Factors Affecting Thermodynamic Stabilities of RNA 3 × 3 Internal Loops†

Gang Chen,‡ Brent M. Znosko,‡,§ Xiaoqi Jiao,‡ and Douglas H. Turner*,‡,||

Department of Chemistry, University of Rochester, Rochester, New York 14627, and Center for Human Genetics and Molecular Pediatric Disease and Department of Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

Received April 25, 2004; Revised Manuscript Received July 19, 2004

ABSTRACT: Internal loops in RNA are important for folding and function. The 3 × 3 nucleotide internal loops are the smallest size symmetric loops with a potential noncanonical base pair (middle pair) flanked on both sides by a noncanonical base pair (loop-terminal pair). Thermodynamic and structural insights acquired for 3 × 3 loops should improve approximations for stabilities of 3 × 3 and larger internal loops. Most natural 3 × 3 internal loops are purine rich, which is also true of other internal loops. A series of oligoribonucleotides containing different 3 × 3 internal loops were studied by UV melting and imino proton NMR. Both loop-terminal and middle pairs contribute to the thermodynamic stabilities of 3 × 3 loops. Extra stabilization of −1.2 kcal/mol was found for a GA middle pair when flanked by at least one non-pyrimidine-pyrimidine loop-terminal pair. A penalty of ∼1 kcal/mol was found for loops with a single loop-terminal GA pair that has a U 3′ to the G of the GA pair. A revised model for predicting stabilities of 3 × 3 loops is derived by multiple linear regression.

RNA was perhaps the first type of molecule involved in the evolution of life and is still involved in many cellular functions (*1*−*3*). Sequencing projects are generating DNA sequence information at a rate of greater than 1 billion nucleotides per year. This is providing a foundation for determining the secondary and tertiary structures of RNAs transcribed from DNAs. This structural information should provide a deeper understanding of structure−function relationships and could also lead to rational design of therapeutics targeting RNA (*4*−*9*).

X-ray diffraction (*10*) and NMR (*10*−*15*) methods are providing an increasing number of RNA structures, but it is not likely that these methods will keep pace with the rate at which interesting sequences are being discovered. Therefore, it is necessary to develop reliable, rapid methods for predicting secondary and three-dimensional structures of RNAs. Such methods will also be useful for designing RNA constructs suitable for X-ray and NMR studies.

On the basis of the principle that structures are more conserved than sequences for function, many RNA secondary structures have been accurately deduced by comparative sequence analysis (*16*−*19*). Often, however, there are not enough sequences available to apply this method. It is also probable that structure will not be largely conserved in coding sequences.

Interactions determining RNA secondary structure, primarily hydrogen bonding between base pairs and stacking between adjacent bases (*20*−*22*), apparently dominate those of tertiary structure. Thus, studies of the properties of short oligonucleotides can provide insight into folding of large RNAs (*23*). Therefore, free energy minimization combined with experimental restraints such as chemical modification (*24*) and sequence comparison (*25*, *26*) could predict RNA secondary structure in a manner largely independent of tertiary structure. Thermodynamics could also facilitate design of therapeutics targeting RNA (*5*, *6*), self-assembling nanostructures based on nucleic acid folding (*27*, *28*), and biosensors based on nucleic acid hybridization (*29*−*33*). Recently, a correlation was found between the thermodynamics of siRNA[1] and miRNA and their ability to induce RNA interference (*34*−*36*).

Internal loops are important motifs for tertiary interactions, protein and small molecule binding, and potential therapeutic targeting. UV melting and imino proton NMR have been combined to study the thermodynamics and structures of small RNA motifs (*21*, *37*, *38*). UV melting is the easiest and most economical way to measure the thermodynamics of a small RNA if it melts in a two-state manner. Imino proton NMR provides basic structural information.

Studies on the thermodynamics of small internal loops show that size symmetric internal loops have more sequence dependence than size asymmetric loops (*38*). For example, for internal loops closed by GC or CG pairs, stabilities of 2 × 2 nucleotide loops range from 2.2 to −2.9 kcal/mol whereas those of 1 × 3 nucleotide loops range from 3.3 to 1.6 kcal/mol at 37 °C (*38*). Presumably, this is because asymmetric loops are less restricted in available stacking patterns. This flexibility may also be reflected in the observation that unbound asymmetric loops are typically

† This work was supported by NIH Grant GM22939 (D.H.T.).

* To whom correspondence should be addressed. Phone: (585) 275-3207. Fax: (585) 506-0205. E-mail: turner@chem.rochester.edu.

‡ University of Rochester.

§ Current address: Department of Chemistry, St. Louis University, St. Louis, MO 63103.

|| University of Rochester School of Medicine and Dentistry.

[1] Abbreviations: C_T, total concentration of all strands of oligonucleotides in solution; LSU, large subunit; miRNA, microRNA; N, any nucleotide, including A, C, G, or U; siRNA, small interfering RNA; SRP, signal recognition particle; SSU, small subunit; T_M, melting temperature in kelvin; T_m, melting temperature in degrees Celsius.

10.1021/bi049168d CCC: $27.50
Published on Web 09/18/2004

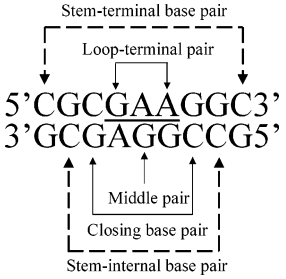

FIGURE 1: Schematic representation of the nomenclature for base pairs in duplexes with 3 × 3 internal loops. Throughout the paper, each top strand is written from 5′ to 3′ from left to right.

relatively unstructured in solution (*15*, *39*−*41*). Current free energy parameters for 3 × 3 internal loops (*23*, *24*) are largely based on the knowledge of 2 × 2 and 2 × 3 internal loops. The 3 × 3 internal loops, however, are the smallest size symmetric loops with a potential noncanonical base pair (middle pair) flanked by a noncanonical base pair (loop-terminal pair) on both sides (Figure 1). The flexibility of internal loops will increase as the loop size increases. Presumably, 3 × 3 internal loops differ from 2 × 2 loops more than from 4 × 4 and larger size symmetric loops. Thus, insights acquired from 3 × 3 loops should improve approximations for stabilities of 3 × 3 and larger internal loops.

The thermodynamic contribution of each noncanonical pair in sequence nonsymmetric 2 × 2 internal loops depends on the identity of the adjacent noncanonical pair (*42*, *43*). Both steric and electrostatic compatibilities probably determine internal loop stability (*38*, *42*−*45*). It is possible to generalize steric effects with "isostericity matrices" as determined from the geometry of all possible canonical and noncanonical base pairs by visual examination of high-resolution X-ray crystal structures (*46*). For example, Leontis and Westhof successfully predicted that the loop E motifs in 5S rRNA that are thermodynamically stable in Mg^{2+} are structurally well conserved in different kingdoms even though the sequences differ considerably (*47*, *48*). High-level *ab initio* calculations and molecular modeling of base stacking and hydrogen bonding in nucleic acids are emerging (*49*, *50*). Experimental studies of the sequence dependence of the thermodynamics and structures of internal loops provide benchmarks for testing these theoretical approaches.

Thermodynamic measurements are required in the determination of the factors that are important for determining stabilities of 3 × 3 and larger internal loops. The thermodynamic model developed here will help improve RNA secondary structure prediction.

MATERIALS AND METHODS

Oligoribonucleotide Synthesis and Purification. Oligoribonucleotides were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer using the phosphoramidite method (*51*, *52*). CPG support and phosphoramidites were acquired from Glen Research or ChemGenes. Base protecting groups and CPG support were removed by incubation in a 3:1 (v/v) ammonia/ethanol mixture at 55 °C overnight (*53*). A disposable filter column was used to separate the oligonucleotides from the solid support. Removal of the silyl protecting groups on the 2′-hydroxyls was achieved with a 1 M triethylamine hydrogen fluoride/pyridine mixture (50 equiv) at 55 °C for 48 h. Alternative deprotection of 2′-hydroxyl groups was carried out via incubation in a 9:1 (v/v) TEA·3HF (98%) (Aldrich)/DMF solution at 55 °C for 2 h, followed by 1-butanol precipitation (*54*). The sample was extracted with diethyl ether to remove most of the organic impurities and then lyophilized before being redissolved in 5 mM ammonium bicarbonate or ammonium acetate at pH 7.0. The solution was loaded onto a Waters Sep-Pak C18 chromatography column to remove excess salts. The oligonucleotide was purified on a large preparative Baker Si500F TLC Silica Gel Plate (20 cm × 20 cm, 500 μm thick) with a 55:35:10 (v/v/v) 1-propanol/ammonia/water running solution. The product was identified by UV shadowing and scraped from the plate. RNA was extracted from the silica with distilled water. The Sep-Pak procedure was repeated to desalt the sample. All oligonucleotides were identified by ESI-MS with a Hewlett-Packard 1100 LC/MS Chemstation. Purities were checked by reverse phase HPLC or an analytical Baker Si500F TLC Silica Gel Plate (250 μm thick), and all were greater than 90% 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 dinucleoside monophosphates (*55*). Small mixing errors for non-self-complementary duplexes do not appreciably affect thermodynamic measurements (*56*). Oligonucleotides were lyophilized and redissolved in 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM disodium EDTA at pH 7.0. Curves of absorbance at 280 nm versus temperature were acquired using a heating rate of 1 °C/min with a Beckman Coulter DU640C spectrophotometer with a high-performance temperature controller cooled with flowing water or with a Gilford 250 spectrophotometer with a Gilford 2527 thermoprogrammer.

Melting curves were fit to a two-state model with MeltWin, assuming linear sloping baselines and temperature-independent $\Delta H°$ and $\Delta S°$ (*44*, *57*, *58*). Additionally, the T_M values at different concentrations were used to calculate thermodynamic parameters according to (*59*)

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

where $a = 1$ for self-complementary duplexes and $a = 4$ for non-self-complementary duplexes. T_M is the temperature in kelvin at which half the strands are in duplex. Multiple linear regression and statistical analysis of free energies for loops at 37 °C were carried out with Microsoft Excel.

For self-complementary duplexes, hairpin secondary structures may compete with internal loop secondary structure. For non-self-complementary duplexes, the self-structure (hairpin and/or duplex) of individual single strands may compete with designed non-self-complementary duplexes. The melting data for single-stranded RNA are listed in Table S1 of the Supporting Information. Our rough standard of sequence design is the fact that the T_m values of the designed duplex 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 a heteroduplex (*60*).

Imino Proton NMR Spectroscopy. All spectra were acquired on a Varian Inova 500 MHz spectrometer. NMR spectra of the exchangeable protons of 0.4−1 mM total single strands were recorded at 0−45 °C in 80 mM NaCl, 3 mM NaH_2PO_4, 7 mM Na_2HPO_4, and 0.5 mM Na_2EDTA (pH 7) in a 90:10 H_2O/D_2O mixture. One-dimensional imino proton spectra were acquired with an S or binomial pulse sequence. Proton spectra were referenced to H_2O or HDO at a known temperature-dependent chemical shift relative to 3,3,3-trimethylsilylpropionate (TSP).

