Consecutive GA Pairs Stabilize Medium-Size RNA Internal Loops†

Gang Chen‡ and Douglas H. Turner*,‡,§

Department of Chemistry, University of Rochester, Rochester, New York 14627, and Center for Pediatric Biomedical Research and Department of Pediatrics, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642

Received October 10, 2005; Revised Manuscript Received December 15, 2005

ABSTRACT: Internal loops in RNA are important for folding and function. Consecutive noncanonical pairs can form in internal loops having at least two nucleotides on each side. Thermodynamic and structural insights into such internal loops should improve approximations for their stabilities and predictions of secondary and three-dimensional structures. Most natural internal loops are purine rich. A series of oligoribonucleotides that form purine-rich internal loops of 5−10 nucleotides, including kink-turn loops, were studied by UV melting, exchangeable proton and phosphorus NMR. Three consecutive GA pairs with the motif $\begin{smallmatrix} 5'\,\mathrm{Y}\,\underline{\mathrm{GGA}} \\ 3'\,\mathrm{R}\,\mathrm{AAG} \end{smallmatrix}$ or $\begin{smallmatrix} \underline{\mathrm{GGA}}\,\mathrm{R}\,3' \\ \mathrm{AAG}\,\mathrm{Y}\,5' \end{smallmatrix}$ (i.e., $\begin{smallmatrix} 5'\,\underline{\mathrm{GGA}}\,3' \\ 3'\,\mathrm{AAG}\,5' \end{smallmatrix}$ closed on at least one side with a CG, UA, or UG pair with Y representing C or U and R representing A or G) stabilize internal loops having 6−10 nucleotides. Certain motifs with two consecutive GA pairs are also stabilizing. In internal loops with three or more nucleotides on each side, the motif $\begin{smallmatrix} 5'\,\mathrm{U}\,\underline{\mathrm{G}} \\ 3'\,\mathrm{G}\,\mathrm{A} \end{smallmatrix}$ has stability similar to $\begin{smallmatrix} 5'\,\mathrm{C}\,\underline{\mathrm{G}} \\ 3'\,\mathrm{G}\,\mathrm{A} \end{smallmatrix}$. A revised model for predicting stabilities of internal loops with 6−10 nucleotides is derived by multiple linear regression. Loops with 2 × 3 nucleotides are predicted well by a previous thermodynamic model.

Sequence-dependent secondary structure interactions in RNA usually dominate energetically over tertiary interactions (*1*−*5*). Thus, free energy parameters derived from studies of short oligonucleotides (*5*−*9*) allow prediction of RNA secondary structures with about 73% accuracy on average without consideration of tertiary structure when the RNA is shorter than about 500 nucleotides (*8*, *9*). This accuracy could be improved with more knowledge of the sequence dependence of stabilities for loops in RNA. This work provides insight into the sequence dependence of stability for internal loops. Internal loops are important elements of tertiary structure (*10*−*17*) and are binding sites for proteins and therapeutics (*17*−*25*).

Thermodynamics and structures of internal loops in oligoribonucleotides have been studied by UV melting and NMR, respectively (*5*, *6*, *26*−*40*). Current free energy parameters derived for internal loops (*9*) are largely based on knowledge of 2 × 2[1] and 2 × 3 internal loops, where "n1 × n2" represents an internal loop with n1 and n2 nucleotides on each side, respectively. Currently, only the thermodynamic effect of the first noncanonical pair on each side of an internal loop is considered in structure prediction algorithms.

Stabilities of size symmetric (i.e., n1 = n2) internal loops are more sequence dependent than size asymmetric (i.e., n1 ≠ n2) loops (*27*). Presumably, this is because asymmetric loops are more flexible. This flexibility is also reflected in the observation that asymmetric loops are typically relatively unstructured in solution (*21*, *24*, *26*, *30*, *41*−*44*). Structured consecutive noncanonical pairs have been observed, however, in large internal loops and hairpins (*31*−*33*, *45*−*49*).

The motif of three consecutive sheared GA pairs $\begin{smallmatrix} 5'\,\underline{\mathrm{GGA}}\,3' \\ 3'\,\mathrm{AAG}\,5' \end{smallmatrix}$ is the most stable among 3 × 3 internal loops (*33*, *34*). Two consecutive GA pairs $\begin{smallmatrix} 5'\,\underline{\mathrm{GA}}\,3' \\ 3'\,\mathrm{AG}\,5' \end{smallmatrix}$ are also the most stable among 2 × 2 internal loops. (Throughout the paper, each top strand is written from 5′ to 3′ going from left to right.) Formation of consecutive sheared GA pairs is well conserved in some size asymmetric loops, including certain types of kink-turn motifs (*17*−*23*, *50*). This suggests that thermodynamic stabilization due to consecutive GA pairs may not be restricted to 2 × 2 and 3 × 3 internal loops. Biophysical and biochemical studies of a kink-turn suggest the formation of three consecutive GA pairs within the 3 × 6 internal loop $\begin{smallmatrix} \mathrm{G}\,\underline{\mathrm{GGA}}\,\mathrm{G} \\ \mathrm{C}\,\mathrm{AAGAAG}\,\mathrm{C} \end{smallmatrix}$ even in the absence of Mg^{2+} (aq) and protein (*51*). Consecutive two or three GA pairs are also found in internal loops in signal recognition particle (SRP) RNA (*33*, *52*); in the substrate loop of VS ribozyme (*53*, *54*); in multibranch loops such as the P5abc domain of the *Tetrahymena thermophila* group I intron (*14*, *55*, *56*); in the putative catalytic site of hammerhead ribozymes (*57*); and in a variety of structural elements in ribosomal RNA (*33*, *37*, *58*−*60*).

Here, the thermodynamics of internal loops with 5−10 nucleotides are presented. Multiple linear regression is used to develop a revised thermodynamic model for loops larger than 2 × 3. Three consecutive GA pairs with the motif

† This work was supported by NIH Grant GM22939.

* To whom correspondence should be addressed. Phone: (585) 275−3207. Fax: (585) 276−0205. E-mail: turner@chem.rochester.edu.

‡ Department of Chemistry, University of Rochester.

§ Center for Pediatric Biomedical Research and Department of Pediatrics, University of Rochester.

[1] Abbreviations: C_T, total concentration of all strands of oligonucleotides in solution; eu, entropy units = cal K^{-1} mol^{-1}; n1 × n2 or $X_{n1} \times W_{n2}$, an internal loop with n1 nucleotides on one side and n2 nucleotides on the opposite side (n1 ≤ n2); T_M, melting temperature in Kelvin; T_m, melting temperature in degrees Celsius; YR, canonical pair of UA, UG, or CG, with Y on the 5′ side and R on the 3′ side of the internal loop; YR/RY, canonical pair of YR or RY; ΔG°_{37}, measured free energy at 37 °C for duplex formation; $\Delta G^\circ_{37,\mathrm{loop}}$, measured free energy at 37 °C for the internal loop formation; $\Delta G^\circ_{\mathrm{predicted}}$, free energy increment of internal loop formation at 37 °C predicted from the model in the RNAstructure algorithm, MFOLD algorithm, or revised thermodynamic model derived here; P, purine.

10.1021/bi052060t CCC: $33.50
Published on Web 03/03/2006

of ${}^{\text{Y}}_{\text{R}}\frac{\text{GGA}}{\text{AAG}}$ or $\frac{\text{GGA}}{\text{AAG}}{}^{\text{R}}_{\text{Y}}$ (i.e., $\frac{\text{GGA}}{\text{AAG}}$ closed on at least one side with a CG, UA, or UG pair with Y representing C or U and R representing A or G) can stabilize $X_3 \times W_{3-6}$ and $X_4 \times W_{4-6}$ internal loops. The motifs of ${}^{\text{R}}_{\text{Y}}\frac{\text{GGA}}{\text{AAG}}$ and $\frac{\text{GGA}}{\text{AAG}}{}^{\text{Y}}_{\text{R}}$ (i.e., $\frac{\text{GGA}}{\text{AAG}}$ motif not closed with at least one YR canonical pair) are also stabilizing but less than those with the closing canonical pair reversed. Two consecutive GA pairs with the motif ${}^{\text{Y}}_{\text{R}}\frac{\text{GA}}{\text{AG}}$, ${}^{\text{R}}_{\text{Y}}\frac{\text{GA}}{\text{AG}}$, ${}^{\text{Y}}_{\text{R}}\frac{\text{GG}}{\text{AA}}$, or ${}^{\text{R}}_{\text{Y}}\frac{\text{GG}}{\text{AA}}$ stabilize 3 × 3, 3 × 4, 4 × 4, and 4 × 5 nucleotide internal loops (i.e., loops with the closing base pair 3′ to the A of a GA pair and with asymmetry $n2 - n1 < 2$). In 3 × 3 and larger loops, the motif ${}^{\text{U}}_{\text{G}}{}^{\text{G}}_{\text{A}}$ has a stability similar to ${}^{\text{C}}_{\text{G}}{}^{\text{G}}_{\text{A}}$ even though UG pairs are usually less stable than CG pairs.

The thermodynamic model developed here will help improve RNA secondary structure prediction, particularly prediction of medium-size internal loops, including those that form kink-turns. Internal loops are also often involved in the formation of tertiary structure (*10*−*17*), and tertiary structure can perturb secondary structure (*55*, *61*, *62*). Thus, the results presented here should also facilitate three-dimensional (3D) structure modeling of RNAs and consideration of tertiary interactions in predicting secondary structure.

MATERIALS AND METHODS

Oligoribonucleotide Synthesis and Purification. Oligoribonucleotides were synthesized using the phosphoramidite method (*63*, *64*) and purified as described before (*33*, *34*). CPG supports and phosphoramidites were acquired from Proligo, Glen Research, or ChemGenes. Purities were checked by reverse phase HPLC or analytical TLC on a Baker Si500F silica gel plate (250 μm thick) and all were greater than 95% pure.

UV Melting Experiments and Thermodynamics. Concentrations of single-stranded oligonucleotides were calculated from the absorbance at 280 nm at 80 °C and extinction coefficients predicted from those of dinucleotide monophosphates and nucleosides (*65*, *66*) with the RNAcalc program (*67*). Purine riboside (P) was assumed to be the same as adenosine for the approximation of extinction coefficients. Small mixing errors for non-self-complementary duplexes do not affect thermodynamic measurements appreciably (*68*).

Oligonucleotides were lyophilized and redissolved in 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM disodium EDTA at pH 7. 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 flowing water.

Melting curves were fit to a two-state model with the MeltWin program (http://www.meltwin.com), assuming linear sloping baselines and temperature-independent $\Delta H°$ and $\Delta S°$ (*7*, *67*, *69*). Additionally, 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 non-self-complementary duplexes according to (*70*):

$$T_M^{-1} = (R/\Delta H°)\ln(C_T/4) + (\Delta S°/\Delta H°) \qquad (1)$$

Here R is the gas constant, 1.987 cal/mol·K. The $\Delta H°$ values from T_M^{-1} versus $\ln(C_T/4)$ plots and from the average of the fits of melting curves to two-state transitions agree within 15% (Table 1), suggesting that the two-state model is a good 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).

Self-structure (hairpin and/or duplex) of individual single strands may compete with designed non-self-complementary duplexes. Melting data for individual single strands are listed in Table S1, Supporting Information. The rough standard of sequence design was that the T_m's of duplexes are 5 °C higher than those of individual single strands and the $\Delta G°_{37\,(\text{duplex with loop})}$ values are at least 1.4 kcal/mol more favorable than those of duplex formation by individual single strands. It is possible, however, to measure reasonable thermodynamic parameters even when the T_m of a competing homoduplex is a few degrees higher than that of the heteroduplex (*71*).

NMR Spectroscopy. All exchangeable proton spectra were acquired on a Varian Inova 500 MHz (^{1}H) spectrometer. One-dimensional (1D) imino proton spectra were acquired with an S pulse sequence and temperatures ranging from 0 to 55 °C. SNOESY (*72*) spectra were recorded with a 150 ms mixing time at 5 or 10 °C. The Felix (2000) software package (Molecular Simulations Inc.) was used to process two-dimensional (2D) spectra. Proton spectra were referenced to H_2O or HDO at a known temperature-dependent chemical shift relative to 3-(trimethylsilyl) tetradeutero sodium propionate (TSP). The 1D ^{1}H-decoupled ^{31}P spectra (referenced to external standard of 85% H_3PO_4 at 0 ppm) were acquired on a Bruker Avance 500 MHz (^{1}H) spectrometer at 30 °C. Sample buffer conditions were 80 mM NaCl, 10 mM sodium phosphate, 0.5 mM Na_2EDTA. Total volumes were 300 μL with 90:10 (v:v) H_2O/D_2O or 100% D_2O.

