5*). This is consistent with crystal structures of $\frac{U}{G}\frac{UC}{CU}\frac{G}{U}$ that reveal *cis* Watson−Crick/Watson−Crick UC pairs with a water-mediated hydrogen bond between the U imino proton and C N3 but without protonated nucleobases (Figure 1j) (*56, 57*). Quantum chemical calculations show that a water-mediated UC pair is energetically preferred over a UC pair with two direct hydrogen bonds (U O4 to C amino and U H3 to C N3) (Figure 1k) (*58*). Molecular dynamics simulations of the loops $\frac{U}{G}\frac{UC}{CU}\frac{G}{U}$ (*59*) and $\frac{C}{G}\frac{UUUC}{UUUU}\frac{A}{U}$ (sequence found in human telomerase RNA) (*60*) reveal dynamics of the water-mediated UC pairs. No significant pH effect is observed for $\frac{GAGC}{CUCG}\frac{CU}{UU}\frac{CGAC}{GCUG}$ ($\Delta\Delta G°_{37,pH} = 0.16$ kcal/mol) and $\frac{GAGC}{CUCG}\frac{UU}{CU}\frac{CGAC}{GCUG}$ ($\Delta\Delta G°_{37,pH} = 0.21$ kcal/mol), which is consistent with the NMR structures of $\frac{C}{G}\frac{CU}{UU}\frac{C}{G}$ (sequence found in a poliovirus 3′-UTR) (*61*) and $\frac{C}{G}\frac{UU}{CU}\frac{C}{G}$ (sequence found in HCV IRES domain II) (*62*), which contain no protonated C^+U and UC^+ pairs (Figure 1l), respectively.

A Watson−Crick-type UC pair with two direct hydrogen bonds (U O4 to C amino and U H3 to C N3) (Figure 1k) was observed in $\frac{C}{G}\frac{UUU}{UCU}\frac{G}{C}$ (loop found in several RNA viruses) by NMR (*63, 64*), which is consistent with the small pH-dependent thermodynamics observed for $\frac{CGC}{GCG}\frac{UUU}{UCU}\frac{GGC}{CCG}$ ($\Delta G°_{37,pH7,loop} = 0.72$ kcal/mol, $\Delta\Delta G°_{37,pH} = 0.57$ kcal/mol). The loop free energies at pH 7 are well-predicted, and no pH stabilization is observed for loops in the duplexes, $\frac{GAGC}{CUCG}\frac{UGC}{UAU}\frac{CGAC}{GCUG}$ ($\Delta G°_{37,pH7,loop} = 1.54$ kcal/mol, $\Delta\Delta G°_{37,pH} = 0.17$ kcal/mol), $\frac{GAGC}{CUCG}\frac{CGA}{UAA}\frac{CGAC}{GCUG}$ ($\Delta G°_{37,pH7,loop} = 1.26$ kcal/mol, $\Delta\Delta G°_{37,pH} = 0.41$ kcal/mol), or $\frac{CGCA}{GCGU}\frac{U}{C}\frac{AGGC}{UCCG}$ ($\Delta G°_{37,pH7,loop} = 2.36$ kcal/mol, $\Delta\Delta G°_{37,pH} = 0.42$ kcal/mol).

No pH Bonus Is Applied to Size-Asymmetric Internal Loops. The duplex, $\frac{CGAC}{GCUG}\frac{CGAG}{AG}\frac{CCAG}{GGUC}$ ($\Delta G°_{37,pH7,loop} = 2.60$ kcal/mol, $\Delta\Delta G°_{37,pH} = -0.73$ kcal/mol), has the 2×4 internal loop from the leadzyme (*1, 65−67*) and is 0.73 kcal/mol more stable at pH 5.5 than pH 7. This stabilization is consistent with the formation of a wobble CA^+ pair in the NMR structure without multivalent metal ions (*1, 65*) but not with the crystal structure with multivalent ions (*66*) and a molecular modeling study of the active conformation (*67*). The molecular model of the active conformation is consistent with kinetic studies, in which different loop G's are forced to be in *syn* glycosidic conformation (*68*).