RESULTS

Phylogenetic Analysis. Most natural internal loops are purine rich (*16*−*18*, *46*), perhaps due to the combination of large stacking areas and many hydrogen bonding donors and acceptors. A search of known RNA secondary structures of 79 group I introns (*61*), 16 RNase P RNAs (*16*), 91 SRP RNAs (*18*), 109 LSU rRNAs (*62*), and 101 SSU rRNAs (*63*) reveals 1079 3 × 3 internal loops (Table S2). The composition percentages of A, G, U, and C in these 3 × 3 loops are ∼46, ∼24, ∼17, and ∼11%, respectively, not including closing base pairs. GA or AG loop-terminal pairs are the most prevalent (∼40%). Almost 70% of closing base pairs are CG or GC. Thus, most of our model loops are purine rich and closed by CG or GC.

Thermodynamic Parameters. Thermodynamic parameters measured for formation of duplexes with internal loops are listed in Table 1. In Table 1 and throughout the paper, each top strand is written from 5′ to 3′ from left to right. Most of the $\Delta H°$ values from T_M^{-1} versus $\ln(C_T/a)$ plots and from 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 three exceptions are $\frac{CGCAAAGCG}{GCG\overline{AAA}CGC}$, $\frac{CGCGCAGGC}{GCG\overline{AAG}CCG}$, and $\frac{CGCAAAGGC}{GCG\overline{AGG}CCG}$ (Table 1). 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). Because of a high correlation between $\Delta H°$ and $\Delta S°$, $\Delta G°_{37}$ is more accurate than either $\Delta H°$ or $\Delta S°$ (*64*).

Thermodynamic parameters for the formation of the internal loops are listed in Table 2 as calculated according to the following equation (*65*):

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

For example,

$$\Delta G°_{37}\left(\begin{smallmatrix}5'\text{CGAAG}3'\\3'\text{G}\overline{\text{AAG}}\text{C}5'\end{smallmatrix}\right) = \Delta G°_{37}\left(\begin{smallmatrix}5'\text{CGCGAAGGC}3'\\3'\text{GCG}\overline{\text{AAG}}\text{CCG}5'\end{smallmatrix}\right) - \Delta G°_{37}\left(\begin{smallmatrix}5'\text{CGCGGC}3'\\3'\text{GCGCCG}5'\end{smallmatrix}\right) + \Delta G°_{37}\left(\begin{smallmatrix}5'\text{CG}3'\\3'\text{GC}5'\end{smallmatrix}\right) \quad (2b)$$

where, $\Delta G°_{37}\left(\begin{smallmatrix}5'\text{CGCGAAGGC}3'\\3'\text{GCG}\overline{\text{AAG}}\text{CCG}5'\end{smallmatrix}\right)$ is the measured value of the duplex containing the internal loop, $\Delta G°_{37}\left(\begin{smallmatrix}5'\text{CGCGGC}3'\\3'\text{GCGCCG}5'\end{smallmatrix}\right)$ is the measured value of the duplex without the loop (*58*), and $\Delta G°_{37}\left(\begin{smallmatrix}5'\text{CG}3'\\3'\text{GC}5'\end{smallmatrix}\right)$ is the free energy increment for the nearest neighbor base pair interaction interrupted by the internal loop (*58*). Identical calculations can be carried out for $\Delta H°_{\text{loop}}$ and $\Delta S°_{\text{loop}}$. All the thermodynamic parameters used in this calculation are derived from T_M^{-1} versus $\ln(C_T/a)$ plots. In both Tables 1 and 2, sequences are ordered from the least stable to most stable according to loop stability at 37 °C, $\Delta G°_{37,\text{loop}}$.

Models for Predicting Thermodynamic Stabilities of RNA 3 × 3 Internal Loops. Results reported here can be compared to predictions from the model in the current RNAstructure 4.0 algorithm (*24*):

$$\Delta G°_{\text{predicted}} = \Delta G°_{\text{loop initiation}} + \Delta G°_{\text{AU penalty}} + \Delta G°_{\text{UU bonus}} + \Delta G°_{\text{GA bonus}} + \Delta G°_{\text{AG bonus}} \quad (3)$$

At 37 °C, $\Delta G°_{\text{loop initiation}} = 1.9 \pm 0.1$ kcal/mol, $\Delta G°_{\text{AU penalty}} = 0.7 \pm 0.05$ kcal/mol, $\Delta G°_{\text{UU bonus}} = -0.7 \pm 0.1$ kcal/mol, $\Delta G°_{\text{GA bonus}} = -1.0 \pm 0.1$ kcal/mol, and $\Delta G°_{\text{AG bonus}} = -0.8 \pm 0.1$ kcal/mol (*24*). When loops closed with at least one GU pair are omitted, comparison with measured values in Table 2 gives an R^2 of 0.29 and a standard deviation of 0.85 kcal/mol. The poor fit (Figure 2) suggests that the parameters and/or eq 3 can be improved.

Fitting the results reported for 37 °C in Table 2 by multiple linear regression to eq 3 gives the following values: $\Delta G°_{\text{loop initiation}} = 1.67 \pm 0.16$ kcal/mol, $\Delta G°_{\text{AU penalty}} = 0.70 \pm 0.25$ kcal/mol, $\Delta G°_{\text{UU bonus}} = -0.41 \pm 0.18$ kcal/mol, $\Delta G°_{\text{GA bonus}} = -0.78 \pm 0.14$ kcal/mol, and $\Delta G°_{\text{AG bonus}} = -0.28 \pm 0.22$ kcal/mol. Comparison with measured values gives an R^2 of 0.32 and a standard deviation of 0.81 kcal/mol. Evidently, fitting the expanded data set to eq 3 yields little improvement, so eq 3 must be improved.

Measured values at 37 °C show that 3 × 3 loops with middle GA pairs are usually more stable than eq 3 would suggest. In loops with a single loop-terminal GA pair, motifs with a U 3′ to the loop-terminal G of a GA pair are always less stable than expected. Thus, additional terms were added to eq 3: $\Delta G°_{\text{middle bonus}}$ for loops with a GA middle pair and at least one non-pyrimidine-pyrimidine loop-terminal pair and $\Delta G°_{5'\text{GU}/3'\text{AN}}$ for loops with a single loop-terminal GA pair that has a U 3′ to the G of the GA pair. This gives eq 4:

$$\Delta G°_{\text{predicted}} = \Delta G°_{\text{loop initiation}} + \Delta G°_{\text{AU penalty}} + \Delta G°_{\text{UU bonus}} + \Delta G°_{\text{GA bonus}} + \Delta G°_{\text{AG bonus}} + \Delta G°_{\text{middle bonus}} + \Delta G°_{5'\text{GU}/3'\text{AN}} \quad (4)$$

The fit of eq 4 to the data in Table 2 gives the following: $\Delta G°_{\text{loop initiation}} = 1.99 \pm 0.11$ kcal/mol, $\Delta G°_{\text{AU penalty}} = 0.84 \pm 0.15$ kcal/mol, $\Delta G°_{\text{UU bonus}} = -0.49 \pm 0.11$ kcal/mol, $\Delta G°_{\text{GA bonus}} = -0.89 \pm 0.09$ kcal/mol, $\Delta G°_{\text{AG bonus}} = -0.37 \pm 0.14$ kcal/mol, $\Delta G°_{\text{middle bonus}} = -1.22 \pm 0.14$ kcal/mol, and $\Delta G°_{5'\text{GU}/3'\text{AN}} = 0.95 \pm 0.22$ kcal/mol (Table 3). Listed in the $\Delta G°_{\text{predicted}}$ column in Table 2 are the free energy increments of internal loop formation at 37 °C predicted by this model. The correlation with measured values gives an R^2 of 0.76 and a standard deviation of 0.49 kcal/mol (Figure 2). A fit with no bonus for AG in the context $\frac{GA}{C\overline{G}}$ gives a similar R^2 of 0.75 with a standard deviation of 0.50 kcal/mol. Evidently, a middle GA pair in a 3 × 3 loop can enhance stability relative to that predicted by eq 3.

Imino Proton NMR Spectra. Imino protons in canonical AU and GC pairs typically resonate in the regions of 13−15 and 12−13.5 ppm, respectively (*66*), which helps confirm the formation of designed secondary structures. Imino protons in imino hydrogen-bonded GA pairs and sheared GA pairs typically resonate near 11.5−12.5 and 10−11 ppm, respectively (*66*−*70*). Non-hydrogen-bonded