RESULTS

Thermodynamics. Measured thermodynamic parameters for duplexes at 1 M NaCl are listed in Table 1. Thermodynamic parameters for formation of the internal loops (Table 2) were calculated from measured parameters of duplexes according to the following equation (*73*):

$$\Delta G°_{37,\text{loop}} = \Delta G°_{37\,(\text{duplex with loop})} - \Delta G°_{37(\text{duplex without loop})} + \Delta G°_{37(\text{interrupted base stack})} \qquad (2a)$$

For example,

$$\Delta G°_{37}\ {}^{\text{U}}_{\text{G}}\frac{\text{GGAA}}{\text{AAG}}{}^{\text{G}}_{\text{C}} = \Delta G°_{37}\ {}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}} - \Delta G°_{37}\ {}^{\text{GGUGGCU}}_{\text{PCCGCCG}} + \Delta G°_{37}\ {}^{\text{UG}}_{\text{GC}} \qquad (2b)$$

Here, $\Delta G°_{37}\ {}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$ is the measured value of the duplex containing the internal loop, $\Delta G°_{37}\ {}^{\text{GGUGGCU}}_{\text{PCCGCCG}}$ is the measured value of the duplex without the loop (*33*), and $\Delta G°_{37}\ {}^{\text{UG}}_{\text{GC}}$ is the free energy increment for the nearest-neighbor interaction interrupted by the internal loop (*7*, *8*). Nearest-neighbor parameters (*7*, *8*) are used to estimate the difference of one or two base pairs compared with ${}^{\text{GGUGGCU}}_{\text{PCCGCCG}}$ (*33*). Identical calculations can be done for measured values for $\Delta H°_{\text{loop}}$ and $\Delta S°_{\text{loop}}$. All the measured thermodynamic parameters used in this calculation are derived from T_M^{-1} versus $\ln(C_T/4)$ plots (eq 1). In Tables 1 and 2, sequences are ordered from smallest to largest according to internal

Table 1: Measured Thermodynamic Parameters for Duplex Formation in 1 M NaCl, pH 7

Sequences	T_M^{-1} vs $\ln(C_T/4)$ plots (eq 1)				Average of melt curve fits			
	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)
2 × 3 internal loops								
GGC**GA** GGCU PCCG**AAG**CCG	84.7±4.3	234.6±12.9	11.92±0.26	58.1	80.1±5.7	220.8±17.1	11.65±0.40	58.2
GGC**GAA**GGCU PCCG**AG** CCG	82.7±1.7	228.7±5.3	11.72±0.10	57.8	74.0±2.3	202.2±7.1	11.23±0.12	58.1
GGC**GGA**GGCU PCCG**AG** CCG	82.6±4.7	228.6±14.5	11.71±0.27	57.7	76.6±4.3	210.3±13.4	11.40±0.17	58.1
GGU**GAA**GGCU PCCG**AG** CCG	83.9±1.7	236.1±5.3	10.70±0.07	53.2	86.4±1.8	243.8±5.4	10.80±0.11	53.1
GGU**GGA**GGCU PCCG**AG** CCG	87.2±1.9	247.3±5.8	10.53±0.08	51.9	77.8±3.9	218.2±12.0	10.11±0.21	52.0
GGU**GA** GGCU PCCG**AAG**CCG	82.9±2.0	233.5±6.3	10.43±0.08	52.3	78.5±3.3	220.2±10.1	10.25±0.14	52.4
GGU**GA** GGCU PCCA**AAG**CCG	79.7±1.9	224.9±5.8	9.96±0.07	51.0	76.2±2.0	214.0±6.2	9.83±0.07	51.0
GGU**GAA**GGCU PCCG**AG** UCG	74.9±2.6	213.8±8.0	8.57±0.06	45.7	71.2±5.4	202.4±17.1	8.45±0.13	45.6
GGU**GA** GGCU PCCG**AAG**UCG	67.8±2.8	192.5±8.8	8.13±0.04	44.5	62.5±2.2	175.3±7.0	8.08±0.07	44.9
GGU**AGA**GGCU PCCG**AG** CCG	75.5±2.0	213.4±6.1	9.27±0.06	48.7	72.2±3.3	203.4±10.4	9.15±0.13	48.7
3 × 3 internal loops								
GGU**GUA**GGCU PCCG**AAG**CCG	85.2±3.1	237.8±9.5	11.41±0.17	55.8	88.5±4.6	247.8±13.8	11.61±0.28	55.9
GGU**GAA**GGCU PCCG**AUG**CCG	88.4±2.2	249.5±6.8	11.07±0.10	53.8	82.9±3.4	232.3±10.5	10.81±0.14	53.9
GGU**AGA**GGCU PCCG**AAG**CCG	86.2±2.9	242.9±8.9	10.90±0.13	53.5	89.9±2.5	254.1±7.9	11.06±0.13	53.5
GGC**GAA**GGCU PCCG**AUG**CCG	78.2±2.0	214.6±6.2	11.66±0.12	58.8	75.1±5.0	205.1±15.3	11.52±0.30	59.1
2 × 4 internal loops								
GGU**GA** GGCU PCCG**AAGG**CCG	75.7±4.7	213.9±14.6	9.36±0.14	49.0	70.5±6.9	197.6±21.3	9.16±0.32	49.0

Table 1 (Continued)

	T_M^{-1} vs $\ln(C_T/4)$ plots (eq 1)				Average of melt curve fits			
Sequences	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)
GGU**GGAA**GGCU PCCG**AG** CCG	75.6±1.8	213.6±5.8	9.32±0.06	48.9	72.3±4.8	203.5±15.0	9.19±0.18	48.8
GGC**GAAA**GGCU PCCG**AG** CCG	68.2±3.1	188.3±9.5	9.82±0.14	52.7	62.0±6.6	168.9±20.7	9.62±0.22	53.3
3 × 4 internal loops								
GGU**GGAA**GGCU[b] PCCG**AAG** CCG	91.9±1.4	258.1±4.3	11.82±0.07	56.0	89.4±3.7	250.5±11.3	11.71±0.19	56.1
GGC**GGA** GGCU PCCG**AAGG**CCG	89.0±1.2	246.5±3.5	12.55±0.07	59.5	83.2±3.2	228.9±9.8	12.22±0.15	59.8
GGU**GGA** GGCU[b] PCCG**AAGG**CCG	91.4±3.0	257.4±9.1	11.56±0.15	55.0	85.4±3.9	238.9±11.9	11.26±0.22	55.2
GGC **GAA**GGCU PCCG**AAGG**CCG	86.2±2.6	238.5±7.8	12.22±0.16	58.9	79.5±3.5	218.2±10.8	11.84±0.18	59.2
GGU **GAA**GGCU PCCG**AAGG**CCG	89.9±1.6	253.8±4.8	11.16±0.07	53.9	82.0±2.5	229.4±7.7	10.81±0.16	54.1
GAGC**GGA** CGAC CUCG**AAGA**GCUG	93.3±3.7	266.5±11.5	10.61±0.15	51.2	92.8±3.6	265.1±11.3	10.60±0.16	51.2
GGC **AAA**GGCU PCCG**AAGG**CCG	77.2±2.1	213.5±6.4	10.96±0.11	55.9	74.8±3.8	206.1±11.6	10.85±0.21	56.0
GAGC **AGA**CGAC CUCG**AAAG**GCUG	88.7±6.4	254.3±20.1	9.80±0.24	48.9	88.7±8.7	254.4±27.0	9.82±0.37	48.9
GGC**GAAA**GGCU PCCG**AAG** CCG	70.4±3.5	193.1±10.9	10.45±0.18	55.3	69.9±7.0	191.4±21.5	10.50±0.34	55.7
GAGC**AAGA**CGAC CUCG **AAG**GCUG	78.3±4.0	222.0±12.5	9.44±0.13	49.0	82.7±3.6	235.9±11.0	9.59±0.20	48.9
GAGC**AGGA**CGAC CUCG **AAG**GCUG	81.6±3.3	232.7±10.5	9.42±0.10	48.4	78.9±5.7	224.2±17.9	9.34±0.20	48.5
GGU**AGA** GGCU PCCG**AAGG**CCG	72.7±2.8	205.4±8.7	9.01±0.09	48.0	64.4±8.7	179.6±26.9	8.68±0.36	47.7
GAGC**AGGA**CGAC CUCG**AUG** GCUG	78.2±2.0	223.3±6.3	8.96±0.05	46.9	88.5±4.5	255.7±14.5	9.18±0.11	46.6

Table 1 (Continued)

Sequences	T_M^{-1} vs ln(C_T/4) plots (eq 1)				Average of melt curve fits			
	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG°$_{37}$ (kcal/mol)	T_m[a] (°C)	−ΔH° (kcal/mol)	−ΔS° (eu)	−ΔG°$_{37}$ (kcal/mol)	T_m[a] (°C)
GAGC**AGAG**CGAC CUCG **AGA**GCUG	52.8±2.9	144.7±9.1	7.94±0.07	45.5	47.0±10.0	126.2±31.5	7.83±0.33	45.8
2 × 5 internal loops								
GGU**GA**___GGCU PCCG**AAGGA**CCG	79.3±6.1	229.4±19.4	8.14±0.15	43.4	68.7±5.9	196.0±18.2	7.95±0.26	43.5
GGC**GA**___GGCU[c] PCCG**AGUAA**CCG	54.4±2.5	147.7±7.8	8.55±0.09	49.0	53.1±5.2	143.6±16.3	8.52±0.14	49.1
4 × 4 internal loops								
GGU**GGAA**GGCU[b] PCCG**AAGG**CCG	108.1±4.1	300.6±12.4	14.91±0.30	63.0	108.5±3.1	301.8±9.4	14.93±0.25	63.0
GGC**GGAU**GGCU[b] PCCG**AAGU**CCG	105.4±1.7	289.3±5.1	15.71±0.14	66.6	108.6±2.6	298.6±7.5	15.97±0.23	66.5
GGC**GAAA**GGCU PCCG**AAGG**CCG	89.6±2.4	248.8±7.3	12.45±0.15	58.9	86.0±6.4	237.8±19.4	12.29±0.36	59.3
GAGC**AGGA**CGAC CUCG**AAAG**GCUG	94.9±2.7	271.6±8.5	10.65±0.10	51.1	92.1±2.8	262.9±8.6	10.56±0.11	51.2
GAGC**AAGA**CGAC CUCG**AAAG**GCUG	91.0±4.0	261.7±12.5	9.83±0.13	48.7	87.9±5.9	251.9±18.3	9.72±0.27	48.7
CGC**GAAA**GGC GCG**AAAG**CCG	54.7±2.5	152.0±7.9	7.56±0.05	42.9	47.1±8.9	127.2±29.2	7.62±0.12	44.3
GAGC**AGAG**CGAC CUCG**AAGA**GCUG	81.7±2.8	235.5±8.8	8.61±0.05	45.1	79.8±4.3	229.8±13.6	8.57±0.08	45.1
CGC**AAAA**GGC GCG**AAAA**CCG	38.3±1.8	103.3±7.0	6.29±0.05	35.0	37.3±7.3	99.5±20.1	6.47±0.01	36.5
GAGC**AAAG**CGAC CUCG**AAGA**GCUG	73.9±2.6	212.9±8.3	7.80±0.04	42.4	69.8±5.2	199.9±16.5	7.75±0.09	42.5
CGG**AAAA**CGC GCC**AAAA**GCG	31.6±1.8	87.4±6.2	4.48±0.15	18.1	30.8±9.9	83.8±34.7	4.78±0.81	20.3
3 × 5 internal loops								
GGC**GGA**__GGCU PCCG**AAGGA**CCG	83.2±3.0	233.6±9.1	10.75±0.14	53.6	72.5±5.0	200.6±15.7	10.28±0.16	53.9
GGU**GGA**__GGCU PCCG**AAGGA**CCG	81.2±1.7	230.8±5.2	9.57±0.05	49.1	77.2±5.7	218.5±17.6	9.42±0.27	49.1

Table 1 (Continued)

Sequences	T_M^{-1} vs $\ln(C_T/4)$ plots (eq 1)				Average of melt curve fits			
	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)
GGC**GAA**__GGCU PCCG**AAGGA**CCG	71.0±2.2	197.4±6.8	9.79±0.09	51.9	65.5±5.4	180.2±17.1	9.63±0.17	52.4
GGU**GAA**__GGCU PCCG**AAGGA**CCG	78.6±2.8	224.7±8.7	8.86±0.06	46.5	74.2±4.6	211.2±14.5	8.74±0.11	46.5
GGC**AAA**__GGCU PCCG**AAGGA**CCG	66.8±2.4	185.9±7.6	9.10±0.08	49.4	65.8±3.4	183.1±10.8	9.06±0.14	49.4
GGU**AGA**__GGCU PCCG**AAGGA**CCG	76.0±5.7	219.4±18.3	7.96±0.11	43.0	68.0±6.9	193.9±21.7	7.87±0.21	43.2
2 × 6 internal loops								
GGC**GA**____GGCU PCCG**AAAAAA**CCG	68.4±5.9	191.2±18.5	9.12±0.24	49.2	64.0±6.1	177.4±19.2	8.95±0.24	49.2
4 × 5 internal loops								
GGU**GGAA**_GGCU[d] PCCG**AAGGA**CCG	91.5±3.5	259.7±10.8	10.93±0.15	52.7	88.1±2.9	249.2±9.1	10.78±0.16	52.7
GGC**GAAA**_GGCU PCCG**AAGGA**CCG	69.0±2.7	190.6±8.4	9.89±0.12	52.9	68.7±4.6	189.5±14.2	9.93±0.19	53.1
3 × 6 internal loops								
GGU**GGA**___GGCU[b,e] PCCG**AAGUUU**CCG	85.3±6.7	239.7±20.5	10.97±0.32	54.0	89.6±3.2	253.0±10.0	11.16±0.21	53.9
GGC**GGA**___GGCU[b] PCCG**AAGUUU**CCG	86.9±3.5	242.1±10.6	11.77±0.18	56.9	81.3±8.2	225.1±25.2	11.54±0.43	57.4
GGU**GGA**___GGCU[b] PCCG**AAGAAA**CCG	90.8±1.9	259.1±6.0	10.47±0.07	51.1	84.5±4.2	239.4±13.0	10.25±0.19	51.3
GGC**GGA**___GGCU[b] PCCG**AAGAAA**CCG	72.9±4.0	202.5±12.3	10.09±0.16	52.9	66.0±5.8	181.1±18.2	9.84±0.20	53.4
GGC**GAA**___GGCU PCCG**AAGAAA**CCG	66.6±2.3	185.5±7.2	9.06±0.07	49.3	58.5±7.2	159.9±22.9	8.88±0.16	50.0
GGC**GGA**___GGCU PCCG**AAAAAA**CCG	56.0±3.7	152.4±11.6	8.76±0.14	49.9	53.6±6.9	144.5±21.7	8.78±0.17	50.6
GGC**GAA**___GGCU PCCG**AAAAAA**CCG	59.9±3.0	165.2±9.3	8.71±0.10	48.7	53.8±5.4	145.9±17.0	8.58±0.18	49.3
GGC**GGA**___GGCU PCCG**AGAAAA**CCG	57.8±5.6	158.9±17.4	8.56±0.24	48.3	50.1±6.3	134.4±19.7	8.41±0.26	49.1