In helix 58 of the large ribosomal subunit of *Haloarcula marismortui* (*40*), *trans* Hoogsteen/sugar AC (Figure 1d) and *trans* Hoogsteen/Hoogsteen AA pairs (Figure 1h) (*43*) form in $\frac{G}{C}\frac{CAUA}{AAG}\frac{G}{C}$. The A in bold is in a *syn* glycosidic conformation, and

the U is bulged out. Apparently, this conformation is more stable in this 3×4 internal loop than a wobble CA^+ pair adjacent to a sheared AA pair. Further thermodynamic and structural studies are needed to see whether the loop structure is preformed or induced by tertiary and protein binding in the ribosome.

Moderate pH effects were found for the 1×2 loop in $\frac{UGAG}{ACUC}\frac{C}{CC}\frac{GUCA}{CAGU}$ ($\Delta\Delta G°_{37,pH} = -1.04$ kcal/mol) and the 2×3 loop in $\frac{UCAG}{AGUC}\frac{CC}{AAU}\frac{GUGA}{CACU}$ ($\Delta\Delta G°_{37,pH} = -0.63$ kcal/mol) (*69*). Pairings within the loop that are sensitive to pH and context are likely in size-asymmetric internal loops, because they provide flexibility (*1, 65−68*). This will make it difficult to determine sequence- and pH-dependent rules for size-asymmetric internal loops and other flexible loops.

Thermodynamics of Internal Loops May Be Useful for Predicting Kinetics. The internal loops of $\frac{GCA}{UCGU}\frac{AGAA}{CUGC}\frac{GGC}{CCG}$ ($\Delta G°_{37,pH7,loop} = 2.13$ kcal/mol), $\frac{GCA}{UCGU}\frac{AGAA}{CAGC}\frac{GGC}{CCG}$ ($\Delta G°_{37,pH7,loop} = -0.09$ kcal/mol), and $\frac{GCA}{UCGU}\frac{AGAA}{CAGG}\frac{GGC}{CCG}$ ($\Delta G°_{37,pH7,loop} = -1.25$ kcal/mol) belong to proposed consensus sequences in loop A of the hairpin ribozymes (*3*). The bold G has to flip out and dock with the loop B domain to form the functional hairpin ribozyme conformation (*70*). On the basis of loop A stability and structure, this step is predicted to be slowest at pH 7 for the $\frac{A}{U}\frac{AGAA}{CAGG}\frac{G}{C}$ loop (Figure 3a) if the transition states for all of the sequences have similar free energies.

CONCLUSION

The pK_a of the A N1 nitrogen in a CA pair depends upon local sequence context, as evidenced by thermodynamic and structural results shown here and previously (*1, 3, 8, 9, 12, 34−40, 42*). In a nearest neighbor of $\frac{CG}{AA}$ or $\frac{CA}{AA}$ with the CA adjacent to a closing canonical pair (including wobble UG pairs), the formation of a wobble CA^+ adjacent to a sheared GA or AA pair stabilizes 3×3 nucleotide and larger size-symmetric internal loops on average by about 1 kcal/mol at 37 °C, pH 7, and 1 M NaCl. Such nearest neighbors with the CA adjacent to a closing Watson−Crick pair are further stabilized on average by 1 kcal/mol at 37 °C when the pH is lowered from 7 to 5.5. Other stabilizing nearest neighbor combinations can exist to shift pK_a. The pK_a may also depend upon global context; e.g., pK_a could be shifted in the middle of a large structure, such as the ribosome. The results presented here along with published NMR and crystal structures provide benchmarks to test free-energy and structural calculations by computational chemists.