Table 1: Thermodynamic Parameters for Duplex Formation

	T_M^{-1} vs ln(C_T/a) plots				average of melt curve fits			
sequence	$-\Delta H°$ (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)
Sequences Used To Derive Parameters for 3 × 3 Loops								
GCG**AAA**CGC CGC**AAA**GCG	36.0 ± 1.4	102.3 ± 4.6	4.44 ± 0.04	26.7	34.4 ± 7.2	95.9 ± 23.4	4.61 ± 0.11	27.6
CGC**AAA**GGC GCG**ACC**CCG	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
CGG**AAA**CGC GCC**AAA**GCG	48.7 ± 1.8	140.9 ± 6.0	5.02 ± 0.07	27.6	41.2 ± 7.4	114.4 ± 25.7	5.42 ± 0.48	28.8
CGC**AUA**GGC GCG**AUG**CCG	41.5 ± 2.6	114.3 ± 8.7	6.06 ± 0.08	33.5	46.0 ± 8.8	129.0 ± 28.8	6.00 ± 0.26	33.5
UGAC**AAA**CUCA[b] ACUG**AAA**GAGU	59.4	170.1	6.67	37.7	57.6	164.1	6.74	38.1
CGC**AUA**GGC GCG**AAA**CCG	52.2 ± 2.8	148.5 ± 9.1	6.15 ± 0.06	34.8	57.6 ± 13.8	165.8 ± 45.0	6.14 ± 0.18	34.9
GAGU**GAA**UGAC CUCA**AGA**ACUG	87.0 ± 10.8	257.9 ± 34.7	6.98 ± 0.28	38.6	81.6 ± 12.7	240.3 ± 41.3	7.08 ± 0.20	39.1
GAGC**AGA**CGAC CUCG**AGA**GCUG	57.7 ± 0.7	159.1 ± 2.3	8.39 ± 0.01	47.3	49.9 ± 3.8	134.2 ± 11.8	8.23 ± 0.14	47.9
CGC**AGA**GGC GCG**AUG**CCG	41.1 ± 2.5	112.1 ± 8.0	6.27 ± 0.08	35.0	45.8 ± 9.0	127.4 ± 29.1	6.29 ± 0.15	35.3
GAGC**GUA**CGAC CUCG**AUA**GCUG	51.9 ± 6.0	140.0 ± 18.9	8.53 ± 0.24	49.4	44.9 ± 3.2	117.6 ± 10.1	8.41 ± 0.13	50.6
CGC**AAA**GGC GCG**AUG**CCG	52.0 ± 2.8	146.5 ± 9.2	6.54 ± 0.06	37.1	52.0 ± 9.6	146.1 ± 31.0	6.64 ± 0.15	37.6
GAGC**AGA**CGAC CUCG**AUA**GCUG	71.9 ± 5.5	203.5 ± 17.2	8.74 ± 0.14	46.8	70.2 ± 5.2	198.2 ± 16.1	8.71 ± 0.17	46.9
UGAC**CAA**CUCA[b] ACUG**AAA**GAGU	67.5	194.6	7.14	39.8	55.1	154.4	7.22	40.9
UGAC**AAA**CUCA[b] ACUG**AAC**GAGU	62.9	179.7	7.17	40.2	61.6	175.2	7.24	40.6
GAGC**AGA**CGAC CUCG**AUG**GCUG	73.4 ± 3.8	208.5 ± 11.8	8.78 ± 0.11	46.8	76.0 ± 7.4	216.7 ± 23.7	8.83 ± 0.19	46.7
CGG**AAA**CCG[b] GCC**AAA**GGC	38.4	108.9	4.64	28.9	36.9	103.5	4.79	29.7
CGG**CAC**CCG GCC**CAC**GGC	32.3 ± 0.8	89.0 ± 2.7	4.66 ± 0.04	27.5	32.6 ± 5.8	89.6 ± 19.9	4.78 ± 0.36	28.7
GAGC**GGA**CGAC CUCG**AUA**GCUG	80.5 ± 8.5	230.9 ± 26.9	8.91 ± 0.28	46.5	72.5 ± 7.5	205.6 ± 23.5	8.76 ± 0.26	46.8
CGC**AAA**GGC GCG**AAA**CCG	49.7 ± 2.5	138.4 ± 8.1	6.77 ± 0.05	38.5	46.6 ± 7.1	127.9 ± 23.3	6.89 ± 0.10	39.4
GCA**GAA**UGC CGU**AAG**ACG	42.7 ± 2.3	122.9 ± 7.8	4.62 ± 0.09	29.5	39.9 ± 13.9	110.7 ± 46.4	5.00 ± 0.56	31.8
CGC**UCU**GGC GCG**UCU**CCG	65.3 ± 3.9	188.3 ± 12.6	6.88 ± 0.07	38.6	67.5 ± 8.0	195.4 ± 25.8	6.94 ± 0.13	38.9
GAGC**AAA**CGAC CUCG**AUG**GCUG	78.1 ± 2.0	222.4 ± 6.3	9.11 ± 0.06	47.6	87.5 ± 4.8	251.8 ± 15.2	9.37 ± 0.17	47.4
CGC**AAA**GGC GCG**AAC**CCG	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
GAGC**CGA**CGAC CUCG**AGA**GCUG	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
GAGC**AAA**CGAC CUCG**AGA**GCUG	79.0 ± 1.8	225.1 ± 5.7	9.18 ± 0.04	47.8	69.3 ± 3.9	194.6 ± 12.2	8.96 ± 0.16	48.3
CGAC**GCA**GCAG GCUG**AAA**CGUC	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
CGG**ACA**CCG GCC**ACA**GGC	31.4 ± 1.1	84.9 ± 3.7	5.04 ± 0.05	30.8	34.3 ± 0.9	94.2 ± 13.0	5.10 ± 0.33	31.9
GAGC**UGC**CGAC CUCG**UAU**GCUG	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
GAGC**CGA**CGAC CUCG**AUA**GCUG	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
GAGC**GAA**CGAC CUCG**AAA**GCUG	81.4 ± 4.4	232.4 ± 13.8	9.29 ± 0.11	47.9	80.4 ± 1.5	229.3 ± 4.8	9.27 ± 0.06	47.9
GCU**GAA**AGC CGA**AAG**UCG	39.8 ± 2.5	111.8 ± 8.4	5.21 ± 0.09	33.4	38.4 ± 1.2	106.4 ± 39.8	5.41 ± 0.32	34.9
GAGC**GAA**CGAC CUCG**AUA**GCUG	82.1 ± 1.0	234.6 ± 3.0	9.35 ± 0.02	48.0	78.8 ± 1.8	224.2 ± 5.6	9.27 ± 0.07	48.2
GAGC**GAG**CGAC CUCG**AUA**GCUG	72.5 ± 2.5	203.3 ± 7.7	9.40 ± 0.07	49.8	66.9 ± 3.5	185.9 ± 10.8	9.23 ± 0.14	50.0
GAGC**CGA**CGAC CUCG**UAA**GCUG	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
GGC**GAA**GCC CCG**AAG**CGG	53.3 ± 3.1	146.9 ± 9.7	7.72 ± 0.11	49.3	53.4 ± 11.8	146.9 ± 37.0	7.86 ± 0.39	50.2
CGG**AAG**CGC GCC**GUA**GCG	33.5 ± 1.3	87.6 ± 4.1	6.28 ± 0.03	34.7	31.2 ± 6.6	79.8 ± 21.9	6.47 ± 0.24	36.4
GAGC**GAG**CGAC CUCG**AAA**GCUG	77.0 ± 4.1	217.3 ± 13.0	9.60 ± 0.12	49.9	70.1 ± 2.9	195.8 ± 8.8	9.39 ± 0.15	50.2
GAGC**GAG**CGAC CUCG**AAG**GCUG	78.9 ± 1.6	223.3 ± 5.2	9.61 ± 0.05	49.6	70.0 ± 1.3	195.4 ± 4.0	9.37 ± 0.09	50.1
CGC**GAA**GGC GCG**ACC**CCG	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
GAGC**UGU**CGAC CUCG**UAU**GCUG	83.1 ± 1.7	236.6 ± 5.4	9.69 ± 0.06	49.3	84.8 ± 4.2	241.9 ± 13.0	9.73 ± 0.14	49.2
GAGC**AGA**CGAC CUCG**AAA**GCUG	81.3 ± 3.1	230.5 ± 9.8	9.77 ± 0.09	49.9	84.5 ± 3.3	240.5 ± 10.4	9.87 ± 0.13	49.8
CGC**GAA**GGC GCG**AAA**CCG	68.8 ± 6.3	197.1 ± 20.2	7.63 ± 0.17	42.0	57.2 ± 4.7	160.1 ± 15.2	7.57 ± 0.20	42.7
CGG**AAG**CGC GCC**GAA**GCG	71.2 ± 3.0	208.3 ± 9.8	6.62 ± 0.04	37.4	62.1 ± 4.9	178.7 ± 15.7	6.66 ± 0.15	37.7
GAGC**UGA**CGAC CUCG**UAA**GCUG	88.8 ± 2.7	254.5 ± 8.3	9.89 ± 0.09	49.2	87.2 ± 2.5	249.5 ± 8.0	9.83 ± 0.07	49.2
GCG**UUU**CGC CGC**UUU**GCG	62.6 ± 2.5	181.0 ± 7.9	6.41 ± 0.03	40.7	63.5 ± 10.7	183.5 ± 34.2	6.55 ± 0.13	41.3

Table 1 (Continued)