Table 1 (Continued)

Sequences	T_M^{-1} vs $\ln(C_T/4)$ plots (eq 1)				Average of melt curve fits			
	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)
GGC**AAA**GGCU PCCG**AAAAAA**CCG	58.2±2.5	160.4±7.9	8.47±0.07	47.7	60.5±7.0	167.6±22.3	8.51±0.20	47.5
GGU**GUA**GGCU PCCG**AAAAAA**CCG	61.2±5.8	173.6±18.5	7.39±0.17	41.4	56.0±3.5	156.9±11.5	7.37±0.16	41.7
4 × 6 internal loops								
GGU**GGAA**GGCU[b] PCCG**AAGAAA**CCG	90.3±2.2	259.3±6.9	9.88±0.07	48.9	82.4±3.9	234.6±12.1	9.64±0.15	49.2
GGU**GGAA**GGCU PCCG**AAAAAA**CCG	77.9±2.4	223.9±7.6	8.46±0.04	44.9	72.0±6.2	205.3±19.3	8.32±0.22	44.9
GGC**GAAA**GGCU PCCG**AAAAAA**CCG	59.5±3.8	163.7±12.0	8.72±0.14	48.8	54.9±7.4	149.0±23.1	8.64±0.25	49.4
GGC**GAAA**GGCU PCCG**AGAAAA**CCG	53.5±5.5	145.1±17.3	8.49±0.24	48.8	48.8±7.8	130.4±24.5	8.41±0.22	49.4

[a] At C_T = 0.1 mM. [b] Imino proton spectra (Figure 2) are consistent with secondary structure shown. [c] Kink-turn in U4 snRNA (*17*, *22*). [d] Kt-58 (*17*). [e] Predicted to be kink-turn in helix 78 of *E. coli* 23S rRNA (*17*, *50*).

loop size, and from the most stable to least stable according to measured loop stability at 37 °C, $\Delta G°_{37,\text{loop}}$. The $\Delta G°_{37,\text{loop}}$ value is also often put in parentheses following each duplex or internal loop in Results and Discussion.

Models for Predicting Thermodynamic Stabilities of Medium-Size RNA Internal Loops. Measured thermodynamic results reported here and previous data on 2 × 3, 2 × 4, and 3 × 3 internal loops (*28*−*30*, *33*, *34*, *68*) can be compared to predictions from the model in the current RNAstructure 4.0 algorithm (*9*), which is also similar to that used in MFOLD (*8*):

$$\Delta G°_{\text{predicted}} = \Delta G°_{\text{loop initiation}}(n) + \text{m1}\Delta G°_{\text{AU/GU penalty}} + |\text{n2} - \text{n1}|\Delta G°_{\text{asym}} + \text{m2}\Delta G°_{\text{UU bonus}} + \text{m3}\Delta G°_{5'\text{YA}/3'\text{RG bonus}} + \text{m4}\Delta G°_{\text{GG bonus}} + \text{m5}\Delta G°_{5'\text{YG}/3'\text{RA bonus}} + \text{m5}'\Delta G°_{5'\text{RG}/3'\text{YA bonus}} \quad (3a)$$

or

$$\Delta G°_{\text{predicted}} = \Delta G°_{\text{loop initiation}}(n) + \text{m1}\Delta G°_{\text{AU/GU penalty}} + |\text{n2} - \text{n1}|\Delta G°_{\text{asym}} + \text{m2}\Delta G°_{\text{UU bonus}} + \text{m3}\Delta G°_{\text{AG bonus}} + \text{m4}\Delta G°_{\text{GG bonus}} + \text{m5}\Delta G°_{\text{GA bonus}} \quad (3b)$$

Here eqs 3a and 3b are for 2 × 3 loops (*28*) and larger loops, respectively (*9*). $\Delta G°_{\text{loop initiation}}(n)$ is the free energy for initiating an internal loop with n nucleotides that is closed by two GC/CG pairs, $\Delta G°_{\text{AU/GU penalty}}$ is the penalty for replacing a closing GC/CG pair with an AU/UA or GU/UG pair, m1 to m5′ are 1 or 2, n1 and n2 are the number of nucleotides on each side of the loop (n = n1 + n2, n1 ≤ n2), and $\Delta G°_{\text{XW bonus}}$ terms are increments applied for particular first noncanonical pairs with X on the 3′ side and W on the 5′ side of the adjacent canonical helix. $\Delta G°_{5'\text{YX}/3'\text{RW bonus}}$ is applied for an XW first noncanonical pair adjacent to a YR canonical pair (defined as UG, UA, or CG with the pyrimidine on the 5′ side of the XW noncanonical pair).

The data in Table 2 and previously published (*28*−*30*, *33*, *34*, *68*) (Table S2, Supporting Information) were fit to eqs 3a and 3b. Comparison with measured values in Table 2 and those previously published gives R^2 = 0.92 and a standard deviation of 0.36 kcal/mol for 2 × 3 loops and R^2 = 0.47 and a standard deviation of 1.11 kcal/mol for larger loops. The good fit for 2 × 3 loops suggests that eq 3a is a good model, so the new data were only used to slightly revise the previous parameters (Table 3). In contrast, the poor fit for loops larger than 2 × 3 (Figure 1) suggests that eq 3b can be improved for such loops.

A previous study of 3 × 3 internal loops concluded that additional terms should be added to eq 3b: $\Delta G°_{\text{middle bonus}}$ for 3 × 3 loops with a middle pair of GA and at least one non-pyrimidine−pyrimidine first noncanonical pair, and $\Delta G°_{5'\text{GU}/3'\text{AN penalty}}$ for 3 × 3 loops with a single first noncanonical GA pair that has a U 3′ to the G of the GA pair (*34*). With the exception of the loop in $\frac{\text{GGC}\ \underline{\text{GGAU}}\ \text{GGCU}}{\text{PCCG}\ \underline{\text{AAGU}}\ \text{CCG}}$ that was omitted, the data for loops larger than 2 × 3 in Table 2 and in previously published sequences (*29*, *33*, *34*,

Table 2: Measured and Predicted Thermodynamic Parameters for Internal Loop Formation in 1 M NaCl, pH7[a]

Sequence	$\Delta G^\circ_{37,\ loop}$ (kcal/mol)	ΔH°_{loop} (kcal/mol)	ΔS°_{loop} (eu)
2 × 3 internal loops			
GGC**GA** GGCU PCCG**AAG**CCG	−0.37±0.66 (−0.23)	−12.4±10.6	−38.7±32.3
GGC**GAA**GGCU PCCG**AG** CCG	−0.17±0.62 (−0.23)	−10.4±9.8	−32.8±30.1
GGC**GGA**GGCU PCCG**AG** CCG	−0.16±0.67 (−0.23)	−10.3±10.8	−32.7±33.0
GGU**GAA**GGCU PCCG**AG** CCG	−0.06±0.53 (0.50)	−13.9±9.4	−44.6±28.9
GGU**GGA**GGCU PCCG**AG** CCG	0.11±0.54 (0.50)	−17.2±9.5	−55.8±29.0
GGU**GA** GGCU PCCG**AAG**CCG	0.21±0.54 (0.50)	−12.9±9.5	−42.0±29.1
GGU**GA** GGCU PCCA**AAG**CCG	0.41±0.52 (0.50)	−10.9±7.9	−36.4±24.2
GGU**GAA**GGCU PCCG**AG** UCG	0.92±0.60 (1.23)	−6.2±10.0	−22.8±30.5
GGU**GA** GGCU PCCG**AAG**UCG	1.36±0.60 (1.23)	0.9±10.1	−1.5±30.7
GGU**AGA**GGCU PCCG**AG** CCG	1.37±0.53 (1.91)	−5.5±9.4	−21.9±29.0
3 × 3 internal loops			
GGU**GGA**GGCU[b,f] PCCG**AAG**CCG	−2.62±0.78 (−2.39)	−24.3±12.4	−69.7±37.5
GGU**GGA**GGCU[b,f] PCCA**AAG**CCG	−2.27±0.59 (−1.44)	−23.9±9.7	−69.5±29.5
GGC**GGA**GGCU[b,f] PCCG**AAG**UCG	−2.00±0.77 (−2.39)	−18.9±11.8	−54.5±35.9
GGU**GUA**GGCU PCCG**AAG**CCG	−0.77±0.56 (−0.03)	−15.2±9.7	−46.3±29.9
GGU**GAA**GGCU[b,f] PCCG**AAG**CCG	−0.48±0.57 (−0.03)	−14.2±11.1	−44.2±34.0
GGU**GAA**GGCU PCCG**AUG**CCG	−0.43±0.54 (−0.03)	−18.4±9.5	−58.0±29.2
GGC**GAA**GGCU[b,f] PCCG**AAG**CCG	−0.37±0.76 (0.18)	−8.9±11.9	−27.5±36.4
GGU**AGA**GGCU PCCG**AAG**CCG	−0.26±0.54 (0.65)	−16.2±9.6	−51.4±29.7

Table 2 (Continued)

Sequence	$\Delta G^\circ_{37,\ loop}$ (kcal/mol)	ΔH°_{loop} (kcal/mol)	ΔS°_{loop} (eu)
GGC**GAA**GGCU PCCG**AUG**CCG	−0.11±0.62 (0.18)	−5.9±9.9	−18.7±30.3
2 × 4 internal loops			
GGU**GA** GGCU PCCG**AAGG**CCG	1.28±0.55 (0.87)	−5.7±10.4	−22.4±31.9
GGU**GGAA**GGCU PCCG**AG** CCG	1.32±0.53 (0.87)	−5.6±9.4	−22.1±28.9
GGC**GAAA**GGCU PCCG**AG** CCG	1.73±0.62 (1.08)	4.1±10.1	7.6±31.1
3 × 4 internal loops			
GGU**GGAA**GGCU[f] PCCG**AAG** CCG	−1.18±0.53 (−0.78)	−21.9±9.4	−66.6±28.7
GGC**GGA** GGCU PCCG**AAGG**CCG	−1.00±0.61 (−0.57)	−16.7±9.8	−50.6±29.9
GGU**GGA** GGCU[f] PCCG**AAGG**CCG	−0.92±0.55 (−0.78)	−21.4±9.7	−65.9±29.8
GGC **GAA**GGCU PCCG**AAGG**CCG	−0.67±0.63 (−0.57)	−13.9±10.0	−42.6±30.7
GGU **GAA**GGCU PCCG**AAGG**CCG	−0.52±0.53 (0.17)	−19.9±9.4	−62.3±28.8
GAGC**GGA** CGAC CUCG**AAGA**GCUG	0.07±0.59 (−0.57)	−26.6±10.4	−85.2±31.3
GGC **AAA**GGCU PCCG**AAGG**CCG	0.59±0.61 (0.61)	−4.9±9.9	−17.6±30.3
GAGC **AGA**CGAC CUCG**AAAG**GCUG	0.88±0.61 (0.61)	−22.0±11.5	−73.0±35.4
GGC**GAAA**GGCU PCCG**AAG** CCG	1.10±0.63 (0.88)	1.9±10.3	2.8±31.6
GAGC**AAGA**CGAC CUCG **AAG**GCUG	1.24±0.58 (0.61)	−11.6±10.4	−40.7±31.7
GAGC**AGGA**CGAC CUCG **AAG**GCUG	1.26±0.58 (0.61)	−14.9±10.2	−51.4±31.0
GGU**AGA** GGCU PCCG**AAGG**CCG	1.63±0.54 (2.53)	−2.7±9.7	−13.9±29.7
GAGC**AGGA**CGAC CUCG**AUG** GCUG	1.72±0.57 (1.79)	−11.5±9.9	−42.0±29.8
GAGC**AGAG**CGAC CUCG **AGA**GCUG	2.74±0.57 (2.70)	13.9±10.1	36.6±30.5

Table 2 (Continued)

Sequence	$\Delta G^{\circ}_{37,\ loop}$ (kcal/mol)	ΔH°_{loop} (kcal/mol)	ΔS°_{loop} (eu)
2 × 5 internal loops			
GGU**GA**___GGCU PCCG**AAGGA**CCG	2.50±0.55 (2.48)	−9.3±11.1	−37.9±34.4
GGC**GA**___GGCU[c] PCCG**AGUAA**CCG	3.00±0.62 (2.69)	17.9±10.0	48.2±30.7
4 × 4 internal loops			
GGU**GGAA**GGCU[f] PCCG**AAGG**CCG	−4.27±0.61 (−3.31)	−38.1±10.1	−109.1±31.0
GGC**GGAU**GGCU[f] PCCG**AAGU**CCG	−4.16±0.62 (−1.52)	−33.1±9.8	−93.4±30.1
GGC**GAAA**GGCU PCCG**AAGG**CCG	−0.90±0.63 (−0.74)	−17.3±10.0	−52.9±30.5
GAGC**AGGA**CGAC CUCG**AAAG**GCUG	0.03±0.57 (0.17)	−28.2±10.0	−90.3±30.3
GAGC**AAGA**CGAC CUCG**AAAG**GCUG	0.85±0.58 (0.17)	−24.3±10.4	−80.4±31.7
CGC**GAAA**GGC GCG**AAAG**CCG	0.96±0.07 (0.44)	−5.6±1.9	−21.3±7.3
GAGC**AGAG**CGAC CUCG**AAGA**GCUG	2.07±0.33 (2.26)	−15.0±10.1	−54.2±30.4
CGC**AAAA**GGC GCG**AAAA**CCG	2.23±0.07 (2.26)	10.8±1.9	27.4±7.3
GAGC**AAAG**CGAC CUCG**AAGA**GCUG	2.88±0.57 (2.26)	−7.2±10.0	−31.6±30.3
CGG**AAAA**CGC GCC**AAAA**GCG	3.01±0.17 (2.26)	11.4±2.2	27.1±7.2
3 × 5 internal loops			
GGC**GGA**__GGCU PCCG**AAGGA**CCG	0.80±0.62 (−0.11)	−10.9±10.1	−37.7±31.0
GGU**GGA**__GGCU PCCG**AAGGA**CCG	1.07±0.53 (−0.32)	−11.2±9.4	−39.3±28.8
GGC**GAA**__GGCU PCCG**AAGGA**CCG	1.76±0.61 (2.25)	1.3±9.9	−1.5±30.4
GGU**GAA**__GGCU PCCG**AAGGA**CCG	1.78±0.53 (2.04)	−8.6±9.7	−33.2±29.7
GGC**AAA**__GGCU PCCG**AAGGA**CCG	2.45±0.61 (3.16)	5.5±10.0	10.0±30.6