SUPPORTING INFORMATION AVAILABLE

Tables of single-strand UV melting results, linear regression data, figures of base stacking and base pairing involving CA pairs, and an exchangeable proton SNOESY spectrum. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Legault, P., and Pardi, A. (1997) Unusual dynamics and pK_a shift at the active site of a lead-dependent ribozyme. *J. Am. Chem. Soc. 119*, 6621−6628.
2. Bevilacqua, P. C. (2003) Mechanistic considerations for general acid−base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme. *Biochemistry 42*, 2259−2265.

3. Cai, Z., and Tinoco, I. Jr. (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry 35*, 6026–6036.

4. Allawi, H. T., and SantaLucia, J. Jr. (1998) Nearest-neighbor thermodynamics of internal AC mismatches in DNA: Sequence dependence and pH effects. *Biochemistry 37*, 9435–9444.

5. SantaLucia, J. Jr., Kierzek, R., and Turner, D. H. (1991) Stabilities of consecutive A·C, C·C, G·G, U·C, and U·U mismatches in RNA internal loops: Evidence for stable hydrogen bonded U·U and C·C$^+$ pairs. *Biochemistry 30*, 8242–8251.

6. Bevilacqua, P. C., Brown, T. S., Chadalavada, D., Lecomte, J., Moody, E., and Nakano, S. (2005) Linkage between proton binding and folding in RNA: Implications for RNA catalysis. *Biochem. Soc. Trans. 33*, 466–470.

7. Moody, E. M., Lecomte, J. T. J., and Bevilacqua, P. C. (2005) Linkage between proton binding and folding in RNA: A thermodynamic framework and its experimental application for investigating pK_a shifting. *RNA 11*, 157–172.

8. Flinders, J., and Dieckmann, T. (2001) A pH controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA. *J. Mol. Biol. 308*, 665–679.

9. Michiels, P. J. A., Schouten, C. H. J., Hilbers, C. W., and Heus, H. A. (2000) Structure of the ribozyme substrate hairpin of Neurospora VS RNA: A close look at the cleavage site. *RNA 6*, 1821–1832.

10. Reiter, N. J., Blad, H., Abildgaard, F., and Butcher, S. E. (2004) Dynamics in the U6 RNA intramolecular stem-loop: A base flipping conformational change. *Biochemistry 43*, 13739–13747.

11. Zimmermann, G. R., Shields, T. P., Jenison, R. D., Wick, C. L., and Pardi, A. (1998) A semiconserved residue inhibits complex formation by stabilizing interactions in the free state of a theophylline-binding RNA. *Biochemistry 37*, 9186–9192.

12. Puglisi, J. D., Wyatt, J. R., and Tinoco, I. Jr. (1990) Solution conformation of an RNA hairpin loop. *Biochemistry 29*, 4215–4226.

13. Durant, P. C., Bajji, A. C., Sundaram, M., Kumar, R. K., and Davis, D. R. (2005) Structural effects of hypermodified nucleosides in the *Escherichia coli* and human tRNA^{Lys} anticodon loop: The effect of nucleosides s^{2}U, mcm^{5}U, mcm^{5}s^{2}U, mnm^{5}s^{2}U, t^{6}A, and ms^{2}t^{6}A. *Biochemistry 44*, 8078–8089.

14. Bink, H. H., Hellendoorn, K., van der Meulen, J., and Pleij, C. W. (2002) Protonation of non-Watson−Crick base pairs and encapsidation of turnip yellow mosaic virus RNA. *Proc. Natl. Acad. Sci. U.S.A. 99*, 13465–13470.

15. Reiche, K., and Stadler, P. F. (2007) RNAstrand: Reading direction of structured RNAs in multiple sequence alignments. *Algorithms Mol. Biol. 2*, 6.

16. Chen, G., and Turner, D. H. (2006) Consecutive GA pairs stabilize medium-size RNA internal loops. *Biochemistry 45*, 4025–4043.