sequence	T_M^{-1} vs ln(C_T/a) plots: $-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)	average of melt curve fits: $-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_m[a] (°C)
Sequences Used To Derive Parameters for 3 × 3 Loops								
CGC**UCU**GGC GCG**UUU**CCG	70.8 ± 3.8	203.2 ± 12.3	7.74 ± 0.07	42.4	67.9 ± 6.5	193.9 ± 20.9	7.77 ± 0.08	42.7
CGAC**GCA**GCAG GCUG**AAG**CGUC	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
CGC**UUU**GGC GCG**UCU**CCG	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
CGC**AAG**GCG GCG**GAA**CGC	39.8 ± 0.9	108.9 ± 3.1	6.05 ± 0.01	39.9	37.7 ± 11.6	101.7 ± 37.5	6.15 ± 0.19	40.9
GCG**GAA**CGG CGC**AUG**GCC	61.3 ± 2.2	176.2 ± 7.1	7.26 ± 0.03	40.7	55.9 ± 4.1	156.8 ± 13.2	7.25 ± 0.07	41.0
CGC**UUU**GGC GCG**UUU**CCG	57.9 ± 5.8	161.4 ± 18.4	7.83 ± 0.15	44.1	64.7 ± 4.5	183.0 ± 14.1	7.90 ± 0.13	43.7
GAGC**CGA**CGAC CUCG**AAA**GCUG	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
GAGC**UGC**CGAC CUCG**UAA**GCUG	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
CGC**AGA**GGC GCG**AAA**CCG	55.0 ± 5.3	151.6 ± 16.8	7.94 ± 0.15	45.1	63.6 ± 2.7	179.3 ± 8.4	8.04 ± 0.15	44.5
CGC**GAA**GGC GCG**AUG**CCG	66.5 ± 2.4	188.7 ± 7.6	7.95 ± 0.04	43.8	60.8 ± 6.3	170.6 ± 20.0	7.91 ± 0.13	44.2
GAGC**GAG**CGAC CUCG**AGA**GCUG	76.8 ± 4.2	215.0 ± 13.0	10.14 ± 0.15	52.3	73.5 ± 3.7	204.8 ± 11.5	10.01 ± 0.16	52.4
CGC**AUA**GGC GCG**AAG**CCG	53.7 ± 1.4	147.4 ± 4.4	8.01 ± 0.02	45.8	59.1 ± 3.0	164.4 ± 9.3	8.08 ± 0.11	45.4
GCG**GAA**CGC CGC**AAG**GCG	60.1 ± 3.9	172.0 ± 12.5	6.71 ± 0.08	42.4	62.5 ± 3.9	179.7 ± 12.3	6.77 ± 0.16	42.5
GCG**GUA**CGG CGC**AUG**GCC	61.6 ± 2.9	173.7 ± 9.1	7.67 ± 0.01	42.9	52.1 ± 7.8	142.3 ± 24.1	7.44 ± 0.40	42.6
CGC**AAA**GGC GCG**AAG**CCG	82.2 ± 6.1	238.4 ± 19.3	8.26 ± 0.14	43.7	74.0 ± 14.5	212.5 ± 46.5	8.04 ± 0.21	43.4
GAGC**GAA**CGAC CUCG**AGA**GCUG	90.1 ± 2.5	256.5 ± 7.9	10.52 ± 0.09	51.4	82.0 ± 3.1	231.5 ± 9.6	10.25 ± 0.13	51.7
CGC**GAA**GGC GCG**AAG**CCG	62.1 ± 4.2	173.1 ± 13.1	8.37 ± 0.13	46.5	68.1 ± 7.4	192.1 ± 23.6	8.56 ± 0.23	46.5
CGAC**CGA**GCAG GCUG**AAA**CGUC	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
GAGC**GAA**CGAC CUCG**AAG**GCUG	86.8 ± 4.0	245.2 ± 12.5	10.75 ± 0.17	52.9	88.7 ± 2.8	251.0 ± 8.7	10.83 ± 0.14	52.8
CGC**GAA**GGC GCG**AAC**CCG	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
GAGC**AGA**CGAC CUCG**AAG**GCUG	82.0 ± 8.9	229.0 ± 27.6	10.92 ± 0.43	54.5	88.9 ± 4.2	250.4 ± 12.5	11.21 ± 0.30	54.2
GAGC**CGA**CGAC CUCG**AAG**GCUG	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
CGC**AGA**GGC GCG**AAG**CCG	60.0 ± 6.6	164.4 ± 20.6	8.97 ± 0.26	50.1	65.4 ± 1.9	181.5 ± 5.8	9.12 ± 0.15	49.8
CGAC**CGA**GCAG GCUG**AAG**CGUC	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
GAGC**GGA**CGAC CUCG**AAA**GCUG	87.4 ± 2.5	245.3 ± 7.6	11.33 ± 0.11	55.0	95.3 ± 1.7	269.6 ± 5.2	11.69 ± 0.08	54.7
CGC**GAA**GGC GCG**AGG**CCG	70.2 ± 4.3	193.2 ± 13.3	10.25 ± 0.20	54.3	74.7 ± 3.6	207.3 ± 11.1	10.44 ± 0.26	54.1
GAGC**GGA**CGAC CUCG**AAG**GCUG	92.1 ± 1.9	256.7 ± 5.9	12.48 ± 0.11	58.4	94.9 ± 4.0	265.1 ± 12.1	12.64 ± 0.25	58.4
Canonically Paired Stems								
UCCGCC AGGUGG	57.0 ± 3.2 (50.1)[c]	162.2 ± 10.5 (134.9)[c]	6.71 ± 0.06 (8.25)[c]	38.0 (48.1)[c]	54.3 ± 4.0	153.4 ± 12.7	6.69 ± 0.13	37.9
GAGUUGAC CUCAACUG	88.7 ± 11.0 (71.8)[c]	251.8 ± 33.8 (198.5)[c]	10.57 ± 0.57 (10.21)[c]	51.8 (54.0)[c]	79.9 ± 5.9	224.6 ± 18.6	10.23 ± 0.29	52.0
Sequences Not Used To Derive Parameters for 3 × 3 Loops								
Sequences with GU Closing Base Pairs								
GAGU**AAA**UGAC CUCG**AAA**GCUG	65.1 ± 6.6	196.5 ± 22.2	4.14 ± 0.31	26.0	64.3 ± 5.0	193.8 ± 16.6	4.21 ± 0.16	26.2
GAGU**AAA**UGAC CUCG**AAG**GCUG	76.9 ± 2.8	230.9 ± 9.2	5.28 ± 0.07	32.1	82.6 ± 3.4	249.6 ± 11.0	5.17 ± 0.06	32.0
GAGU**GAA**UGAC CUCG**AAA**GCUG	76.3 ± 5.6	228.4 ± 18.5	5.46 ± 0.15	32.7	78.9 ± 8.6	237.0 ± 28.0	5.43 ± 0.13	32.7
GCU**AAA**GGC CGG**AAA**UCG	32.1 ± 2.3	90.5 ± 7.9	4.00 ± 0.15	21.6	38.8 ± 10.2	112.7 ± 34.0	3.81 ± 0.43	22.8
GAGU**GAA**UGAC CUCG**AAG**GCUG	103.4 ± 5.3	306.9 ± 17.0	8.25 ± 0.08	42.2	96.5 ± 10.1	284.7 ± 32.3	8.23 ± 0.14	42.6
CCAGC**CAA**GUCCU [d] GGUUG**ACG**UAGGA	94.5 ± 1.9	272.1 ± 5.9	10.1 ± 0.1	49.2	98.9 ± 1.2	285.8 ± 3.6	10.3 ± 0.1	49.2
UCC**GAA**GCC AGG**AGG**UGG	74.8 ± 5.8	213.1 ± 18.3	8.72 ± 0.18	46.4	83.8 ± 7.3	241.4 ± 22.6	8.97 ± 0.37	46.3
Non-Two-State Sequences								
CGC**AAA**GCG[b] GCG**AAA**CGC	45.9	132.3	4.88	31.7	30.1	80.3	5.19	32.1
CGC**GCA**GGC GCG**AAG**CCG	64.1 ± 4.7	179.3 ± 14.9	8.51 ± 0.11	46.9	41.2 ± 2.9	106.5 ± 9.7	8.17 ± 0.21	49.9
CGC**AAA**GGC GCG**AGG**CCG	85.9 ± 4.4	244.9 ± 13.7	9.92 ± 0.13	49.8	61.2 ± 3.1	167.6 ± 9.6	9.21 ± 0.23	51.2

[a] At a C_T of 0.1 mM. [b] Data from ref *56*. [c] Parameters listed in parentheses are calculated from nearest neighbor model in ref *58*. [d] From ref *72*.

imino protons have broad peaks or no observable peaks due to exchange with water.

Figure 3 shows imino proton spectra for some of the duplexes studied here. For symmetric self-complementary

Table 2: Thermodynamic Parameters for Internal Loop Formation[a]

sequence	$\Delta G^\circ_{37,\mathrm{loop}}$ (kcal/mol)	$\Delta H^\circ_{\mathrm{loop}}$ (kcal/mol)	$\Delta S^\circ_{\mathrm{loop}}$ (eu)	$\Delta G^\circ_{\mathrm{predicted}}$ (kcal/mol; from Table 3)
Sequences Used To Derive Parameters for 3 × 3 Loops				
GCG**AAA**CGC CGC**AAA**GCG	2.76 ± 0.43	15.1 ± 8.2	39.3 ± 25.0	1.99
CGC**AAA**GGC GCG**ACC**CCG	2.65 ± 0.45	2.4 ± 7.8	−1.0 ± 23.6	1.99
CGG**AAA**CGC GCC**AAA**GCG	2.47 ± 1.31	−5.7 ± 7.3	−26.5 ± 21.8	1.99
CGC**AUA**GGC GCG**AUG**CCG	2.46 ± 0.45	7.6 ± 7.8	16.4 ± 23.5	2.05
UGAC**AAA**CUCA$_b$ ACUG**AAA**GAGU	2.41 ± 0.54	3.3 ± 11.0	2.8 ± 32.8	1.99
CGC**AUA**GGC GCG**AAA**CCG	2.37 ± 0.45	−3.2 ± 7.9	−17.8 ± 23.7	1.99
GAGU**GAA**UGAC CUCA**AGA**ACUG	2.30 ± 0.32[c]	−22.0 ± 11.3[c]	−78.4 ± 36.1[c]	1.56
GAGC**AGA**CGAC CUCG**AGA**GCUG	2.29 ± 0.56	9.0 ± 9.7	22.2 ± 29.2	1.99
CGC**AGA**GGC GCG**AUG**CCG	2.25 ± 0.45	8.0 ± 7.8	18.6 ± 23.2	2.05
GAGC**GUA**CGAC CUCG**AUA**GCUG	2.15 ± 0.61	14.8 ± 11.4	41.3 ± 34.7	2.05
CGC**AAA**GGC GCG**AUG**CCG	1.98 ± 0.45	−3.0 ± 7.9	−15.8 ± 23.7	2.05
GAGC**AGA**CGAC CUCG**AUA**GCUG	1.94 ± 0.58	−5.2 ± 11.1	−22.2 ± 33.8	1.99
UGAC**CAA**CUCA$_b$ ACUG**AAA**GAGU	1.94 ± 0.54	−4.8 ± 11.4	−21.7 ± 34.1	1.99
UGAC**AAA**CUCA$_b$ ACUG**AAC**GAGU	1.91 ± 0.54	−0.2 ± 11.2	−6.8 ± 33.3	1.99
GAGC**AGA**CGAC CUCG**AUG**GCUG	1.90 ± 0.57	−6.7 ± 10.4	−27.2 ± 31.4	2.05
CGG**AAA**CCG$_b$ GCC**AAA**GGC	1.84 ± 0.43	0.8 ± 7.7	−3.2 ± 22.7	1.99
CGG**CAC**CCG GCC**CAC**GGC	1.82 ± 0.41	6.9 ± 6.7	16.7 ± 20.1	1.99
GAGC**GGA**CGAC CUCG**AUA**GCUG	1.77 ± 0.63	−13.8 ± 12.9	−49.6 ± 39.7	1.10
CGC**AAA**GGC GCG**AAA**CCG	1.75 ± 0.45	−0.7 ± 7.8	−7.7 ± 23.3	1.99
GCA**GAA**UGC CGU**AAG**ACG	1.66 ± 0.32	10.3 ± 8.0	27.6 ± 25.7	1.89
CGC**UCU**GGC GCG**UCU**CCG	1.64 ± 0.45	−16.3 ± 8.3	−57.6 ± 25.2	1.01
GAGC**AAA**CGAC CUCG**AUG**GCUG	1.57 ± 0.57	−11.4 ± 9.9	−41.1 ± 29.8	2.05
CGC**AAA**GGC GCG**AAC**CCG	1.56 ± 0.45	−12.1 ± 8.5	−43.8 ± 25.9	1.99
GAGC**CGA**CGAC CUCG**AGA**GCUG	1.52 ± 0.57	−3.1 ± 9.9	−14.3 ± 30.0	1.99
GAGC**AAA**CGAC CUCG**AGA**GCUG	1.50 ± 0.56	−12.3 ± 9.9	−43.8 ± 29.7	0.77
CGAC**GCA**GCAG GCUG**AAA**CGUC	1.50 ± 0.51	−16.2 ± 9.2	−57.5 ± 27.7	1.10
CGG**ACA**CCG GCC**ACA**GGC	1.44 ± 0.41	7.8 ± 6.8	20.8 ± 20.2	1.99
GAGC**UGC**CGAC CUCG**UAU**GCUG	1.44 ± 0.56	−14.3 ± 9.9	−50.0 ± 29.7	1.50
GAGC**CGA**CGAC CUCG**AUA**GCUG	1.43 ± 0.58	−13.0 ± 10.9	−45.9 ± 32.9	1.99
GAGC**GAA**CGAC CUCG**AAA**GCUG	1.39 ± 0.57	−14.7 ± 10.6	−51.1 ± 32.2	1.10
GCU**GAA**AGC CGA**AAG**UCG	1.38 ± 0.34	11.6 ± 7.8	32.8 ± 24.6	1.89
GAGC**GAA**CGAC CUCG**AUA**GCUG	1.33 ± 0.56	−15.4 ± 9.7	−53.3 ± 29.3	1.10
GAGC**GAG**CGAC CUCG**AUA**GCUG	1.28 ± 0.57	−5.8 ± 10.0	−22.0 ± 30.1	0.73
GAGC**CGA**CGAC CUCG**UAA**GCUG	1.26 ± 0.57	−12.9 ± 10.2	−44.9 ± 30.7	0.77
GGC**GAA**GCC CCG**AAG**CGG	1.25 ± 0.48	3.8 ± 8.9	8.4 ± 26.9	0.21
CGG**AAG**CGC GCC**GUA**GCG	1.21 ± 1.31	9.5 ± 7.2	26.8 ± 21.4	1.25
GAGC**GAG**CGAC CUCG**AAA**GCUG	1.08 ± 0.58	−10.3 ± 10.5	−36.0 ± 31.9	0.73
GAGC**GAG**CGAC CUCG**AAG**GCUG	1.07 ± 0.56	−12.2 ± 9.8	−42.0 ± 29.6	1.10
CGC**GAA**GGC GCG**ACC**CCG	1.00 ± 0.47	−9.9 ± 8.8	−35.1 ± 26.8	1.10
GAGC**UGU**CGAC CUCG**UAU**GCUG	0.99 ± 0.57	−16.4 ± 9.9	−55.3 ± 29.6	1.01
GAGC**AGA**CGAC CUCG**AAA**GCUG	0.91 ± 0.57	−14.6 ± 10.2	−49.2 ± 30.7	0.77
CGC**GAA**GGC GCG**AAA**CCG	0.89 ± 0.48	−19.8 ± 9.7	−66.4 ± 29.7	1.10
CGG**AAG**CGC GCC**GAA**GCG	0.87 ± 1.31	−28.2 ± 7.7	−93.9 ± 23.2	1.25
GAGC**UGA**CGAC CUCG**UAA**GCUG	0.79 ± 0.57	−22.1 ± 10.1	−73.2 ± 30.3	0.28
GCG**UUU**CGC CGC**UUU**GCG	0.79 ± 0.43	−11.5 ± 8.5	−39.4 ± 25.8	1.01
CGC**UCU**GGC GCG**UUU**CCG	0.78 ± 0.45	−21.8 ± 8.3	−72.5 ± 25.1	1.01