Table 2 (Continued)

Sequence	$\Delta G°_{37,\ loop}$ (kcal/mol)	$\Delta H°_{loop}$ (kcal/mol)	$\Delta S°_{loop}$ (eu)
GGU**AGA** GGCU PCCG**AAGGA**CCG	2.68±0.54 (3.90)	−6.0±10.8	−27.9±33.7
2 × 6 internal loops			
GGC**GA** GGCU PCCG**AAAAAA**CCG	2.43±0.65 (3.15)	3.9±11.3	4.7±35.0
4 × 5 internal loops			
GGU**GGAA** GGCU[d] PCCG**AAGGA**CCG	−0.29±0.55 (−0.70)	−21.5±9.9	−68.2±30.4
GGC**GAAA** GGCU PCCG**AAGGA**CCG	1.66±0.62 (1.87)	3.3±10.1	5.3±30.8
3 × 6 internal loops			
GGU**GGA** GGCU[e,f] PCCG**AAGUUU**CCG	−0.33±0.62 (0.20)	−15.3±11.4	−48.2±35.0
GGC**GGA** GGCU[f] PCCG**AAGUUU**CCG	−0.22±0.63 (0.41)	−14.6±10.3	−46.2±31.5
GGU**GGA** GGCU[f] PCCG**AAGAAA**CCG	0.17±0.54 (0.20)	−20.8±9.4	−67.6±29.0
GGC**GGA** GGCU[f] PCCG**AAGAAA**CCG	1.46±0.62 (0.41)	−0.6±10.4	−6.6±32.1
GGC**GAA** GGCU PCCG**AAGAAA**CCG	2.49±0.60 (2.77)	5.7±9.9	10.4±30.5
GGC**GGA** GGCU PCCG**AAAAAA**CCG	2.79±0.62 (2.77)	16.3±10.4	43.5±31.8
GGC**GAA** GGCU PCCG**AAAAAA**CCG	2.84±0.62 (2.77)	12.4±10.1	30.7±31.1
GGC**GGA** GGCU PCCG**AGAAAA**CCG	2.99±0.64 (2.77)	14.5±11.1	37.0±34.4
GGC**AAA** GGCU PCCG**AAAAAA**CCG	3.08±0.61 (3.68)	14.1±10.0	35.5±30.7
GGU**GUA** GGCU PCCG**AAAAAA**CCG	3.25±0.56 (2.56)	8.8±10.8	17.9±33.9
4 × 6 internal loops			
GGU**GGAA** GGCU[f] PCCG**AAGAAA**CCG	0.76±0.54 (0.32)	−20.3±9.5	−67.8±29.2
GGU**GGAA** GGCU PCCG**AAAAAA**CCG	2.18±0.53 (2.68)	−7.9±9.6	−32.4±29.4
GGC**GAAA** GGCU PCCG**AAAAAA**CCG	2.83±0.62 (2.89)	12.8±10.4	32.2±32.0
GGC**GAAA** GGCU PCCG**AGAAAA**CCG	3.06±0.64 (2.89)	18.8±11.0	50.8±34.4

[a] Calculated from eq 2a and data in Table 1 unless otherwise noted. Experimental error 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 *7*. Values in parentheses are $\Delta G°_{predicted}$, predicted according to eq 3a for 2 × 3 loops and eq 4 for other loops. [b] Data from ref *33*. [c] Kink-turn in U4 snRNA (*17*, *22*). [d] Kt-58 (*17*). [e] Predicted to be kink-turn in helix 78 of *E. coli* 23S rRNA (*17*, *50*). [f] Imino proton spectra (Figure 2) are consistent with secondary structure shown.

Table 3: Free Energy Parameters at 37 °C (kcal/mol) for Medium-Size Internal Loops in 1 M NaCl[a]

free energy parameters	2 × 3 loops	n1 + n2 > 5 loops
$\Delta G^\circ_{\text{loop initiation}}(5)$	2.15 ± 0.14[j]	
$\Delta G^\circ_{\text{loop initiation}}(6)$		2.00 ± 0.11
$\Delta G^\circ_{\text{loop initiation}}(7)$		2.25 ± 0.20
$\Delta G^\circ_{\text{loop initiation}}(8)$		2.26 ± 0.18
$\Delta G^\circ_{\text{loop initiation}}(9)$		2.33 ± 0.28
$\Delta G^\circ_{\text{loop initiation}}(10)$		2.90 ± 0.34
$\Delta G^\circ_{\text{AU/GU penalty}}$	0.73 ± 0.07	0.74 ± 0.11
$\Delta G^\circ_{\text{asym}}$	0.45 ± 0.08[k]	0.45 ± 0.08
$\Delta G^\circ_{\text{UU bonus}}$[b]	−0.34 ± 0.15	−0.51 ± 0.12
$\Delta G^\circ_{\text{5}'\text{YA/3}'\text{RG bonus}}$[b,c]	−0.39 ± 0.22	−0.65 ± 0.29
$\Delta G^\circ_{\text{GG bonus}}$[b]	−0.74 ± 0.28	
$\Delta G^\circ_{\text{5}'\text{YG/3}'\text{RA bonus}}$[b]	−1.41 ± 0.08	
$\Delta G^\circ_{\text{5}'\text{RG/3}'\text{YA bonus}}$[b]	−1.06 ± 0.17	
$\Delta G^\circ_{\text{GA bonus}}$[b]		−0.91 ± 0.08
$\Delta G^\circ_{\text{middle GA bonus (3×3 loop)}}$[d]		−1.07 ± 0.23
$\Delta G^\circ_{\text{5}'\text{GU/3}'\text{AN penalty (3×3 loop)}}$[e]		0.96 ± 0.25
$\Delta G^\circ_{\text{2×(5}'\text{GA/3}'\text{CG)bonus(3×3 loop)}}$[b,f]		−0.96 ± 0.42
$\Delta G^\circ_{\text{2GA bonus}}$[g]		−1.18 ± 0.16
$\Delta G^\circ_{\text{3GA bonus}}$[h]		−2.36 ± 0.15
$\Delta G^\circ_{\text{5}'\text{UG/3}'\text{GA bonus}}$[b,i]		−0.95 ± 0.16

[a] These parameters are used to predict the free energy of 2 × 3 (left) and larger (right) internal loops having more than one nucleotide on each side in 1 M NaCl according to eqs 3a and 4, respectively. [b] Applied for first noncanonical pair. [c] Applied for an AG first noncanonical pair adjacent to a YR canonical pair (defined as UG, UA, or CG with the pyrimidine on the 5′ side of AG pair). [d] Applied for 3 × 3 loops with a middle pair of GA and at least one non-pyrimidine-pyrimidine first noncanonical pair unless a $\Delta G^\circ_{\text{2GA bonus}}$ or $\Delta G^\circ_{\text{3GA bonus}}$ has been used. [e] Applied for 3 × 3 loops with a single first noncanonical GA pair that has a U 3′ to the G of the GA pair. [f] Applied for loops with two motifs of 5′GA/3′CG in 3 × 3 loops. Note that this parameter is only applied once to a loop. [g] Applied to loops with the motif 5′YGA/3′RAG, 5′RGA/3′YAG, 5′YGG/3′RAA, or 5′RGG/3′YAA in 3 × 3, 3 × 4, 4 × 4, and 4 × 5 loops (i.e., loops with the closing base pair 3′ to the A of a GA pair) unless the motif has been represented by a 3GA bonus. Note that this parameter is applied for 5′RGGA/3′YAAG or 5′GGAY/3′AAGR (i.e., $\frac{\text{GGA}}{\text{AAG}}$ not closed with at least one YR canonical pair), which are not represented by a 3GA bonus. This parameter is also applied for an unusually stable 4 × 4 loop, $\frac{\text{U}}{\text{G}}\frac{\text{GGAA}}{\text{AAGG}}\frac{\text{G}}{\text{C}}$ (see text). [h] Applied for loops with the motif of 5′YGGA/3′RAAG or 5′GGAR/3′AAGY (i.e., $\frac{\text{GGA}}{\text{AAG}}$ closed on at least one side with a YR canonical pair). [i] Applied for 3 × 3 and larger loops with the motif of 5′UG/3′GA. [j] Calculated from the fitted value (2.59 ± 0.11 kcal/mol) of $\Delta G^\circ_{\text{loop initiation}}(5)$ + $\Delta G^\circ_{\text{asym}}$ in 2 × 3 loops minus the fitted value (0.45 ± 0.08) of $\Delta G^\circ_{\text{asym}}$ in loops larger than 2 × 3. [k] Value fit in loops larger than 2 × 3.

68) (Table S2, Supporting Information) are fit well if four additional bonus parameters are added to give eq 4:

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

Only the first six parameters are currently included in structure prediction programs such as MFOLD (*8*) and RNAstructure (*9*). Here, $\Delta G^\circ_{2\times(5'\text{GA}/3'\text{CG})\text{ bonus (3×3 loop)}}$ is applied for loops with two motifs $\frac{\text{G A}}{\text{C G}}$ in 3 × 3 loops, e.g., $\frac{\text{CGG}}{\text{GCC}}\frac{\text{AAG}}{\text{GUA}}\frac{\text{CGC}}{\text{GCG}}$ (1.21 kcal/mol) and $\frac{\text{CGG}}{\text{GCC}}\frac{\text{AAG}}{\text{GAA}}\frac{\text{CGC}}{\text{GCG}}$ (0.87 kcal/mol) (Table S2, Supporting Information) (*34*). Note that this parameter is only applied once to a loop. Unless the motifs have been represented by a 3GA bonus, the $\Delta G^\circ_{\text{2GA bonus}}$ is applied for $X_3 \times W_{3-4}$ and $X_4 \times W_{4-5}$ loops with the motifs $\frac{\text{Y}}{\text{R}}\frac{\text{GA}}{\text{AG}}$, $\frac{\text{R}}{\text{Y}}\frac{\text{GA}}{\text{AG}}$, $\frac{\text{Y}}{\text{R}}\frac{\text{GG}}{\text{AA}}$, or $\frac{\text{R}}{\text{Y}}\frac{\text{GG}}{\text{AA}}$ (i.e., loops with the closing base pair 3′ to the A of a GA pair) and additionally for $X_3 \times W_{5-6}$ and $X_4 \times W_6$ loops with the motifs $\frac{\text{R}}{\text{Y}}\frac{\text{GGA}}{\text{AAG}}$ and $\frac{\text{GGA}}{\text{AAG}}\frac{\text{Y}}{\text{R}}$ (i.e., $\frac{\text{GGA}}{\text{AAG}}$ motif not closed with at least one YR canonical pair). The $\Delta G^\circ_{\text{3GA bonus}}$ is applied for loops with the motifs $\frac{\text{Y}}{\text{R}}\frac{\text{GGA}}{\text{AAG}}$ or $\frac{\text{GGA}}{\text{AAG}}\frac{\text{R}}{\text{Y}}$ (i.e., $\frac{\text{GGA}}{\text{AAG}}$ closed on at least one side with a YR canonical pair). If a $\Delta G^\circ_{\text{2GA bonus}}$ or $\Delta G^\circ_{\text{3GA bonus}}$ has been used, then the $\Delta G^\circ_{\text{middle GA bonus(3×3 loop)}}$ is not applied. The $\Delta G^\circ_{5'\text{UG}/3'\text{GA bonus}}$ is applied for each $\frac{\text{UG}}{\text{GA}}$ motif at a loop terminus in loops larger than 2 × 3, so m6 is 1 or 2. Table 3 lists the values of these fitted parameters. Attempts were made to fit the data with fewer parameters, but that always resulted in certain classes of sequences being predicted poorly. For example, at least three consecutive GA pairs are required to provide extra stability to loops with |n2 − n1| > 1, so the stabilizing effect of only two consecutive GA pairs is restricted to internal loops with |n2 − n1| < 2. This is apparently a non-nearest-neighbor effect. The detailed multiple linear regression analysis is given in Table S3, Supporting Information.

A $\Delta G^\circ_{\text{GG bonus}}$ is not used in eq 4 because only the loop in $\frac{\text{GAGC}}{\text{CUCG}}\frac{\text{GAG}}{\text{AAG}}\frac{\text{CGAC}}{\text{GCUG}}$ has a GG first noncanonical pair, and its stability is predicted well without including a GG bonus. A GG bonus has been found for 2 × 2 loops (*74*). The result for the 3 × 3 loop studied here suggests that any GG bonus is context dependent.

Listed in parentheses in Table 2 are the free energy increments at 37 °C for internal loops larger than 2 × 3 as predicted by eq 4 with the parameters from Table 3. The correlation with measured values gives $R^2 = 0.86$ and a standard deviation of 0.57 kcal/mol (Figure 1). The average absolute difference per nucleotide between measured and predicted values is 0.05 kcal/mol. Evidently, the new parameters in eq 4 are justified. The revised values for parameters that appear in both eqs 3 and 4 are within experimental error of those determined previously (*9*).