17. Usman, N., Ogilvie, K. K., Jiang, M. Y., and Cedergren, R. J. (1987) Automated chemical synthesis of long oligoribonucleotides using 2'-O-silylated ribonucleoside 3'-O-phosphoramidites on a controlled-pore glass support: Synthesis of a 43-nucleotide sequence similar to the 3'-half molecule of an *Escherichia coli* formylmethionine tRNA. *J. Am. Chem. Soc. 109*, 7845–7854.

18. Wincott, F., Direnzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) Synthesis, deprotection, analysis and purification of RNA and ribozymes. *Nucleic Acids Res. 23*, 2677–2684.

19. Chen, G., Znosko, B. M., Jiao, X. Q., and Turner, D. H. (2004) Factors affecting thermodynamic stabilities of RNA 3 × 3 internal loops. *Biochemistry 43*, 12865–12876.

20. Chen, G., Znosko, B. M., Kennedy, S. D., Krugh, T. R., and Turner, D. H. (2005) Solution structure of an RNA internal loop with three consecutive sheared GA pairs. *Biochemistry 44*, 2845–2856.

21. Borer, P. N. (1975) Optical properties of nucleic acids, absorption and circular dichroism spectra, in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd ed. (Fasman, G. D., Ed.) pp 589−595, CRC Press, Cleveland, OH.

22. Richards, E. G. (1975) Use of tables in calculation of absorption, optical rotatory dispersion and circular dichroism of polyribonucleotides, in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 3rd ed. (Fasman, G. D., Ed.) pp 596−603, CRC Press, Cleveland, OH.

23. McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAGGUCUC)$_2$ by two-dimensional NMR and simulated annealing. *Biochemistry 35*, 14077–14089.

24. Petersheim, M., and Turner, D. H. (1983) Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry 22*, 256–263.

25. Xia, T., SantaLucia, J. Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson−Crick base pairs. *Biochemistry 37*, 14719–14735.

26. Borer, P. N., Dengler, B., Tinoco, I. Jr., and Uhlenbeck, O. C. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol. 86*, 843–853.

27. Lukavsky, P. J., and Puglisi, J. D. (2001) RNAPack: An integrated NMR approach to RNA structure determination. *Methods 25*, 316–332.

28. Smallcombe, S. H. (1993) Solvent suppression with symmetrically shifted pulses. *J. Am. Chem. Soc. 115*, 4776–4785.

29. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A. 101*, 7287–7292.

30. Gralla, J., and Crothers, D. M. (1973) Free energy of imperfect nucleic acid helices: III. Small internal loops resulting from mismatches. *J. Mol. Biol. 78*, 301–319.

31. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol. 288*, 911–940.

32. Blake, R. D. (1996) Denaturation of DNA, in *Encyclopedia of Molecular Biology and Molecular Medicine* (Meyers, R. A., Ed.) pp 1−19, VCH, New York.

33. Record, M. T. Jr. (1967) Electrostatic effects on polynucleotide transitions. II. Behavior of titrated systems. *Biopolymers 5*, 993–1008.

34. Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z. D., Zhao, Q., and Fox, G. E. (2002) NCIR: A database of non-canonical interactions in known RNA structures. *Nucleic Acids Res. 30*, 395–397.

35. Tamura, M., Hendrix, D. K., Klosterman, P. S., Schimmelman, N. R., Brenner, S. E., and Holbrook, S. R. (2004) SCOR: Structural classification of RNA, version 2.0. *Nucleic Acids Res. 32*, D182–D184.

36. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. M., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature 430*, 45–50.

37. Butcher, S. E., Allain, F. H. T., and Feigon, J. (1999) Solution structure of the loop B domain from the hairpin ribozyme. *Nat. Struct. Biol. 6*, 212–216.

38. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. D. (2005) Structures of the bacterial ribosome at 3.5 Å resolution. *Science 310*, 827–834.

39. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature 407*, 327–339.

40. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science 289*, 905–920.