Table 2 (Continued)

sequence	$\Delta G^\circ_{37,loop}$ (kcal/mol)	ΔH°_{loop} (kcal/mol)	ΔS°_{loop} (eu)	$\Delta G^\circ_{predicted}$ (kcal/mol; from Table 3)
		Sequences Used To Derive Parameters for 3 × 3 Loops		
CGAC**GCA**GCAG GCUG**AAG**CGUC	0.73 ± 0.50	−16.0 ± 8.8	−54.3 ± 26.3	0.21
CGC**UUU**GGC GCG**UCU**CCG	0.72 ± 0.45	−6.3 ± 8.1	−22.3 ± 24.2	1.01
CGC**AAG**GCG GCG**GAA**CGC	0.71 ± 0.38	4.1 ± 6.8	10.8 ± 20.6	1.25
GCG**GAA**CGG CGC**AUG**GCC	0.70 ± 0.46	−5.0 ± 9.0	−20.4 ± 27.4	0.21
CGC**UUU**GGC GCG**UUU**CCG	0.69 ± 0.47	−8.9 ± 9.4	−30.7 ± 28.6	1.01
GAGC**CGA**CGAC CUCG**AAA**GCUG	0.69 ± 0.58	−16.4 ± 10.2	−54.6 ± 30.8	0.77
GAGC**UGC**CGAC CUCG**UAA**GCUG	0.63 ± 0.59	−20.1 ± 10.8	−66.3 ± 32.6	0.28
CGC**AGA**GGC GCG**AAA**CCG	0.58 ± 0.47	−6.0 ± 9.1	−20.9 ± 27.6	0.77
CGC**GAA**GGC GCG**AUG**CCG	0.57 ± 0.45	−17.5 ± 7.7	−58.0 ± 23.1	0.21
GAGC**GAG**CGAC CUCG**AGA**GCUG	0.54 ± 0.58	−10.1 ± 10.6	−33.7 ± 31.9	−0.49
CGC**AUA**GGC GCG**AAG**CCG	0.51 ± 0.45	−4.7 ± 7.5	−16.7 ± 22.3	1.10
GCG**GAA**CGC CGC**AAG**GCG	0.49 ± 0.44	−9.0 ± 9.0	−30.4 ± 27.6	0.21
GCG**GUA**CGG CGC**AUG**GCC	0.29 ± 0.46	−5.3 ± 9.2	−17.9 ± 28.0	0.21
CGC**AAA**GGC GCG**AAG**CCG	0.26 ± 0.47	−33.2 ± 9.6	−107.7 ± 29.1	1.10
GAGC**GAA**CGAC CUCG**AGA**GCUG	0.16 ± 0.57	−23.4 ± 10.0	−75.2 ± 30.2	−0.12
CGC**GAA**GGC GCG**AAG**CCG	0.15 ± 0.46	−13.1 ± 8.5	−42.4 ± 25.5	0.21
CGAC**CGA**GCAG GCUG**AAA**CGUC	0.05 ± 0.51	−25.3 ± 8.8	−82.2 ± 26.5	0.77
GAGC**GAA**CGAC CUCG**AAG**GCUG	−0.07 ± 0.59	−20.1 ± 10.5	−63.9 ± 31.7	0.21
CGC**GAA**GGC GCG**AAC**CCG	−0.13 ± 0.46	−31.7 ± 9.1	−101.6 ± 27.7	1.10
GAGC**AGA**CGAC CUCG**AAG**GCUG	−0.24 ± 0.71	−15.3 ± 13.2	−47.7 ± 40.1	−0.12
GAGC**CGA**CGAC CUCG**AAG**GCUG	−0.36 ± 0.58	−6.6 ± 10.1	−19.5 ± 30.3	−0.12
CGC**AGA**GGC GCG**AAG**CCG	−0.45 ± 0.52	−11.0 ± 9.9	−33.7 ± 30.0	−0.12
CGAC**CGA**GCAG GCUG**AAG**CGUC	−0.60 ± 0.53	−19.6 ± 9.4	−61.6 ± 28.3	−0.12
GAGC**GGA**CGAC CUCG**AAA**GCUG	−0.65 ± 0.57	−20.7 ± 10.0	−64.0 ± 30.1	−0.05
CGC**GAA**GGC GCG**AGG**CCG	−1.73 ± 0.49	−21.2 ± 8.5	−62.5 ± 25.6	−1.01
GAGC**GGA**CGAC CUCG**AAG**GCUG	−1.80 ± 0.57	−25.4 ± 9.9	−75.4 ± 29.7	−1.01
		Sequences Not Used To Derive Parameters for 3 × 3 Loops		
		Sequences with GU Closing Base Pairs[d]		
GAGU**AAA**UGAC CUCG**AAA**GCUG	4.73 ± 1.06	−3.8 ± 13.9	−27.2 ± 44.6	3.67
GAGU**AAA**UGAC CUCG**AAG**GCUG	3.59 ± 1.02	−15.6 ± 12.6	−61.6 ± 39.7	2.78
GAGU**GAA**UGAC CUCG**AAA**GCUG	3.41 ± 1.03	−15.0 ± 13.5	−59.1 ± 42.9	2.78
GCU**AAA**GGC CGG**AAA**UCG	2.79 ± 0.57	17.7 ± 8.6	48.4 ± 27.2	3.67
GAGU**GAA**UGAC CUCG**AAG**GCUG	0.62 ± 1.02	−42.1 ± 13.4	−137.6 ± 42.2	1.89
CCAGC**CAA**GUCCU [e] GGUUG**ACG**UAGGA	0.1 ± 0.3	−12.2 ± 4.4	−39.6 ± 34.4	1.94
UCC**GAA**GCC AGG**AGG**UGG	−3.42 ± 0.40	−23.4 ± 9.2	−64.4 ± 29.3	−0.17
		Non-Two-State Sequences		
CGC**AAA**GCG[b] GCG**AAA**CGC	1.88 ± 0.4	−2.0 ± 8.2	−12.6 ± 24.3	1.99
CGC**GCA**GGC GCG**AAG**CCG	0.01 ± 0.46	−15.1 ± 8.7	−48.6 ± 26.4	0.21
CGC**AAA**GGC GCG**AGG**CCG	−1.40 ± 0.46	−36.9 ± 8.6	−114.2 ± 25.8	−0.12

[a] Internal loops are in bold with a line in the middle of the loop. Experimental error for ΔG°_{37}, ΔH°, and ΔS° for the canonical stems are estimated to be 4, 12, and 13.5%, respectively, according to ref *58*. [b] Data from ref *56*. [c] Nearest neighbor value of Watson−Crick stem used to derive loop thermodynamics. [d] $\mathbf{\Delta}G^\circ_{37(predicted)}$ is predicted assuming the penalty of GU is the same as that of AU. [e] From ref *72*.

duplexes, each resonance is attributed to two imino protons. Some assignments are based on NMR melting and comparison with similar duplexes having 2 × 2 and 2 × 3 internal loops. For example, resonances from closing or stem-terminal GC base pairs are relatively broader (Figure 3g,n) and disappear sooner than resonances from stem-internal GC pairs when the temperature is increased (data not shown). Imino proton resonances from canonical AU pairs closing internal loops can shift upfield beyond 13 ppm (Figure 3n), which was also observed in 2 × 2 loops (*71*, *72*). The imino

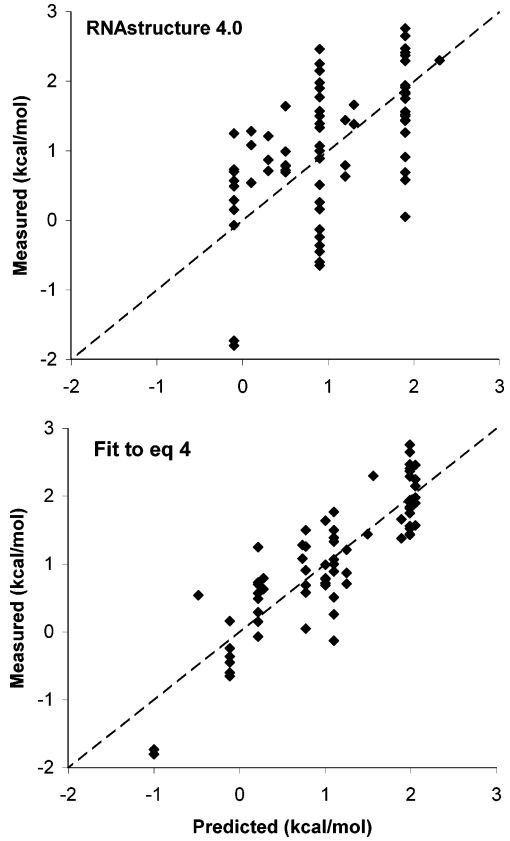

FIGURE 2: Comparisons between predicted and measured free energies for 3 × 3 loops in the current RNAstructure 4.0 program (eq 3; $R^2 = 0.29$, standard deviation = 0.85 kcal/mol) and the model of eq 4 ($R^2 = 0.76$, standard deviation = 0.49 kcal/mol).