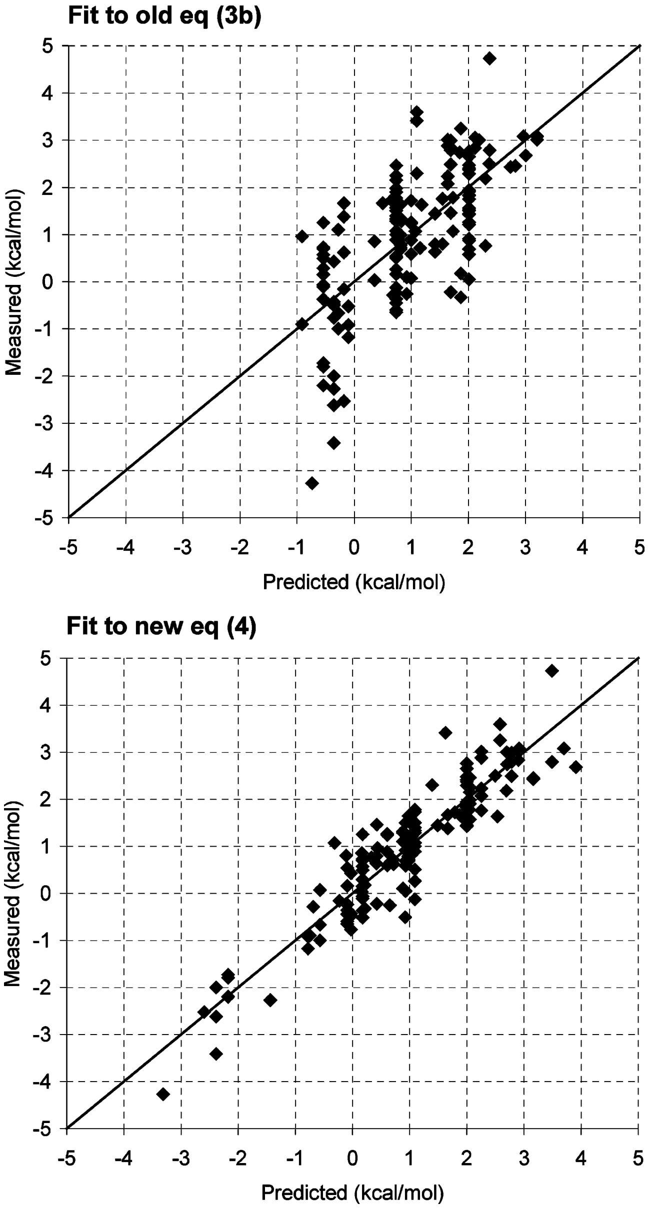

FIGURE 1: Comparisons between predicted and measured free energies for 3 × 3 and larger loops for model of eq 3b as used in current RNAstructure 4.0 program (*9*) (R^2 = 0.47, standard deviation = 1.11 kcal/mol) and model of eq 4 (R^2 = 0.86, standard deviation = 0.57 kcal/mol).

Exchangeable Proton and Phosphorus-31 NMR Spectra. For several loops with interesting stabilities and/or sequences expected to give interesting structures, 1D imino proton NMR spectra confirm that the expected canonical base pairs are present (Figure 2, left panel). Some preliminary assignments are based on NMR melting and comparison with similar duplexes having 3 × 3 internal loops (*33*). The 2D SNOESY spectra (Figure S1, Supporting Information) were also used to confirm assignments and secondary structure. The 1D ^{1}H-decoupled ^{31}P NMR spectra (Figure 2, right panel) were used to probe backbone structural features of the duplexes. Several unusual downfield ^{31}P resonances are likely due to the phosphorus residues at 5′GpA3′ nearest neighbors in 5′GA/3′AG motifs. Tandem GA pairs often have a trans ζ phosphate configuration that gives a downfield phosphorus resonance (*53*, *75*, *76*). These resonances are not observed for all loops with this motif, however, suggesting that the backbone structure and dynamics depend on context.

DISCUSSION

Thermodynamic models and parameters for internal loops are important for the prediction of RNA secondary structure (*8*, *9*, *77*−*82*). In turn, RNA secondary structure is the first step for modeling 3D structure (*10*−*12*, *21*, *83*, *84*) and facilitates interpretation of experimental studies, such as folding (*13*, *15*, *23*, *51*, *55*, *56*, *61*) and ribozyme kinetics (*85*−*87*). It may also allow prediction of sites suitable for rational design of therapeutics.

Loops of 2 × 3 Nucleotides. Loops with 2 × 3 nucleotides are predicted well by the previous thermodynamic model of eq 3a (*9*, *28*, *29*). Linear regression of measured free energy increments on 2 × 3 nucleotide loops reported here (Table 2) and previously (*28*−*30*) gives parameters within experimental error of those previously published (*9*, *28*).

An NMR structure of the 2 × 3 internal loop ${}^{\text{C}}_{\text{G}}\frac{\text{GA}}{\text{AAG}}{}^{\text{G}}_{\text{C}}$ revealed a unique structure with a "shared sheared GA" motif having Hoogsteen edges of two A's forming base pairs with one G (*88*). In contrast, flexibility was observed for the loop ${}^{\text{U}}_{\text{G}}\frac{\text{GA}}{\text{AAG}}{}^{\text{U}}_{\text{G}}$ by NMR, possibly due to the destabilizing effect of an $\frac{\text{AU}}{\text{GG}}$ motif as discussed below (*30*). Interestingly, the stabilities of ${}^{\text{C}}_{\text{G}}\frac{\text{GA}}{\text{AAG}}{}^{\text{G}}_{\text{C}}$ (averaging −0.10 kcal/mol) and ${}^{\text{U}}_{\text{G}}\frac{\text{GA}}{\text{AAG}}{}^{\text{U}}_{\text{G}}$ (1.90 kcal/mol) (Table 2, Supporting Information Table S2) (*29*, *30*) are predicted well by eq 3a with parameters from Table 3, which give values of −0.23 and 1.58 kcal/mol, respectively. Thus, even though the 3D structures are more complex, a simple thermodynamic model works well.

Loops Larger than 2 × 3 Nucleotides. The motif of three consecutive sheared GA pairs $\frac{\text{GGA}}{\text{AAG}}$ is the most stable among 3 × 3 internal loops (*33*, *34*). To explore the effect of consecutive GA pairs on stability of larger internal loops including kink-turns, sequences were studied with two and three potentially consecutive GA pairs. From the comparison of measured and predicted free energy increments for internal loops larger than 2 × 3 (Table 2 and Figure 1), the model of eq 4 is sufficient for purine-rich internal loops, with the exception of ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GGAU}}{\text{AAGU}}{}^{\text{GGCU}}_{\text{CCG}}$ (−4.16 kcal/mol). The terms in eq 4 are discussed below. A sample calculation is shown in Figure 3 for predicting the free energy for formation of a 3 × 3 internal loop. More sample calculations are shown in Figure S2, Supporting Information. In parentheses in Table 2 are predicted free energies for all the loops studied.

Bonus for Three Consecutive GA Pairs in X_{3-4} × W_{3-6} Internal Loops. All of the internal loops studied here with a ${}^{\text{Y}}_{\text{R}}\frac{\text{GGA}}{\text{AAG}}$ or $\frac{\text{GGA}}{\text{AAG}}{}^{\text{R}}_{\text{Y}}$ motif are more stable than expected from the model in the current MFOLD and RNAstructure 4.0 programs (eq 3b). Thus, a bonus parameter for the $\frac{\text{GGA}}{\text{AAG}}$ motif is included in eq 4. The bonus value of −2.36 kcal/mol is approximately twice that for first noncanonical GA pairs, $\Delta G^\circ_{\text{GA bonus}}$ (−0.91 kcal/mol) and for two consecutive GA pairs, $\Delta G^\circ_{\text{2GA bonus}}$ (−1.18 kcal/mol), presumably reflecting interactions of two nearest neighbors in $\frac{\text{GGA}}{\text{AAG}}$. Note that 3 × 3 loops with potentially three consecutive GA pairs are given a $\Delta G^\circ_{\text{3GA bonus}}$ and two $\Delta G^\circ_{\text{GA bonus}}$ but no $\Delta G^\circ_{\text{middle GA bonus}}$. The $\Delta G^\circ_{\text{3GA bonus}}$ accounts for the stacking between GA pairs in $\frac{\text{GGA}}{\text{AAG}}$ and the two $\Delta G^\circ_{\text{GA bonus}}$ increments account for stacking between each first noncanonical GA pair and adjacent closing base pair.

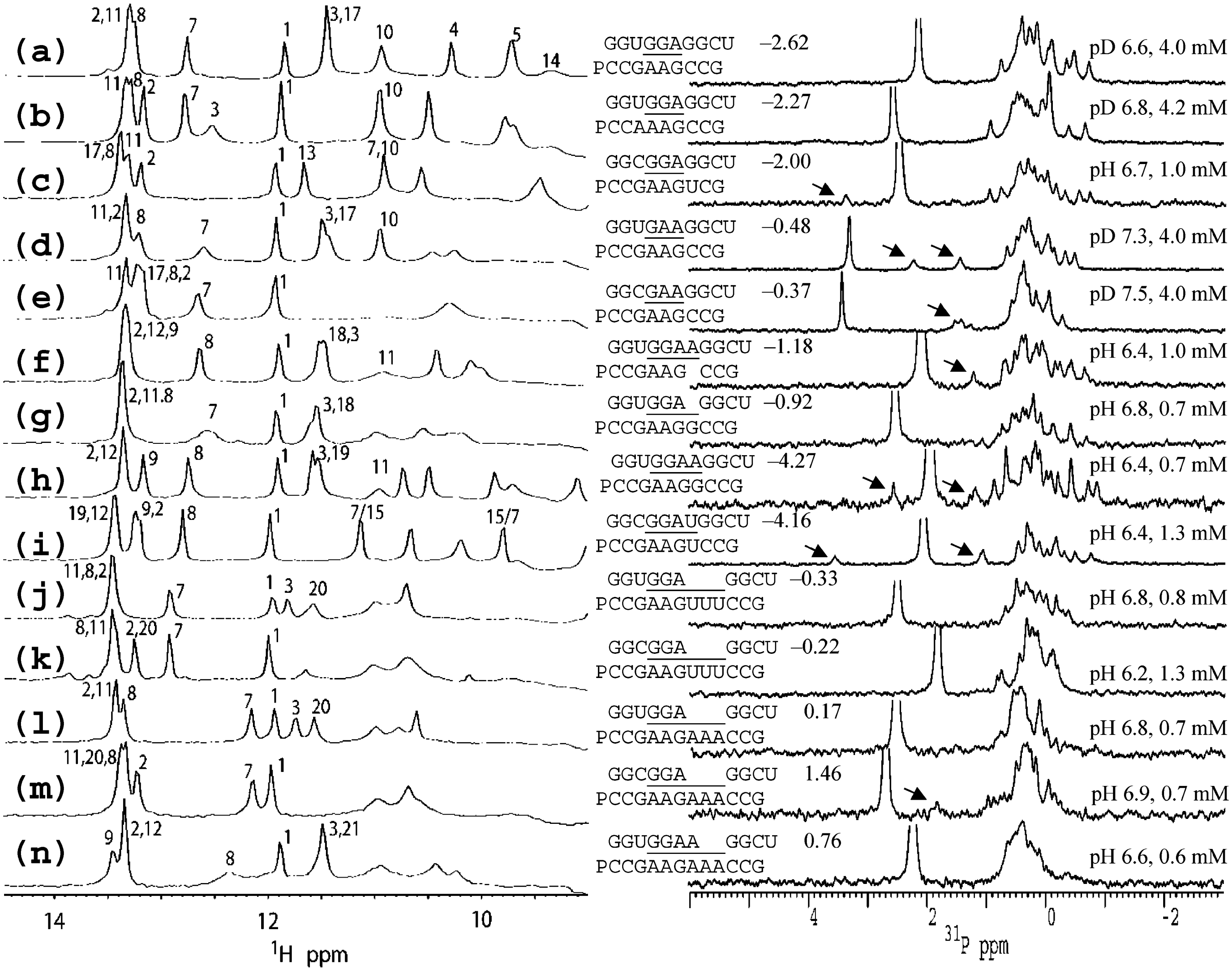

FIGURE 2: NMR spectra of (a) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$ (0 °C (*33*)), (b) ${}^{\text{GGU}}_{\text{PCCA}}\frac{\text{GGA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$, (c) ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{UCG}}$, (d) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$, (e) ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$, (f) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$, (g) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAAG}}{}^{\text{GGCU}}_{\text{CCG}}$, (h) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$, (i) ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GGAU}}{\text{AAGU}}{}^{\text{GGCU}}_{\text{CCG}}$ (10 °C), (j) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGUUU}}{}^{\text{GGCU}}_{\text{CCG}}$, (k) ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGUUU}}{}^{\text{GGCU}}_{\text{CCG}}$, (l) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGAAA}}{}^{\text{GGCU}}_{\text{CCG}}$, (m) ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGAAA}}{}^{\text{GGCU}}_{\text{CCG}}$, and (n) ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAGAAA}}{}^{\text{GGCU}}_{\text{CCG}}$, in 80 mM NaCl, 10 mM sodium phosphate, 0.5 mM Na_2EDTA. Left panel: One-dimensional imino proton spectra in 90:10 (v:v) H_2O/D_2O at 5 °C unless noted otherwise. The sample pHs are given in the right panel, except for spectra (a) pH 5.9 (b) pH 5.4, (d) pH 5.1, and (e) pH 6.7. Numbers on spectra correspond to assignments with numbering starting at left most (5′) nucleotide of top strand and ending at left most (3′) nucleotide of bottom strand. Right panel: The 1D ^{1}H-decoupled ^{31}P spectra at 30 °C in 90:10 (v:v) H_2O/D_2O or 100% D_2O. The spectra were referenced to external standard at 30 °C of 85% H_3PO_4 at 0 ppm. Note that the chemical shift of the phosphate resonance (the highest peak except for (d)) depends on pH or pD. The downfield resonances are labeled with arrows. Values adjacent to sequences are $\Delta G^\circ_{37,\text{loop}}$ (kcal/mol) from Table 2.