41. Acharya, S., Barman, J., Cheruku, P., Chatterjee, S., Acharya, P., Isaksson, J., and Chattopadhyaya, J. (2004) Significant pK_a perturbation of nucleobases is an intrinsic property of the sequence context in DNA and RNA. *J. Am. Chem. Soc. 126*, 8674–8681.

42. Pan, B., Mitra, S. N., and Sundaralingam, M. (1998) Structure of a 16-mer RNA duplex r(GCAGACUUAAAUCUGC)$_2$ with wobble CA$^+$ mismatches. *J. Mol. Biol. 283*, 977–984.

43. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) The non-Watson−Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res. 30*, 3497–3531.

44. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry 38*, 14214–14223.

45. Wang, C., Gao, H., Gaffney, B. L., and Jones, R. A. (1991) Nitrogen-15-labeled oligodeoxynucleotides. 3. Protonation of the adenine N1 in the AC and AG mispairs of the duplexes {d[CG(^{15}N^1) AGAATTCCCG]}$_2$ and {d[CGGGAATTC(^{15}N^1)ACG]}$_2$. *J. Am. Chem. Soc. 113*, 5486–5488.

46. Wong, S. K., Sato, S., and Lazinski, D. W. (2001) Substrate recognition by ADAR1 and ADAR2. *RNA 7*, 846–858.

47. Yildirim, I., and Turner, D. H. (2005) RNA challenges for computational chemists. *Biochemistry 44*, 13225–13234.

48. Tang, C. L., Alexov, E., Pyle, A. M., and Honig, B. (2007) Calculation of pK_as in RNA: On the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol. 366*, 1475–1496.

49. Mizuno, H., and Sundaralingam, M. (1978) Stacking of Crick wobble pair and Watson−Crick pair: Stability rules of G−U pairs at ends of helical stems in tRNAs and the relation to codon−-anticodon wobble interaction. *Nucleic Acids Res. 5*, 4451–4461.

50. Flinders, J., and Dieckmann, T. (2004) The solution structure of the VS ribozyme active site loop reveals a dynamic "hot-spot". *J. Mol. Biol. 341*, 935–949.

51. Jang, S. B., Hung, L. W., Chi, Y. I., Holbrook, E. L., Carter, R. J., and Holbrook, S. R. (1998) Structure of an RNA internal loop consisting of tandem CA$^+$ base pairs. *Biochemistry 37*, 11726–11731.

52. Znosko, B. M., Kennedy, S. D., Wille, P. C., Krugh, T. R., and Turner, D. H. (2004) Structural features and thermodynamics of the J4/5 loop from the *Candida albicans* and *Candida dubliniensis* group I introns. *Biochemistry 43*, 15822–15837.

53. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for G·U pairs: One hydrogen bond for each G·U pair in r(GGCGUGCC)$_2$ and two for each G·U pair in r(GAGUGCUC)$_2$. Biochemistry 39, 8970–8982.

54. O'Toole, A. S., Miller, S., Haines, N., Zink, M. C., and Serra, M. J. (2006) Comprehensive thermodynamic analysis of 3′ double-nucleotide overhangs neighboring Watson−Crick terminal base pairs. *Nucleic Acids Res. 34*, 3338–3344.

55. Ohmichi, T., Nakano, S., Miyoshi, D., and Sugimoto, N. (2002) Long RNA dangling end has large energetic contribution to duplex stability. *J. Am. Chem. Soc. 124*, 10367–10372.

56. Holbrook, S. R., Cheong, C. J., Tinoco, I. Jr., and Kim, S. H. (1991) Crystal structure of an RNA double helix incorporating a track of non-Watson−Crick base pairs. *Nature 353*, 579–581.

57. Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prange, T., and Kennard, O. (1994) Structure of a mispaired RNA double helix at 1.6 Å resolution and implications for the prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A. 91*, 4160–4164.