Table 3: Free Energy Parameters at 37 °C for 3 × 3 Internal Loops[a]

$\Delta G^\circ_{\text{loop initiation}}$ (kcal/mol)	1.99 ± 0.11
$\Delta G^\circ_{\text{AU penalty}}$ (kcal/mol)	0.84 ± 0.15
$\Delta G^\circ_{\text{UU bonus}}$ (kcal/mol)[b]	−0.49 ± 0.11
$\Delta G^\circ_{\text{GA bonus}}$ (kcal/mol)[b]	−0.89 ± 0.09
$\Delta G^\circ_{\text{AG bonus}}$ (kcal/mol)[b]	−0.37 ± 0.14
$\Delta G^\circ_{\text{middle bonus}}$ (kcal/mol)[c]	−1.22 ± 0.14
$\Delta G^\circ_{\text{5′GU/3′AN}}$ (kcal/mol)[d]	0.95 ± 0.22

[a] These parameters are used to predict the free energy of 3 × 3 internal loops according to eq 4. [b] Applied for loop-terminal pairs. [c] Applied for the GA middle pair with at least one non-pyrimidine-pyrimidine loop-terminal pair. [d] Applied for a single GA loop-terminal pair with a U 3′ to the G.

proton NMR spectra of two particularly stable internal loops with three consecutive GA pairs, $\frac{\text{CGAAG}}{\text{GAGGC}}$ (−1.73 kcal/mol) and $\frac{\text{CGAAG}}{\text{GAGGU}}$ (−3.42 kcal/mol), have sharp peaks for the loops (Figure 3h,l), suggesting structure. For $\frac{\text{CGCGAAGGC}}{\text{GCGAGGCCG}}$, a two-dimensional SNOESY spectrum (Figure S1 of the Supporting Information) suggests that four and two peaks are overlapped at 13.2 and 12.7 ppm, respectively, indicating the expected number of base pairs.

For some designed sequences, extremely broad peaks or more than the expected number of peaks in the region of 13−15 ppm were observed (data not shown), indicating dynamics between multiple secondary structures. All these sequences were excluded from thermodynamic analysis.

DISCUSSION

Determination of RNA native structures is necessary for understanding their mechanisms of action in cellular processes and for the design of therapeutics that target RNA. The first stage is determination of RNA secondary structure. Thermodynamic parameters of internal loops are important for the prediction of RNA secondary structure (*23*, *24*, *73*−*77*).

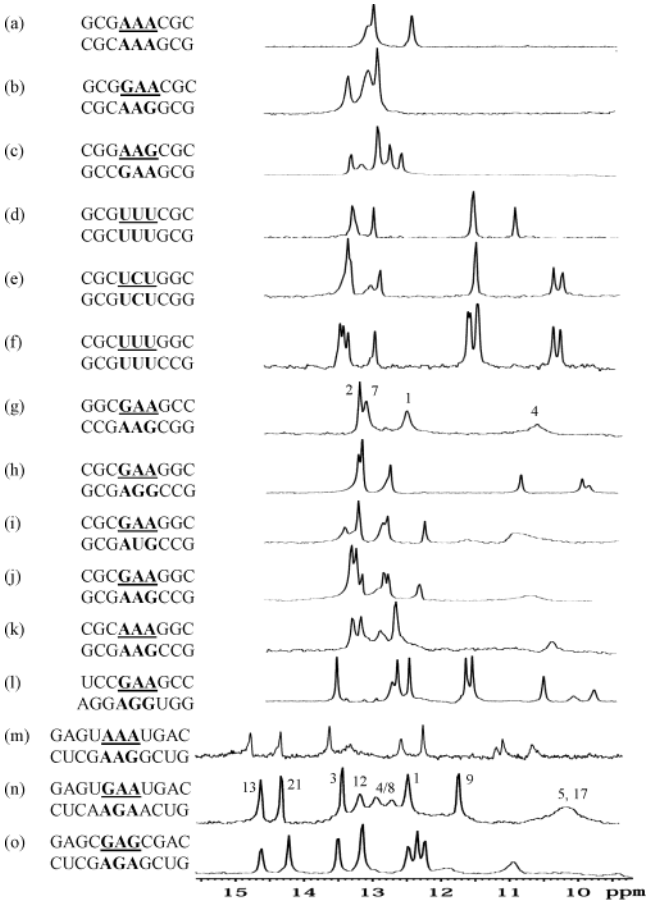

FIGURE 3: One-dimensional imino proton NMR spectra at 10 °C unless noted otherwise of the duplexes (a) $\frac{\text{GCGAAACGC}}{\text{CGCAAAGCG}}$, (b) $\frac{\text{GCGGAACGC}}{\text{CGCAAGGCG}}$, (c) $\frac{\text{CGGAAGCGC}}{\text{GCCGAAGCG}}$, (d) $\frac{\text{GCGUUUCGC}}{\text{CGCUUUGCG}}$, (e) $\frac{\text{CGCUCUGGC}}{\text{GCGUCUCCG}}$, (f) $\frac{\text{CGCUUUGGC}}{\text{GCGUUUCCG}}$, (g) $\frac{\text{GGCGAAGCC}}{\text{CCGAAGCGG}}$, (h) $\frac{\text{CGCGAAGGC}}{\text{GCGAGGCCG}}$, (i) $\frac{\text{CGCGAAGGC}}{\text{GCGAUGCCG}}$, (j) $\frac{\text{CGCGAAGGC}}{\text{GCGAAGCCG}}$, (k) $\frac{\text{CGCAAAGGC}}{\text{GCGAAGCCG}}$ (5 °C), (l) $\frac{\text{UCCGAAGCC}}{\text{AGGAGGUGG}}$ (15 °C), (m) $\frac{\text{GAGUAAAUGAC}}{\text{CUCGAAGGCUG}}$, (n) $\frac{\text{GAGUGAAUGAC}}{\text{CUCAAGAACUG}}$ (15 °C), and (o) $\frac{\text{GAGCGAGCGAC}}{\text{CUCGAAGGCUG}}$ (15 °C) in 80 mM NaCl, 3 mM NaH_2PO_4, 7 mM Na_2HPO_4, and 0.5 mM Na_2EDTA (pH 7). Numbers on spectra correspond to assignments with the numbering starting at the left-most (5′) nucleotide of the top strand and ending at the left-most (3′) nucleotide of the bottom strand. The spectrum for $\frac{\text{CCAGCCAAGUCCU}}{\text{GGUUGACGUAGGA}}$ is in the Supporting Information of ref *72*.

As RNA chains increase in length, it is more likely that they can form multiple secondary structures. Thus, it is difficult to design a series of sequences that form designed internal loops (*42*, *43*, *72*, *78*). For the sequences studied here, the concentration-dependent melting temperatures confirm the formation of duplexes instead of unimolecular structures, such as triloop hairpins, which would have concentration-independent melting temperatures. For several non-self-complementary sequences, imino proton NMR spectra also confirm that the designed structures are forming.

The results in Table 2 show that the stability of a 3 × 3 internal loop closed by GC or CG pairs can range from 2.8 to −1.8 kcal/mol at 37 °C, corresponding to a roughly 2000-fold range for the equilibrium constant for folding. Adequate modeling of the sequence dependence of loop stability is

important for predicting RNA secondary structure on the basis of thermodynamics. A reasonable goal is to predict the stability of most 3 × 3 loops within 0.8 kcal/mol, which is 0.1 kcal/mol of difference per nucleotide since half of the closing base pairs are included in the loop stability in the nearest neighbor model (Figure 1). The thermodynamic data and imino proton NMR spectra provide some clues for the thermodynamic modeling of 3 × 3 internal loops. From the comparison of measured and predicted free energy increments for 3 × 3 internal loops (Table 2 and Figure 2), the relatively simple model of eq 4 is usually sufficient. The terms in eq 4 are discussed below. It is evident that there are additional elements of stability remaining to be discovered, however.

Loop Initiation Free Energy. Previous studies have shown that internal loop formation generally becomes less favorable as loop size increases (*38*, *56*, *65*, *79*). The initiation free energy of 1.99 ± 0.11 kcal/mol for 3 × 3 internal loops is approximately 1 kcal/mol less favorable than that for 2 × 2 internal loops (*24*) and agrees well with that used in the current RNAstructure 4.0 algorithm (1.9 ± 0.1 kcal/mol) (*24*). As shown in Table 2, sequences without stabilizing loop-terminal pairs, i.e., without GA, AG, and UU pairs, are predicted well by this loop initiation value. Note that a potential AU middle pair in $\frac{\text{CGCAUAGGC}}{\text{GCGAAACCG}}$ (2.37 kcal/mol), a GU middle pair in $\frac{\text{GAGCAGACGAC}}{\text{CUCGAUAGCUG}}$ (1.94 kcal/mol) and $\frac{\text{GAGCCGACGAC}}{\text{CUCGAUAGCUG}}$ (1.43 kcal/mol), and a GG middle pair in $\frac{\text{GAGCAGACGAC}}{\text{CUCGAGAGCUG}}$ (2.29 kcal/mol) and $\frac{\text{GAGCCGACGAC}}{\text{CUCGAGAGCUG}}$ (1.52 kcal/mol) have no clear stabilizing effect in the corresponding loops.

Bonus for GA and AG Loop-Terminal Pairs. To explore the effect of the loop-terminal noncanonical pair on stability, sequences were studied with GA, AG, and AA pairs flanking a middle AA pair. The order of thermodynamic stability is as follows: GA > AG > AA. From all the data fitted from Table 2, free energy increments of −0.89 ± 0.09 and −0.37 ± 0.14 kcal/mol were derived for GA and AG, respectively (Table 3). There is no clear dependence on the orientation of GC closing base pairs, e.g., $\frac{\text{GCGGAACGC}}{\text{CGCAAGGCG}}$ (0.49 kcal/mol) versus $\frac{\text{CGCGAAGGC}}{\text{GCGAAGCCG}}$ (0.15 kcal/mol) and $\frac{\text{GAGCGAACGAC}}{\text{CUCGAAGGCUG}}$ (−0.07 kcal/mol) as well as $\frac{\text{CGGAAGCGC}}{\text{GCCGAAGCG}}$ (0.87 kcal/mol) versus $\frac{\text{CGCAAGGCG}}{\text{GCGGAACGC}}$ (0.71 kcal/mol).

The imino proton spectra provide clues about the structural basis of the free energy bonuses for loop-terminal GA and AG pairs. As shown in Figure 3g−l,n,o, duplexes with the $\frac{\text{CG}}{\text{GA}}$, $\frac{\text{UG}}{\text{AA}}$ or $\frac{\text{UG}}{\text{GA}}$ motif always have a resonance between 10 and 11 ppm, which is expected for a sheared GA pair (N7−amino, amino−N3; trans Hoogsteen/Sugar edge) (*67*, *68*). Such a resonance is not observed, however, for the $\frac{\text{GCGGAACGC}}{\text{CGCAAGGCG}}$ (Figure 3b) and $\frac{\text{CGGAAGCGC}}{\text{GCGGAAGCG}}$ (Figure 3c) duplexes, which have $\frac{\text{GG}}{\text{CA}}$ and $\frac{\text{GA}}{\text{CG}}$ motifs, respectively. Due to overlap of imino proton regions between imino GA and Watson−Crick GC pairs, no specific assignment was made for these sequences. Presumably, both motifs result in imino GA pairs (N1−N1, carbonyl−amino; cis Watson−Crick/Watson−Crick), as observed for these motifs in 2 × 2 internal loops (*69*, *70*). The $\frac{\text{GA}}{\text{CG}}$ motif does not allow formation of a sheared GA pair because there is a Watson−Crick pair 5′ to the A (*80*).