Loop	ΔG°$_{37\ \text{loop}}$ kcal/mol	ΔG°$_{\text{predicted}}$ kcal/mol
5'UGGAG3' 3'GAAGC5'	-2.62	-2.39

$\Delta G^\circ_{\text{predicted}} =$

$$\Delta G^\circ_{\text{loop initiation}}(6) + \Delta G^\circ_{\text{AU/GU penalty}} + 2\Delta G^\circ_{\text{GA bonus}} + \Delta G^\circ_{\text{3GA}} + \Delta G^\circ_{\text{5'UG/3'GA}}$$

$$= 2.00 \quad + \quad 0.74 \quad + \quad 2(-0.91) \quad + (-2.36) \; + \; (-0.95)$$

$$= -2.39 \text{ kcal/mol}$$

FIGURE 3: A sample calculation for predicting the free energy at 37 °C for formation of an internal loop. Additional sample calculations are shown in Supporting Information.

Note that besides the $\Delta G^\circ_{\text{3GA bonus}}$, only one bonus parameter of a first noncanonical GA pair, $\Delta G^\circ_{\text{GA bonus}}$, is applied for 3 × 4 loops, ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−1.18 kcal/mol), ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−1.00 kcal/mol), ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−0.92 kcal/mol), ${}^{\text{GGC}}_{\text{PCCG}}\frac{\text{GAA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−0.67 kcal/mol), and ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GAA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−0.52 kcal/mol) because the 3GA bonus is more favorable than a second first noncanonical GA bonus coupled with a 2GA bonus, and formation of three consecutive GA pairs would preclude formation of a second first noncanonical GA pair. Similarly, $\Delta G^\circ_{\text{5'UG/3'GA bonus}}$ was not applied for the loop in ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GAA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−0.52 kcal/mol) because the $\frac{\text{GGA}}{\text{AAG}}$ motif was assumed to be adjacent to the closing CG pair. The imino proton resonances of G8 in ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−1.18 kcal/mol) (Figure 2f) and G7 in ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−0.92 kcal/mol) (Figure 2g) are relatively broader than G7 in ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−2.62 kcal/mol) (Figure 2a) (*33*) and ${}^{\text{GGU}}_{\text{PCCA}}\frac{\text{GGA}}{\text{AAG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−2.27 kcal/mol) (Figure 2b) and G8 in ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$ (−4.27 kcal/mol) (Figure 2h). This is consistent with the stacking assumed in the thermodynamic model.

The 4 × 4 loop in ${}^{\text{GGU}}_{\text{PCCG}}\frac{\text{GGAA}}{\text{AAGG}}{}^{\text{GGCU}}_{\text{CCG}}$(−4.27 kcal/mol) is exceptionally stable but is predicted reasonably well (−3.31 kcal/mol) by applying $\Delta G^\circ_{\text{3GA bonus}}$, $\Delta G^\circ_{\text{2GA bonus}}$, $2\Delta G^\circ_{\text{GA bonus}}$

and $\Delta G^\circ_{5'\mathrm{UG}/3'\mathrm{GA\,bonus}}$. Applying all these bonuses is consistent with the total of five nearest neighbor interactions observed in the crystal structure of a similar loop $\begin{smallmatrix}\text{C GGAA G}\\ \text{G AAGG C}\end{smallmatrix}$ (*89*). The free energy difference between measurement and prediction may be due to highly coupled base stacking and hydrogen-bonding interactions in this loop as indicated by sharp imino proton resonances (Figure 2h). No $\Delta G^\circ_{\mathrm{2GA\,bonus}}$ is added to $\Delta G^\circ_{\mathrm{3GA\,bonus}}$ for the 4 × 5 loop in $\begin{smallmatrix}\text{GGU GGAA GGCU}\\ \text{PCCG AAGGA CCG}\end{smallmatrix}$ (−0.29 kcal/mol, kt-58 (*17*)), however, presumably due to its asymmetry.

The exceptionally stable 4 × 4 loop $\begin{smallmatrix}\text{GGC GGAU GGCU}\\ \text{PCCG AAGU CCG}\end{smallmatrix}$ (−4.16 kcal/mol) was not included in the linear regression analysis. The ^{31}P spectrum for this duplex (Figure 2i) shows large dispersion, similar to that observed for the duplex $\begin{smallmatrix}\text{GGU GGAA GGCU}\\ \text{PCCG AAGG CCG}\end{smallmatrix}$ (Figure 2h), which also has an exceptionally stable 4 × 4 loop (−4.27 kcal/mol). This suggests that the terminal UU pair also supports a favorable, rigid structure. The 1D imino (Figure 2i) and 2D SNOESY (Figure S1, Supporting Information) spectra indicate formation of a UU pair (with two U imino protons hydrogen bonded to carbonyl groups) (cis Watson−Crick/Watson−Crick UU) besides three GA pairs. This is also observed in conserved loops $\begin{smallmatrix}\text{C GGAU U}\\ \text{G AAGU A}\end{smallmatrix}$ in helix 42 of small subunit rRNA (*90*), $\begin{smallmatrix}\text{G UGGAU U}\\ \text{U AAAGU A}\end{smallmatrix}$ in helix 2 of large subunit rRNA (*91*), and a kink-turn loop $\begin{smallmatrix}\text{C UGA C}\\ \text{G UAGUGC G}\end{smallmatrix}$ (*20*). In the UU pairs of these structures, the U's 3′ and 5′ of the adjacent closing base pair are shifted to the major and minor groove, respectively, which is favorable for base stacking with the adjacent sheared GA pair. Note that an $\frac{\text{AU}}{\text{GU}}$ nearest neighbor is not thermodynamically favorable in 2 × 2 loops (*39*), probably because there is geometric incompatibility when a UU pair is adjacent to an imino AG pair (cis Watson−Crick/Watson−Crick AG). An imino AG pair forms when the A of an AG pair is 3′ of the closing Watson−Crick pair, as is the case in an $\frac{\text{AU}}{\text{GU}}$ 2 × 2 loop.

On the basis of NMR spectra, three consecutive sheared GA pairs form in two 3 × 6 internal loops, $\begin{smallmatrix}\text{GGU GGA GGCU}\\ \text{PCCG AAGUUU CCG}\end{smallmatrix}$ (predicted to be a kink-turn in helix 78 of *Escherichia coli* 23 rRNA (*17*, *50*)) and $\begin{smallmatrix}\text{GGC GGA GGCU}\\ \text{PCCG AAGUUU CCG}\end{smallmatrix}$, instead of two GU and one AU pair $\begin{smallmatrix}\text{GGU GGA GGCU}\\ \text{PCCG AAGUUU CCG}\end{smallmatrix}$ and $\begin{smallmatrix}\text{GGC GGA GGCU}\\ \text{PCCG AAGUUU CCG}\end{smallmatrix}$. There are no imino proton resonances that indicate the formation of $\frac{\text{GGA}}{\text{UUU}}$ in the 1D proton (Figure 2j,k) and 2D SNOSY spectra (Figure S1j,k, Supporting Information). This is further confirmed by the relatively small changes in 1D imino proton spectra when the UUU triplets are replaced by AAA (compare $\begin{smallmatrix}\text{GGC GGA GGCU}\\ \text{PCCG AAGAAA CCG}\end{smallmatrix}$ (1.46 kcal/mol, Figure 2m) and $\begin{smallmatrix}\text{GGC GGA GGCU}\\ \text{PCCG AAGUUU CCG}\end{smallmatrix}$ (−0.22 kcal/mol, Figure 2k); $\begin{smallmatrix}\text{GGU GGA GGCU}\\ \text{PCCG AAGAAA CCG}\end{smallmatrix}$ (0.17 kcal/mol, Figure 2l) and $\begin{smallmatrix}\text{GGU GGA GGCU}\\ \text{PCCG AAGUUU CCG}\end{smallmatrix}$ (−0.33 kcal/mol, Figure 2j)). Several weak downfield peaks are probably due to minor conformations (Figures 2j,k and S1j,k, Supporting Information). This structural preference is predicted by the thermodynamic model: $\begin{smallmatrix}\text{C GGA G}\\ \text{G AAGUUU C}\end{smallmatrix}$ (0.41 kcal/mol predicted) and $\begin{smallmatrix}\text{U GGA G}\\ \text{G AAGUUU C}\end{smallmatrix}$ (0.20 kcal/mol predicted) vs $\begin{smallmatrix}\text{C GGAG}\\ \text{G AAG UUUC}\end{smallmatrix}$ (1.02 kcal/mol predicted) and $\begin{smallmatrix}\text{U GGAG}\\ \text{G AAG UUUC}\end{smallmatrix}$ (1.72 kcal/mol predicted). Thus, including a stabilization effect for consecutive GA pairs can help predict multiple GA motifs, including kink-turns (*17*, *50*).

Bonus for Two Consecutive GA Pairs in 3 × 3, 3 × 4, 4 × 4, and 4 × 5 Internal Loops. Two consecutive GA pairs with the motifs $\begin{smallmatrix}\text{Y GA}\\ \text{R AG}\end{smallmatrix}$, $\begin{smallmatrix}\text{R GA}\\ \text{Y AG}\end{smallmatrix}$, $\begin{smallmatrix}\text{Y GG}\\ \text{R AA}\end{smallmatrix}$, or $\begin{smallmatrix}\text{R GG}\\ \text{Y AA}\end{smallmatrix}$ (i.e., $\frac{\text{GA}}{\text{AG}}$ or $\frac{\text{GG}}{\text{AA}}$ closed on at least one side with a canonical pair that is 5′ to the G of a GA pair) only stabilize certain types of internal loops, including 3 × 3, 3 × 4, 4 × 4, and 4 × 5 loops. For 3 × 3 loops such as $\begin{smallmatrix}\text{GAGU GAA UGAC}\\ \text{CUCA AGA ACUG}\end{smallmatrix}$ (2.30 kcal/mol), $\begin{smallmatrix}\text{GAGC GAG CGAC}\\ \text{CUCG AGA GCUG}\end{smallmatrix}$ (0.54 kcal/mol), $\begin{smallmatrix}\text{GAGC GAA CGAC}\\ \text{CUCG AGA GCUG}\end{smallmatrix}$ (0.16 kcal/mol), $\begin{smallmatrix}\text{GAGC AGA CGAC}\\ \text{CUCG AAG GCUG}\end{smallmatrix}$ (−0.24 kcal/mol), $\begin{smallmatrix}\text{GAGC CGA CGAC}\\ \text{CUCG AAG GCUG}\end{smallmatrix}$ (−0.36 kcal/mol), $\begin{smallmatrix}\text{CGC AGA GGC}\\ \text{GCG AAG CCG}\end{smallmatrix}$ (−0.45 kcal/mol), $\begin{smallmatrix}\text{CGAC CGA GCAG}\\ \text{GCUG AAG CGUC}\end{smallmatrix}$ (−0.60 kcal/mol), and $\begin{smallmatrix}\text{GAGC GGA CGAC}\\ \text{CUCG AAA GCUG}\end{smallmatrix}$ (−0.65 kcal/mol), the $\Delta G^\circ_{\mathrm{2GA\,bonus}}$ and $\Delta G^\circ_{\mathrm{GA\,bonus}}$ are applied without adding a $\Delta G^\circ_{\mathrm{middle\,GA\,bonus}}$.

No extra stabilization was observed for two consecutive GA pairs within 2 × 4, 2 × 5, 2 × 6, 3 × 5, 3 × 6, and 4 × 6 loops. Evidently, this term is restricted to internal loops with asymmetry less than two nucleotides. This is in agreement with previous experimental and theoretical modeling studies showing that two consecutive sheared GA pairs in a 2 × 5 internal loop bound to a protein are not present in the free RNA without protein (*18*, *21*). The different contexts exhibiting stabilization for three and two consecutive GA pairs might explain the intolerance of mutation in the $\frac{\text{GGA}}{\text{AAG}}$ motif in the kink-turn (kt-7) $\begin{smallmatrix}\text{G GGA G}\\ \text{C AAGAAG C}\end{smallmatrix}$ (*17*, *51*).

Structurally, there is a possibility that 2 × 4 internal loops could form consecutive GA pairs or two independent first noncanonical GA pairs. At this stage, two bonus parameters of first noncanonical GA pairs, $\Delta G^\circ_{\mathrm{GA\,bonus}}$, are applied for $\begin{smallmatrix}\text{GGC GAAA GGCU}\\ \text{PCCG A G CCG}\end{smallmatrix}$ because the asymmetry is two. This might not reflect the structure, however.

Thermodynamically stable consecutive GA pairs are an important secondary structure motif, providing preorganized functional groups for tertiary interactions such as the A-minor motif (*15*−*17*, *90*) and binding of ligands such as protein (*17*−*23*) and Mg^{2+} (aq) (*88*, *92*, *93*). Similar stabilizing effects are likely in large hairpin loops as well as multibranch loops. Binding of Mg^{2+} (aq) is not expected to significantly stabilize this motif, however, because the thermodynamics of $\begin{smallmatrix}\text{GGU GGA GGCU}\\ \text{PCCG AAG CCG}\end{smallmatrix}$ was essentially identical in 1 M NaCl and in 150 mM KCl with 10 mM $MgCl_2$ (*33*). Binding of protein might stabilize two consecutive GA pairs within size asymmetric internal loops, however (*18*, *21*, *23*).