58. Brandl, M., Meyer, M., and Suhnel, J. (1999) Quantum−chemical study of a water-mediated uracil−cytosine base pair. *J. Am. Chem. Soc. 121*, 2605–2606.

59. Schneider, C., Brandl, M., and Suhnel, J. (2001) Molecular dynamics simulation reveals conformational switching of water-mediated uracil−cytosine base-pairs in an RNA duplex. *J. Mol. Biol. 305*, 659–667.

60. Yingling, Y. G., and Shapiro, B. A. (2005) Dynamic behavior of the telomerase RNA hairpin structure and its relationship to *Dyskeratosis congenita*. *J. Mol. Biol. 348*, 27–42.

61. Lescrinier, E., Tessari, M., van Kuppeveld, F. J. M., Melchers, W. J. G., Hilbers, C. W., and Heus, H. A. (2003) Structure of the pyrimidine-rich internal loop in the poliovirus 3′-UTR: The importance of maintaining pseudo-2-fold symmetry in RNA helices containing two adjacent non-canonical base-pairs. *J. Mol. Biol. 331*, 759–769.

62. Lukavsky, P. J., Kim, I., Otto, G. A., and Puglisi, J. D. (2003) Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol. 10*, 1033–1038.

63. Du, Z. H., Yu, J. H., Ulyanov, N. B., Andino, R., and James, T. L. (2004) Solution structure of a consensus stem-loop D RNA domain that plays important roles in regulating translation and replication in enteroviruses and rhinoviruses. *Biochemistry 43*, 11959–11972.

64. Ohlenschlager, O., Wohnert, J., Bucci, E., Seitz, S., Hafner, S., Ramachandran, R., Zell, R., and Gorlach, M. (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure 12*, 237–248.

65. Hoogstraten, C. G., Legault, P., and Pardi, A. (1998) NMR solution structure of the lead-dependent ribozyme: Evidence for dynamics in RNA catalysis. *J. Mol. Biol. 284*, 337–350.

66. Wedekind, J. E., and McKay, D. B. (2003) Crystal structure of the leadzyme at 1.8 Å resolution: Metal ion binding and the implications for catalytic mechanism and allo site ion regulation. *Biochemistry 42*, 9554–9563.

67. Lemieux, S., Chartrand, P., Cedergren, R., and Major, F. (1998) Modeling active RNA structures using the intersection of conformational space: Application to the lead-activated ribozyme. *RNA 4*, 739–749.

68. Yajima, R., Proctor, D. J., Kierzek, R., Kierzek, E., and Bevilacqua, P. C. (2007) A conformationally restricted guanosine analog reveals the catalytic relevance of three structures of an RNA enzyme. *Chem. Biol. 14*, 23–30.

69. Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry 39*, 9257–9274.

70. Ferre-D'Amare, A. R. (2004) The hairpin ribozyme. *Biopolymers 73*, 71–78.

71. Weichenrieder, O., Wild, K., Strub, K., and Cusack, S. (2000) Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature 408*, 167–173.

72. Chen, G., Kennedy, S. D., Qiao, J., Krugh, T. R., and Turner, D. H. (2006) An alternating sheared AA pair and elements of stability for a single sheared purine−purine pair flanked by sheared GA pairs in RNA. *Biochemistry 45*, 6889–6903.

73. Heus, H. A., Wijmenga, S. S., Hoppe, H., and Hilbers, C. W. (1997) The detailed structure of tandem G·A mismatched base-pair motifs in RNA duplexes is context dependent. *J. Mol. Biol. 271*, 147–158.

74. Tolbert, B. S., Kennedy, S. D., Schroeder, S. J., Krugh, T. R., and Turner, D. H. (2007) NMR structures of (rGCUGAGGCU)$_2$ and (rGCGGAUGCU)$_2$: Probing the structural features that shape the thermodynamic stability of GA pairs. *Biochemistry 46*, 1511–1522.

75. Lu, X. J., and Olson, W. K. (2003) 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res. 31*, 5108–5121.