Bonus for UU Loop-Terminal Pairs. As shown in Table 2, pyrimidine rich internal loops with UU loop-terminal pairs are less sequence dependent than purine rich loops. The range of free energies for loops with two UU loop-terminal pairs is 1.64−0.69 kcal/mol with different CC, UC, GA, and UU middle pairs. The bonus of −0.49 ± 0.11 kcal/mol for each loop-terminal UU pair gives good predictions. There are four peaks in the imino proton NMR spectrum for the self-complementary duplex, $\frac{\text{GCGUUUCGC}}{\text{CGCUUUGCG}}$ (Figure 3d); presumably, two G and two U imino protons are overlapped at 13.2 and 11.4 ppm, respectively, on the basis of the integration of the peaks. Similar chemical shifts of two loop-terminal U imino protons in the same 3 × 3 loop with a longer stem have been observed previously (*81*). Shown in parts e and f of Figure 3 are spectra of pyrimidine rich loops within non-self-complementary duplexes. Observation of imino proton resonances between 10.0 and 11.5 ppm is consistent with hydrogen-bonded UU pairs (*14*, *82*).

Penalty for AU or UA Closing Base Pairs. The three duplexes studied here with loops closed by AU or UA, $\frac{\text{GCAGAAUGC}}{\text{CGUAAGACG}}$, $\frac{\text{GCUGAAAGC}}{\text{CGAAAGUCG}}$, and $\frac{\text{GAGUGAAUGAC}}{\text{CUCAAGAACUG}}$, have $\Delta G^\circ_{37,\text{loop}}$ values of 1.66, 1.38, and 2.30 kcal/mol, respectively. Equation 4 gives predictions of 1.89, 1.89, and 1.56 kcal/mol, respectively. The AU penalty currently employed by *mfold* (*23*) and the most recent RNAstructure 4.0 program (*24*) is 0.7 kcal/mol, which is within experimental error of the fitted AU penalty of 0.84 ± 0.15 kcal/mol (Table 3).

Effect of the Middle Pair. Most of the internal loops studied with a middle GA pair flanked by at least one non-pyrimidine-pyrimidine loop-terminal pair are more stable than would be expected from the model in the current *mfold* and RNAstructure 4.0 programs (eq 3). The two exceptions are $\frac{\text{GAGUGAAUGAC}}{\text{CUCAAGAACUG}}$ (2.30 kcal/mol) and $\frac{\text{GAGCGAGCGAC}}{\text{CUCGAGAGCUG}}$ (0.54 kcal/mol). No extra stabilization is found for other middle pairs, however. This includes the potential Watson−Crick AU pair in $\frac{\text{CGAAG}}{\text{GAUGC}}$, which does not form on the basis of the imino proton spectrum (Figure 3i). Thus, a bonus parameter for a GA middle pair flanked by at least one non-pyrimidine-pyrimidine loop-terminal pair is included in eq 4. The bonus value of −1.22 kcal/mol for a middle GA pair is similar to that for loop-terminal GA pairs (−0.89 kcal/mol), presumably reflecting similar stacking and hydrogen bonding. A GA middle pair flanked by two pyrimidine-pyrimidine loop-terminal pairs has no extra stabilization besides the contribution of UU loop-terminal pairs, e.g., $\frac{\text{GAGCUGCCGAC}}{\text{CUCGUAUGCUG}}$ (1.44 kcal/mol) and $\frac{\text{GAGCUGUCGAC}}{\text{CUCGUAUGCUG}}$ (0.99 kcal/mol).

Thermodynamic stabilization due to a GA pair flanked by two noncanonical pairs may not be restricted to 3 × 3 internal loops. For example, a GA middle pair is well-conserved in the "noncanonical stem" of the kink−turn motif (*83*, *84*). Biochemical and biophysical studies of this motif suggest that the noncanonical stem is structured even in the absence of Mg^{2+} and protein (*85*).

Loops with a single GA loop-terminal pair not adjacent to a middle GA pair are more complicated. A penalty term $\Delta G^\circ_{5'\text{GU}/3'\text{AN}}$ (0.95 kcal/mol) was added for loops with a single GA loop-terminal pair with a U 3′ to the G of the GA pair, e.g., $\frac{\text{CGCAUAGGC}}{\text{GCGAUGCCG}}$ (2.46 kcal/mol), $\frac{\text{CGCAGAGGC}}{\text{GCGAUGCCG}}$ (2.25 kcal/mol), $\frac{\text{GAGCGUACGAC}}{\text{CUCGAUAGCUG}}$ (2.15 kcal/mol), $\frac{\text{CGCAAAGGC}}{\text{GCGAUGCCG}}$ (1.98 kcal/mol), $\frac{\text{GAGCAGACGAC}}{\text{CUCGAUGGCUG}}$ (1.90 kcal/mol), and $\frac{\text{GAGCAAACGAC}}{\text{CUCGAUGGCUG}}$ (1.57 kcal/mol). All these sequences contain a 5′GUA3′

sequence that may have more favorable stacking in the single strand than in the duplex. It is also possible, however, that stacking of a U 3′ to a G of a GA pair is unfavorable unless there is severe distortion of the backbone. Interestingly, relatively less well-defined local structure was observed in G36•A26 and U37•C25 base pairs in an NMR structure of the loop B domain from a hairpin ribozyme (*86*). Dynamics were also found for U21, which is 3′ to the G of a GA pair in loop A from a hairpin ribozyme (*87*). In asymmetric loop E, the G 5′ to U was found to bulge out forming an S turn rather than forming one of two possible sheared GA pairs (*88*, *89*). A similar choice is found in asymmetric internal loops in domains IIb and IIId of the HCV virus (*13*, *90*). A 3 × 3 loop, $^{\mathrm{CGUAG}}_{\mathrm{GAAAC}}$, in helix 38 of 23S rRNA interacts with 5S rRNA, and the loop-terminal G (shown in bold) is found to form a pair with the U to make a base triple rather than making a GA pair (*91*). Note that a $\Delta G^{\circ}_{5'\mathrm{GU}/3'\mathrm{AN}}$ penalty is not applied for loops with two GA loop-terminal pairs, e.g., $^{\mathrm{GCGGAACGG}}_{\mathrm{CGCAUGGCC}}$ (0.70 kcal/mol), $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGAUGCCG}}$ (0.57 kcal/mol), and $^{\mathrm{GCGGUACGG}}_{\mathrm{CGCAUGGCC}}$ (0.29 kcal/mol). The stabilizing effect of two loop-terminal GA pairs agrees with the formation of two sheared GA pairs in the internal loop, $^{\mathrm{UGUAC}}_{\mathrm{GAUGG}}$, in helix 96 of 23S rRNA from *Haloarcula marismortui* (*83*). The other loops in Table 2 with a single GA loop-terminal pair are predicted reasonably well to be 1.10 kcal/mol: $^{\mathrm{GAGCGGACGAC}}_{\mathrm{CUCGAUAGCUG}}$ (1.77 kcal/mol), $^{\mathrm{CGACGCAGCAG}}_{\mathrm{GCUGAAACGUC}}$ (1.50 kcal/mol), $^{\mathrm{GAGCGAACGAC}}_{\mathrm{CUCGAAAGCUG}}$ (1.39 kcal/mol), $^{\mathrm{GAGCGAACGAC}}_{\mathrm{CUCGAUAGCUG}}$ (1.33 kcal/mol), $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGACCCCG}}$ (1.00 kcal/mol), $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGAAACCG}}$ (0.89 kcal/mol), $^{\mathrm{CGCAUAGGC}}_{\mathrm{GCGAAGCCG}}$ (0.51 kcal/mol), $^{\mathrm{CGCAAAGGC}}_{\mathrm{GCGAAGCCG}}$ (0.26 kcal/mol), and $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGAACCCG}}$ (−0.13 kcal/mol).

GU Pairs Closing Loops Are Idiosyncratic. The thermodynamic parameters in folding algorithms typically assume that GU pairs closing internal loops are equivalent to AU pairs. A comparison of measured and predicted values in Table 2 suggests that this is not a good approximation. The most surprising loop is $^{\mathrm{CGAAG}}_{\mathrm{GAGGU}}$ (−3.42 kcal/mol) which is more stable than the $^{\mathrm{CGAAG}}_{\mathrm{GAGGC}}$ loop (−1.73 kcal/mol) even though it has a GU closing pair. The imino proton spectrum (Figure 3l) shows three relatively sharp imino proton peaks from the loop, suggesting a structured loop that protects the imino protons from exchange with water. The $^{\mathrm{UG}}_{\mathrm{GA}}$ motif is thermodynamically relatively stable in 2 × 2 loops when compared with the $^{\mathrm{UG}}_{\mathrm{AA}}$ motif but not when compared with the $^{\mathrm{CG}}_{\mathrm{GA}}$ motif (*72*). Thus, the $^{\mathrm{CGAAG}}_{\mathrm{GAGGU}}$ loop is unusually stable. This loop is found in several SSU rRNA and SRP RNA secondary structures (*17*, *18*). Data for 2 × 2 loops closed with at least one GU pair are also not predicted well assuming GU is thermodynamically equal to AU (*72*). Evidently, stacking interactions with GU pairs differ from those with AU pairs.

Nearest Neighbor Approximation. One important assumption of the nearest neighbor free energy minimization method for predicting RNA secondary structure is pairwise additivity. One test of the model is the free energy comparison between identical loops with different stems and/or orientations: $^{\mathrm{GCGAAACGC}}_{\mathrm{CGCAAAGCG}}$ (2.76 kcal/mol) versus $^{\mathrm{CGGAAACGC}}_{\mathrm{GCCAAAGCG}}$ (2.47 kcal/mol) and $^{\mathrm{CGGAAACCG}}_{\mathrm{GCCAAAGGC}}$ (1.84 kcal/mol); $^{\mathrm{GGCGAAGCC}}_{\mathrm{CCGAAGCGG}}$ (1.25 kcal/mol) versus $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGAAGCCG}}$ (0.15 kcal/mol); $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGAAACCG}}$ (0.89 kcal/mol) versus $^{\mathrm{CGCAAAGGC}}_{\mathrm{GCGAAGCCG}}$ (0.26 kcal/mol); and $^{\mathrm{CGCUCUGGC}}_{\mathrm{GCGUUUCCG}}$ (0.78 kcal/mol) versus $^{\mathrm{CGCUUUGGC}}_{\mathrm{GCGUCUCCG}}$ (0.72 kcal/mol). The comparisons indicate that the nearest neighbor model is usually a reasonable approximation.