Sequence-Dependent Stabilizing Effect of Consecutive GA Pairs. Closing canonical base pairs are important for stabilizing sheared GA pairs. The $\Delta G^\circ_{\mathrm{3GA\,bonus}}$ is applied only for loops closed by at least one YR canonical pair (i.e., UG, UA, or CG with the U or C on the 5′ side of the G of a first GA pair), which is favorable for the formation of sheared GA pairs as observed in 2 × 2 loops (*36*). For loops with three potentially consecutive GA pairs but not closed with at least one YR canonical pair, e.g., $\begin{smallmatrix}\text{GAGC AGGA CGAC}\\ \text{CUCG AAG GCUG}\end{smallmatrix}$ (1.26 kcal/mol) and $\begin{smallmatrix}\text{GAGC AGGA CGAC}\\ \text{CUCG AAAG GCUG}\end{smallmatrix}$ (0.03 kcal/mol), the $\Delta G^\circ_{\mathrm{2GA\,bonus}}$ is applied. Presumably this is due to the destabilizing effect in changing a CG to a GC closing pair adjacent to a sheared GA pair. Destabilization of 1.27 kcal/mol was observed previously for changing the tetraloop hairpin C<u>GAAA</u>G to G<u>GAAA</u>C (*94*) even though both sequences have sheared GA pairs (*14*, *44*, *93*, *95*, *96*). Formation of the kink-turn (kt-7) $\begin{smallmatrix}\text{G GGA G}\\ \text{C AAGAAG C}\end{smallmatrix}$ instead of $\begin{smallmatrix}\text{G GGA G}\\ \text{C AAGAAG C}\end{smallmatrix}$ might be due to binding of protein and/or tertiary interactions (*17*), even though a sheared GA pair is thermodynamically more favorable with a CG closing base pair. Alternatively, the $\frac{\text{GGA}}{\text{AAG}}$ motif is thermodynamically more favorable closed on the 5′ GG 3′ side (i.e., $\begin{smallmatrix}\text{Y GGA}\\ \text{R AAG}\end{smallmatrix}$ or $\begin{smallmatrix}\text{R GGA}\\ \text{Y AAG}\end{smallmatrix}$) than on the 5′ GA 3′ side (i.e., $\begin{smallmatrix}\text{GGA R}\\ \text{AAG Y}\end{smallmatrix}$ or $\begin{smallmatrix}\text{GGA Y}\\ \text{AAG R}\end{smallmatrix}$).

Sequence-Dependent Stabilizing Effect of AG Pairs. No bonus parameter for an AG first noncanonical pair is applied for loops with a single $\frac{\text{R A}}{\text{Y G}}$ motif (RY is canonical pair AU, GU, or GC) such as $\frac{\text{GAGC GAG CGAC}}{\text{CUCG AUA GCUG}}$ (1.28 kcal/mol) and $\frac{\text{GAGC GAG CGAC}}{\text{CUCG AAA GCUG}}$ (1.08 kcal/mol). In addition, no bonus parameter is applied for loops even with potentially consecutive AG pairs: $\frac{\text{GAGC AGAG CGAC}}{\text{CUCG AGA GCUG}}$ (2.74 kcal/mol), $\frac{\text{GAGC AGAG CGAC}}{\text{CUCG AAGA GCUG}}$ (2.07 kcal/mol), and $\frac{\text{GAGC AAAG CGAC}}{\text{CUCG AAGA GCUG}}$ (2.88 kcal/mol). This is consistent with the fact that no stabilization effect is found for the $\frac{\text{RA}}{\text{YG}}$ motif in 2 × 3 loops (Tables 3 and S2, Supporting Information) (*9*, *28*, *29*). The $\Delta G^{\circ}_{\text{3GA bonus}}$ is not applied for a 3 × 3 loop $\frac{\text{GAGC GAG CGAC}}{\text{CUCG AGA GCUG}}$ (0.54 kcal/mol) with one AG first noncanonical pair, because backbone narrowing prohibits formation of a Watson−Crick pair 5′ of the A in a sheared GA pair (*38*, *60*, *92*). Thus, only bonus parameters of $\Delta G^{\circ}_{\text{2GA bonus}}$ and $\Delta G^{\circ}_{\text{GA bonus}}$ are applied for the loop in $\frac{\text{GAGC GAG CGAC}}{\text{CUCG AGA GCUG}}$. Nevertheless, bonus parameters are applied for 3 × 3 loops with two $\frac{\text{GA}}{\text{CG}}$ motifs: $\frac{\text{CGG AAG CGC}}{\text{GCC GUA GCG}}$ (1.21 kcal/mol) and $\frac{\text{CGG AAG CGC}}{\text{GCC GAA GCG}}$ (0.87 kcal/mol), presumably due to geometric compatibility of consecutive face-to-face pairs as observed in 2 × 2 loops (*38*, *92*).

Bonus for $\frac{\text{U G}}{\text{G A}}$ *Motif.* The thermodynamic parameters in folding algorithms typically assume that UG/GU pairs closing internal loops are equivalent to UA/AU pairs (*8*, *9*). As shown in Table 2, loops with $\frac{\text{U G}}{\text{G A}}$ and $\frac{\text{C G}}{\text{G A}}$ motifs have similar stabilities. A stabilization effect of the $\frac{\text{U G}}{\text{G A}}$ motif was also found in 3 × 3 loops (*33*, *34*). The motif of $\frac{\text{U G}}{\text{G A}}$ is thermodynamically relatively stable in 2 × 2 loops when compared with the motif of $\frac{\text{U G}}{\text{A A}}$ but not when compared with $\frac{\text{C G}}{\text{G A}}$ (*29*), suggesting that 2 × 2 loops have less flexibility than 3 × 3 loops. The parameter $\Delta G^{\circ}_{\text{5'UG/3'GA bonus}}$ of −0.95 kcal/mol compensates for the penalty term of 0.7 kcal/mol for UG closure in current folding algorithms. This correlates with more extensive stacking of $\frac{\text{U G}}{\text{G A}}$ than $\frac{\text{C G}}{\text{G A}}$ and a hydrogen bond between the G amino group from a wobble UG pair and GO4′ of the sheared GA pair as shown in an NMR structure with the loop $\frac{\text{U GGA G}}{\text{G AAG C}}$ which contains both a $\frac{\text{U G}}{\text{G A}}$ motif and a $\frac{\text{C G}}{\text{G A}}$ motif (*33*). Formation of UG wobble pairs is consistent with the relatively sharp resonances of the imino protons of G and U from UG pairs (Figures 2 and S1). Note that the U3 imino proton from a UA pair adjacent to a GA pair is relatively broad and shifted upfield (Figure 2b) relative to the usual range of 13−15 ppm for a Watson−Crick UA pair. A similar upfield shift was observed previously in the 2 × 2 loop $\frac{\text{U GA A}}{\text{A AG U}}$ (*75*). Several 3 × 3 loops with $\frac{\text{G G}}{\text{U A}}$ and $\frac{\text{G A}}{\text{U A}}$ motifs are typically less stable than predicted by eq 4, which is consistent with previous thermodynamic and NMR studies in 2 × 2 and 2 × 3 loops (*29*, *30*, *40*) and joint X-ray and NMR studies of a kink-turn loop, $\frac{\text{CG GA G}}{\text{GU AGAGA C}}$ with protein binding (*18*). Thus, depending on the orientation, GU/UG closing base pairs can either destabilize or stabilize internal loops as compared with AU/UA closing pairs.

Comparisons with Other Loops. It is likely that additional elements of stability remain to be discovered. For example, the loop in $\frac{\text{GGC GGAU CGCU}}{\text{PCCG AAGU CCG}}$ (−4.16 kcal/mol), which was not included in the regression analysis, is 2.64 kcal/mol more stable than predicted by eq 4. Further studies are needed for the stable $\frac{\text{AU}}{\text{GU}}$ nearest neighbor. Moreover, NMR studies have revealed structured internal loops with consecutive UU and UC pairs (*32*, *34*, *49*, *97*), although no thermodynamically significant bonus stabilizing effect has been found yet for consecutive UU and UC pairs (*32*, *34*).

When eq 2a is applied to previous data (*98*) on the size asymmetric loop E motif, $\frac{\text{C GA A C}}{\text{G AUGA G}}$, the $\Delta G^{\circ}_{37,\text{loop}}$ is calculated to be 0.37 kcal/mol in 1 M NaCl, which is 1.42 kcal/mol more stable than predicted by eq 4. This size asymmetric loop E motif (G bulge motif) forms in the absence of Mg^{2+} (aq) or Ca^{2+} (aq) as shown by NMR studies (*31*, *45*, *46*).

There are unstable 3 × 3 loops with a single first GA pair that has a U 3′ to the G of the GA pair, e.g., $\frac{\text{GAGC AAA CGAC}}{\text{CUCG AUG GCUG}}$ (1.57 kcal/mol) and $\frac{\text{CGC AAA GGC}}{\text{GCG AUG CCG}}$ (1.98 kcal/mol), which give rise to the $\Delta G^{\circ}_{\text{5'GU/3'AN penalty}}$ (Table S2, Supporting Information) (*34*). These loops are found in crystal structures. In the conserved 3 × 3 loop $\frac{\text{C AAA C}}{\text{G AUG G}}$ in helix 24 of 16S rRNA, three noncanonical pairs form (trans Hoogsteen/sugar edge AA, AG first noncanonical pairs, and trans Watson−Crick/Hoogsteen UA middle pair) but with very little base overlap (*90*). In the other case, the noncanonical G in $\frac{\text{C AAA G}}{\text{G AUG C}}$ (in helix 38 of 23S rRNA) forms a pair with U to make a base triple (*16*). The results show that the thermodynamic penalty of $\Delta G^{\circ}_{\text{5'GU/3'AN penalty}}$ works well even when the crystal structures are different.

The size symmetric 7 × 7 loop E motif, $\frac{\text{C GAUGGUA G}}{\text{G AUGAGAG C}}$, with calculated loop free energy of 0.57 kcal/mol at 37 °C in 1 M NaCl on the basis of published data (*98*) is predicted well (0.99 kcal/mol) by eq 4 with the loop initiation penalty of 2.81 kcal/mol extrapolated with the equation $\Delta G^{\circ}_{\text{loop initiation}}(n) = \Delta G^{\circ}_{\text{loop initiation}}\ (9) + 1.75 \times RT \times \ln(14/9)$ from a loop initiation penalty of 2.33 kcal/mol for internal loops with nine nucleotides. The value was extrapolated from $\Delta G^{\circ}_{\text{loop initiation}}$ (9) because that was the largest loop size represented by at least 10 sequences.

Melting Transition Cooperativity and Enthalpy Changes. As shown in Table 2, asymmetric internal loops typically have less favorable enthalpy changes than symmetric internal loops. This is consistent with previous UV melting studies of bulges and asymmetric internal loops (*71*, *99*, *100*) and indicates less cooperativity in duplex melting when the loop is asymmetric.

CONCLUSION

The stabilizing effect of consecutive sheared GA pairs within internal loops larger than five nucleotides is sequence and size dependent. Consecutive sheared GA pairs can form in motifs other than internal loops. Including this thermodynamic effect quantitatively or semiquantitatively should help model RNA secondary and tertiary structures.

ACKNOWLEDGMENT

We thank Dr. Susan J. Schroeder for suggesting some experiments, Zhi Lu and Prof. David H. Mathews for providing a linear regression file of previous thermodynamic data, Dr. Sandip K. Sur, Prof. Scott D. Kennedy, and Prof. Thomas R. Krugh for help with ^{31}P spectra, and Prof. Martin J. Serra for discussions on Mg^{2+} (aq) dependence.

SUPPORTING INFORMATION AVAILABLE

Tables of single strand melting results, comparison of measured and predicted internal loop free energies for the entire database, multiple linear regression of loops larger than 2 × 3, SNOESY 2D spectra, and sample calculations for predicting the free energies for formation of internal loops. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Crothers, D. M., Cole, P. E., Hilbers, C. W., and Shulman, R. G. (1974) Molecular mechanism of thermal unfolding of *Escherichia coli* formylmethionine transfer RNA, *J. Mol. Biol. 87*, 63−88.
2. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) RNA structure prediction, *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
3. Banerjee, A. R., Jaeger, J. A., and Turner, D. H. (1993) Thermal unfolding of a group I ribozyme: the low temperature transition is primarily disruption of tertiary structure, *Biochemistry 32*, 153−163.
4. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) The interactions that shape RNA structure, in *The RNA World*, 2nd ed. (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) pp 233−264, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. Turner, D. H. (2000) Conformational Changes, in *Nucleic Acids: Structures, Properties, and Functions* (Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr., Eds.) pp 259−334, University Science Books, Sausalito, California.
6. Xia, T., Mathews, D. H., and Turner, D. H. (1999) Thermodynamics of RNA secondary structure formation, in *Prebiotic Chemistry, Molecular Fossils, Nucleotides, and RNA* (Soll, D., Moore, P. B., and Nishimura, S., Eds.) pp 21−47, Elsevier Science Ltd., Oxford.
7. 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.
8. 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.
9. 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.
10. Michel, F., and Westhof, E. (1990) Modeling of the 3-dimensional architecture of group I catalytic introns based on comparative sequence analysis, *J. Mol. Biol. 216*, 585−610.
11. Costa, M., Christian, E. L., and Michel, F. (1998) Differential chemical probing of a group II self-splicing intron identifies bases involved in tertiary interactions and supports an alternative secondary structure model of domain V, *RNA 4*, 1055−1068.
12. Masquida, B., and Westhof, E. (2006) A modular and hierarchical approach for all-atom RNA modeling, in *The RNA World*, 3rd ed. (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) pp 659−681, Cold Spring Harbor Laboratory Press, Woodbury, NY.
13. Klostermeier, D., and Millar, D. P. (2002) Energetics of hydrogen bond networks in RNA: Hydrogen bonds surrounding G+1 and U42 are the major determinants for the tertiary structure stability of the hairpin ribozyme, *Biochemistry 41*, 14095−14102.
14. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K. H., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) Crystal structure of a group I ribozyme domain: Principles of RNA packing, *Science 273*, 1678−1685.
15. Doherty, E. A., Batey, R. T., Masquida, B., and Doudna, J. A. (2001) A universal mode of helix packing in RNA, *Nat. Struct. Biol. 8*, 339−343.
16. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) RNA tertiary interactions in the large ribosomal subunit: The A-minor motif, *Proc. Natl. Acad. Sci. U.S.A. 98*, 4899−4903.
17. Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001) The kink-turn: A new RNA secondary structure motif, *EMBO J. 20*, 4214−4221.
18. Chao, J. A., and Williamson, J. R. (2004) Joint X-ray and NMR refinement of the yeast L30e-mRNA complex, *Structure 12*, 1165−1176.
19. Hamma, T., and Ferre-D'Amare, A. R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 angstrom resolution, *Structure 12*, 893−903.
20. Moore, T., Zhang, Y. M., Fenley, M. O., and Li, H. (2004) Molecular basis of box C/D RNA-protein interactions: Cocrystal structure of archaeal L7Ae and a box C/D RNA, *Structure 12*, 807−818.
21. Cojocaru, V., Nottrott, S., Klement, R., and Jovin, T. M. (2005) The snRNP 15.5K protein folds its cognate K-turn RNA: A combined theoretical and biochemical study, *RNA 11*, 197−209.
22. Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R., and Ficner, R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment, *Mol. Cell. 6*, 1331−1342.
23. Szewczak, L. B. W., DeGregorio, S. J., Strobel, S. A., and Steitz, J. A. (2002) Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP, *Chem. Biol. 9*, 1095−1107.
24. Yuan, Y. Q., Kerwood, D. J., Paoletti, A. C., Shubsda, M. F., and Borer, P. N. (2003) Stem of SL1 RNA in HIV-1: Structure and nucleocapsid protein binding for a 1 × 3 internal loop, *Biochemistry 42*, 5259−5269.
25. Lynch, S. R., Gonzalez, R. L., and Puglisi, J. D. (2003) Comparison of X-ray crystal structure of the 30S subunit-antibiotic complex with NMR structure of decoding site oligonucleotide-paromomycin complex, *Structure 11*, 43−53.
26. Greatorex, J., Gallego, J., Varani, G., and Lever, A. (2002) Structure and stability of wild-type and mutant RNA internal loops from the SL-1 domain of the HIV-1 packaging signal, *J. Mol. Biol. 322*, 543−557.
27. Schroeder, S. J., Burkard, M. E., and Turner, D. H. (1999) The energetics of small internal loops in RNA, *Biopolymers 52*, 157−167.
28. Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, *Biochemistry 39*, 9257−9274.
29. Schroeder, S. J., and Turner, D. H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA, *Biochemistry 40*, 11509−11517.
30. Schroeder, S. J., Fountain, M. A., Kennedy, S. D., Lukavsky, P. J., Puglisi, J. D., Krugh, T. R., and Turner, D. H. (2003) Thermodynamic stability and structural features of the J4/5 loop in a *Pneumocystis carinii* group I intron, *Biochemistry 42*, 14184−14196.
31. Wimberly, B., Varani, G., and Tinoco, I. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA, *Biochemistry 32*, 1078−1087.
32. Theimer, C. A., Finger, L. D., Trantirek, L., and Feigon, J. (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA, *Proc. Natl. Acad. Sci. U.S.A. 100*, 449−454.
33. 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.
34. 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.
35. Sashital, D. G., Allmann, A. M., Van Doren, S. R., and Butcher, S. E. (2003) Structural basis for a lethal mutation in U6 RNA, *Biochemistry 42*, 1470−1477.
36. SantaLucia, J., Jr., and Turner, D. H. (1993) Structure of $(\mathrm{rGGC\underline{GA}GCC})_2$ in solution from NMR and restrained molecular dynamics, *Biochemistry 32*, 12612−12623.
37. Walter, A. E., Wu, M., and Turner, D. H. (1994) The stability and structure of tandem GA mismatches in RNA depend on closing base pairs, *Biochemistry 33*, 11349−11354.
38. Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) Solution structure of $(\mathrm{rGGC\underline{AG}GCC})_2$ by two-dimensional NMR and the iterative relaxation matrix approach, *Biochemistry 36*, 4449−4460.
39. Xia, T., McDowell, J. A., and Turner, D. H. (1997) Thermodynamics of nonsymmetric tandem mismatches adjacent to G·C base pairs in RNA, *Biochemistry 36*, 12486−12497.
40. 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.
41. Jiang, F., Kumar, R. A., Jones, R. A., and Patel, D. J. (1996) Structural basis of RNA folding and recognition in an AMP−RNA aptamer complex, *Nature 382*, 183−186.
42. Dieckmann, T., Suzuki, E., Nakamura, G. K., and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold, *RNA 2*, 628−640.

43. Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P., and Pardi, A. (1997) Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA, *Nat. Struct. Biol. 4*, 644−649.
44. Lawrence, D. C., Stover, C. C., Noznitsky, J., Wu, Z. R., and Summers, M. F. (2003) Structure of the intact stem and bulge of HIV-1 Ψ-RNA stem-loop SL1, *J. Mol. Biol. 326*, 529−542.
45. Bouvet, P., Allain, F. H. T., Finger, L. D., Dieckmann, T., and Feigon, J. (2001) Recognition of pre-formed and flexible elements of an RNA stem-loop by nucleolin, *J. Mol. Biol. 309*, 763−775.
46. Szewczak, A. A., Moore, P. B., Chan, Y. L., and Wool, I. G. (1993) The conformation of the sarcin/ricin loop from 28S ribosomal RNA, *Proc. Natl. Acad. Sci. U.S.A. 90*, 9581−9585.
47. Cai, Z., and Tinoco, I. (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite, *Biochemistry 35*, 6026−6036.
48. 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.
49. 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.
50. Winkler, W. C., Grundy, F. J., Murphy, B. A., and Henkin, T. M. (2001) The GA motif: An RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs, *RNA 7*, 1165−1172.
51. Goody, T. A., Melcher, S. E., Norman, D. G., and Lilley, D. M. J. (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent, *RNA 10*, 254−264.
52. Kaine, B. P. (1990) Structure of the archaebacterial 7S RNA molecule, *Mol. Gen. Genet. 221*, 315−321.
53. 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.
54. 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.
55. Wu, M., and Tinoco, I. (1998) RNA folding causes secondary structure rearrangement, *Proc. Natl. Acad. Sci. U.S.A. 95*, 11555−11560.
56. Zheng, M. X., Wu, M., and Tinoco, I. (2001) Formation of a GNRA tetraloop in P5abc can disrupt an interdomain interaction in the Tetrahymena group I ribozyme, *Proc. Natl. Acad. Sci. U.S.A. 98*, 3695−3700.
57. Han, J., and Burke, J. M. (2005) Model for general acid-base catalysis by the hammerhead ribozyme: pH-activity relationships of G8 and G12 variants at the putative active site, *Biochemistry 44*, 7864−7870.
58. Elgavish, T., Cannone, J. J., Lee, J. C., Harvey, S. C., and Gutell, R. R. (2001) AA.AG@Helix.Ends: A:A and A:G base-pairs at the ends of 16 and 23 S rRNA helices, *J. Mol. Biol. 310*, 735−753.
59. Gutell, R. R., Cannone, J. J., Shang, Z., Du, Y., and Serra, M. J. (2000) A story: Unpaired adenosine bases in ribosomal RNAs, *J. Mol. Biol. 304*, 335−354.
60. Gautheret, D. F., Konings, D., and Gutell, R. R. (1994) A major family of motifs involving G·A mismatches in ribosomal RNA, *J. Mol. Biol. 242*, 1−8.
61. Silverman, K. S., Zheng, M. X., Wu, M., Tinoco, I., and Cech, T. R. (1999) Quantifying the energetic interplay of RNA tertiary and secondary structure interactions, *RNA 5*, 1665−1674.
62. Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2005) RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts, *J. Am. Chem. Soc. 127*, 4659−4667.
63. 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.
64. 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.
65. 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.
66. 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.
67. McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAG$\underline{\text{GU}}$CUC)$_2$ by two-dimensional NMR and simulated annealing, *Biochemistry 35*, 14077−14089.
68. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) Thermodynamic study of internal loops in oligoribonucleotides: Symmetrical loops are more stable than asymmetric loops, *Biochemistry 30*, 6428−6436.
69. 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.
70. 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.
71. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry 29*, 278−285.
72. Lukavsky, P. J., and Puglisi, J. D. (2001) RNAPack: An integrated NMR approach to RNA structure determination, *Methods 25*, 316−332.
73. Gralla, J., and Crothers, D. M. (1973) Free energy of imperfect nucleic acid helices.3. small internal loops resulting from mismatches, *J. Mol. Biol. 78*, 301−319.
74. Burkard, M. E., Xia, T., and Turner, D. H. (2001) Thermodynamics of RNA internal loops with a guanosine-guanosine pair adjacent to another noncanonical pair, *Biochemistry 40*, 2478−2483.
75. 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.
76. Chou, S. H., Zhu, L., and Reid, B. R. (1997) Sheared purine·purine pairing in biology, *J. Mol. Biol. 267*, 1055−1067.
77. Gultyaev, A. P., Vanbatenburg, F. H. D., and Pleij, C. W. A. (1995) The computer simulation of RNA folding pathways using a genetic algorithm, *J. Mol. Biol. 250*, 37−51.
78. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., and Gralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids, *Nature New Biol. 246*, 40−41.
79. Rivas, E., and Eddy, S. R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots, *J. Mol. Biol. 285*, 2053−2068.
80. Wuchty, S., Fontana, W., Hofacker, I. L., and Schuster, P. (1999) Complete suboptimal folding of RNA and the stability of secondary structures, *Biopolymers 49*, 145−165.
81. Ding, Y., and Lawrence, C. E. (2003) A statistical sampling algorithm for RNA secondary structure prediction, *Nucleic Acids Res. 31*, 7280−7301.
82. Dirks, R. M., and Pierce, N. A. (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots, *J. Comput. Chem. 24*, 1664−1677.
83. Razga, F., Spackova, N., Reblova, K., Koca, J., Leontis, N. B., and Sponer, J. (2004) Ribosomal RNA kink-turn motif − A flexible molecular hinge, *J. Biomol. Struct. Dyn. 22*, 183−193.
84. Pinard, R., Lambert, D., Heckman, J. E., Esteban, J. A., Gundlach, C. W., Hampel, K. J., Glick, G. D., Walter, N. G., Major, F., and Burke, J. M. (2001) The hairpin ribozyme substrate binding-domain: A highly constrained D-shaped conformation, *J. Mol. Biol. 307*, 51−65.
85. Brown, T. S., Chadalavada, D. M., and Bevilacqua, P. C. (2004) Design of a highly reactive HDV ribozyme sequence uncovers facilitation of RNA folding by alternative pairings and physiological ionic strength, *J. Mol. Biol. 341*, 695−712.
86. Chadalavada, D. M., Knudsen, S. M., Nakano, S., and Bevilacqua, P. C. (2000) A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme, *J. Mol. Biol. 301*, 349−367.

87. Kraut, D. A., Carroll, K. S., and Herschlag, D. (2003) Challenges in enzyme mechanism and energetics, *Annu. Rev. Biochem. 72*, 517−571.
88. Hoffmann, B., Mitchell, G. T., Gendron, P., Major, F., Andersen, A. A., Collins, R. A., and Legault, P. (2003) NMR structure of the active conformation of the Varkud satellite ribozyme cleavage site, *Proc. Natl. Acad. Sci. U.S.A. 100*, 7003−7008.
89. Jovine, L., Hainzl, T., Oubridge, C., Scott, W. G., Li, J., Sixma, T. K., Wonacott, A., Skarzynski, T., and Nagai, K. (2000) Crystal structure of the Ffh and EF-G binding sites in the conserved domain IV of *Escherichia coli* 4.5S RNA, *Struct. Folding Des. 8*, 527−540.
90. 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.
91. 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 angstrom resolution, *Science 289*, 905−920.
92. Rudisser, S., and Tinoco, I. (2000) Solution structure of cobalt-(III) hexammine complexed to the GAAA tetraloop, and metal ion binding to G·A mismatches, *J. Mol. Biol. 295*, 1211−1223.
93. Kieft, J. S., and Tinoco, I. (1997) Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine, *Structure 5*, 713−721.
94. Moody, E. M., Feerrar, J. C., and Bevilacqua, P. C. (2004) Evidence that folding of an RNA tetraloop hairpin is less cooperative than its DNA counterpart, *Biochemistry 43*, 7992−7998.
95. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M., and Pardi, A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops, *J. Mol. Biol. 264*, 968−980.
96. Huang, H. C., Nagaswamy, U., and Fox, G. E. (2005) The application of cluster analysis in the intercomparison of loop structures in RNA, *RNA 11*, 412−423.
97. Furtig, B., Richter, C., Wohnert, J., and Schwalbe, H. (2003) NMR spectroscopy of RNA, *ChemBioChem 4*, 936−962.
98. Serra, M. J., Baird, J. D., Dale, T., Fey, B. L., Retatagos, K., and Westhof, E. (2002) Effects of magnesium ions on the stabilization of RNA oligomers of defined structures, *RNA 8*, 307−323.
99. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) Improved predictions of secondary structures for RNA, *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.
100. Weeks, K. M., and Crothers, D. M. (1993) Major groove accessibility of RNA, *Science 261*, 1574−1577.

BI052060T