ACKNOWLEDGMENT

We thank Prof. D. H. Mathews for help with linear regression, Prof. R. Kierzek for suggesting the method of 2′-hydroxyl deprotection, and Dr. S. D. Kennedy for NMR technical support. G.C. thanks Drs. M. D. Disney and J. L. Childs for the first UV melting experiment.

SUPPORTING INFORMATION AVAILABLE

Tables of single-strand melting results and number of occurrences of noncanonical base pairs and closing base pairs in 3 × 3 loops in known secondary structures and a figure of an SNOESY two-dimensional spectrum of $^{\mathrm{CGCGAAGGC}}_{\mathrm{GCGAGGCCG}}$• This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hanczyc, M. M., Fujikawa, S. M., and Szostak, J. W. (2003) Experimental models of primitive cellular compartments: Encapsulation, growth, and division, *Science 302*, 618−622.
2. Joyce, G. F. (2002) The antiquity of RNA-based evolution, *Nature 418*, 214−221.
3. Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds. (1999) *The RNA World*, 2nd ed., Cold Spring Harbor Laboratory Press, Plainview, NY.
4. 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.
5. Childs, J. L., Disney, M. D., and Turner, D. H. (2002) Oligonucleotide directed misfolding of RNA inhibits *Candida albicans* group I intron splicing, *Proc. Natl. Acad. Sci. U.S.A. 99*, 11091−11096.
6. Disney, M. D., Childs, J. L., and Turner, D. H. (2004) New approaches to targeting RNA with oligonucleotides: Inhibition of group I intron self-splicing, *Biopolymers 73*, 151−161.
7. Gallego, J., and Varani, G. (2001) Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges, *Acc. Chem. Res. 34*, 836−843.
8. Tor, Y. (2003) Targeting RNA with small molecules, *ChemBioChem 4*, 998−1007.
9. Wilson, W. D., and Li, K. (2000) Targeting RNA with small molecules, *Curr. Med. Chem. 7*, 73−98.
10. Leontis, N. B., and Westhof, E. (2003) Analysis of RNA motifs, *Curr. Opin. Struct. Biol. 13*, 300−308.
11. Al-Hashimi, H. M., Gorin, A., Majumdar, A., Gosser, Y., and Patel, D. J. (2002) Towards structural genomics of RNA: Rapid NMR resonance assignment and simultaneous RNA tertiary structure determination using residual dipolar couplings, *J. Mol. Biol. 318*, 637−649.
12. Bondensgaard, K., Mollova, E. T., and Pardi, A. (2002) The global conformation of the hammerhead ribozyme determined using residual dipolar couplings, *Biochemistry 41*, 11532−11542.
13. 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.
14. 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.
15. 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.
16. Brown, J. W. (1999) The Ribonuclease P Database, *Nucleic Acids Res. 27*, 314.
17. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R.

(2002) The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs, *BMC Bioinf. 3*, 2.
18. Rosenblad, M. A., Gorodkin, J., Knudsen, B., Zwieb, C., and Samuelsson, T. (2003) SRPDB: Signal Recognition Particle Database, *Nucleic Acids Res. 31*, 363−364.
19. Pace, N. R., Thomas, B. C., and Woese, C. R. (1999) Probing RNA structure, function, and history by comparative analysis, in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 113−141, Cold Spring Harbor Laboratory Press, Plainview, NY.
20. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) RNA structure prediction, *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
21. 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, CA.
22. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) The interactions that shape RNA structure, in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 233−264, Cold Spring Harbor Laboratory Press, Plainview, NY.
23. 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.
24. 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.
25. Mathews, D. H., and Turner, D. H. (2002) Dynalign: An algorithm for finding the secondary structure common to two RNA sequences, *J. Mol. Biol. 317*, 191−203.
26. Chen, J. H., Le, S. Y., and Maizel, J. V. (2000) Prediction of common secondary structures of RNAs: a genetic algorithm approach, *Nucleic Acids Res. 28*, 991−999.
27. Jaeger, L., Westhof, E., and Leontis, N. B. (2001) TectoRNA: modular assembly units for the construction of RNA nano-objects, *Nucleic Acids Res. 29*, 455−463.
28. Seeman, N. C. (2003) Biochemistry and structural DNA nanotechnology: An evolving symbiotic relationship, *Biochemistry 42*, 7259−7269.
29. Breaker, R. R. (2002) Engineered allosteric ribozymes as biosensor components, *Curr. Opin. Biotechnol. 13*, 31−39.
30. Robertson, M. P., and Ellington, A. D. (1999) In vitro selection of an allosteric ribozyme that transduces analytes to amplicons, *Nat. Biotechnol. 17*, 62−66.
31. Komatsu, Y., Yamashita, S., Kazama, N., Nobuoka, K., and Ohtsuka, E. (2000) Construction of new ribozymes requiring short regulator oligonucleotides as a cofactor, *J. Mol. Biol. 299*, 1231−1243.
32. Vaish, N. K., Jadhav, V. R., Kossen, K., Pasko, C., Andrews, L. E., McSwiggen, J. A., Polisky, B., and Seiwert, S. D. (2003) Zeptomole detection of a viral nucleic acid using a target-activated ribozyme, *RNA 9*, 1058−1072.
33. Du, H., Disney, M. D., Miller, B. L., and Krauss, T. D. (2003) Hybridization-based unquenching of DNA hairpins on Au surfaces: Prototypical "molecular beacon" biosensors, *J. Am. Chem. Soc. 125*, 4012−4013.
34. Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003) Asymmetry in the assembly of the RNAi enzyme complex, *Cell 115*, 199−208.
35. Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003) Functional siRNAs and miRNAs exhibit strand bias, *Cell 115*, 209−216.
36. Silva, J. M., Sachidanandam, R., and Hannon, G. J. (2003) Free energy lights the path toward more effective RNAi, *Nat. Genet. 35*, 303−305.
37. 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, U.K.
38. Schroeder, S. J., Burkard, M. E., and Turner, D. H. (1999) The energetics of small internal loops in RNA, *Biopolymers 52*, 157−167.
39. Puglisi, J. D., Tan, R. Y., Calnan, B. J., Frankel, A. D., and Williamson, J. R. (1992) Conformation of the TAR RNA-arginine complex by NMR spectroscopy, *Science 257*, 76−80.
40. 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.
41. 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.
42. 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.
43. 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.
44. McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of $(rGAG\underline{GU}CUC)_2$ by two-dimensional NMR and simulated annealing, *Biochemistry 35*, 14077−14089.
45. 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 $r(GGC\underline{GU}GCC)_2$ and two for each G·U pair in $r(GA\underline{GU}\underline{GU}CUC)_2$, *Biochemistry 39*, 8970−8982.
46. 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.
47. Leontis, N. B., and Westhof, E. (1998) The 5S rRNA loop E: Chemical probing and phylogenetic data versus crystal structure, *RNA 4*, 1134−1153.
48. Vallurupalli, P., and Moore, P. B. (2003) The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts, *J. Mol. Biol. 325*, 843−856.
49. Cieplak, P., Caldwell, J., and Kollman, P. (2001) Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and *N*-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases, *J. Comput. Chem. 22*, 1048−1057.
50. Sponer, J., Leszczynski, J., and Hobza, P. (1996) Nature of nucleic acid−base stacking: Nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs, *J. Phys. Chem. 100*, 5590−5596.
51. 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.
52. 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.
53. Stawinski, J., Stromberg, R., Thelin, M., and Westman, E. (1988) Evaluation of the use of the *tert*-butyldimethylsilyl group for 2′-protection in RNA: Synthesis via the H-phosphonate approach, *Nucleosides Nucleotides 7*, 779−782.
54. Pirrung, M. C., Shuey, S. W., Lever, D. C., and Fallon, L. (1994) A convenient procedure for the deprotection of silylated nucleosides and nucleotides using triethylamine trihydrofluoride, *Bioorg. Med. Chem. Lett. 4*, 1345−1346.
55. Borer, P. N. (1975), in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) 3rd ed., pp 597, CRC Press, Cleveland, OH.
56. 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.
57. 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.
58. 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.
59. 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.
60. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry 29*, 278−285.
61. Damberger, S. H., and Gutell, R. R. (1994) A comparative database of group I intron structures, *Nucleic Acids Res. 22*, 3508−3510.
62. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) A compilation of large subunit (23S- and 23S-like) ribosomal RNA structures: 1993, *Nucleic Acids Res. 21*, 3055−3074.
63. Gutell, R. R. (1994) Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994, *Nucleic Acids Res. 22*, 3502−3507.
64. SantaLucia, J., Jr., and Turner, D. H. (1997) Measuring the thermodynamics of RNA secondary structure formation, *Biopolymers 44*, 309−319.
65. 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.
66. Varani, G., Aboulela, F., and Allain, F. H. T. (1996) NMR investigation of RNA structure, *Prog. Nucl. Magn. Reson. Spectrosc. 29*, 51−127.
67. Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops, *Science 253*, 191−194.
68. SantaLucia, J., Jr., and Turner, D. H. (1993) Structure of $(rGGCGAGCC)_2$ in solution from NMR and restrained molecular dynamics, *Biochemistry 32*, 12612−12623.
69. Wu, M., and Turner, D. H. (1996) Solution structure of $(rGCGGACGC)_2$ by two-dimensional NMR and the iterative relaxation matrix approach, *Biochemistry 35*, 9677−9689.
70. Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) Solution structure of $(rGGCAGGCC)_2$ by two-dimensional NMR and the iterative relaxation matrix approach, *Biochemistry 36*, 4449−4460.
71. 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.
72. Schroeder, S. J., and Turner, D. H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA, *Biochemistry 40*, 11509−11517.
73. 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.
74. 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, *Nat. New Biol. 246*, 40−41.
75. Rivas, E., and Eddy, S. R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots, *J. Mol. Biol. 285*, 2053−2068.
76. 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.
77. Ding, Y., and Lawrence, C. E. (2003) A statistical sampling algorithm for RNA secondary structure prediction, *Nucleic Acids Res. 31*, 7280−7301.
78. Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA, *Biochemistry 39*, 9257−9274.
79. Weeks, K. M., and Crothers, D. M. (1993) Major groove accessibility of RNA, *Science 261*, 1574−1577.
80. 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.
81. Sich, C., Ohlenschlager, O., Ramachandran, R., Gorlach, M., and Brown, L. R. (1997) Structure of an RNA hairpin loop with a 5′-CGUUUCG-3′ loop motif by heteronuclear NMR spectroscopy and distance geometry, *Biochemistry 36*, 13989−14002.
82. 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.
83. 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.
84. 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.
85. 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.
86. 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.
87. Cai, Z., and Tinoco, I. (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite, *Biochemistry 35*, 6026−6036.
88. Wimberly, B., Varani, G., and Tinoco, I. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA, *Biochemistry 32*, 1078−1087.
89. 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.
90. Lukavsky, P. J., Otto, G. A., Lancaster, A. M., Sarnow, P., and Puglisi, J. D. (2000) Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function, *Nat. Struct. Biol. 7*, 1105−1110.
91. 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.

BI049168D