Experimentally Derived Nearest-Neighbor Parameters for the Stability of RNA Three- and Four-Way Multibranch Loops†

David H. Mathews and Douglas H. Turner*

Department of Chemistry, University of Rochester, Rochester, New York 14627-0216

Received July 10, 2001

ABSTRACT: Algorithms for predicting RNA secondary structure require approximations for the free energies of multibranch loops, also called junctions. The stabilities of 62 RNA duplexes with three- and four-way multibranch loops were determined by optical melting. To account for the observed sequence dependence, a revised loop free-energy approximation is proposed that accounts for the strain in three-way junctions with fewer than two unpaired nucleotides, penalizes asymmetry in the distribution of unpaired nucleotides, and gives a bonus for four-way loops relative to three-way loops. Parameters for this equation were determined by linear regression.

Multibranch loops, also called junctions, are nucleic acid loops from which three or more helices exit. In RNA, these loops play a central role in most known secondary structures. For example, the 5S rRNA and the cloverleaf tRNA structures are built around central three-way and four-way multibranch loops, respectively (*1*, *2*). Larger secondary structures, such as the small and large rRNA subunits, contain many multibranch loops with various numbers of branching helices of up to 12 (*3*, *4*). Native tertiary structure formation of the hairpin ribozyme is favored by the natural four-way multibranch loop when compared to loops with either two or three branches (*5*). Comparative sequence analysis of homologous RNAs has shown that conserved multibranch loops often have a characteristic pattern in the number and placement of nucleotides in the loop that are not involved in canonical pairs. For example, the 5S rRNA most often has four nucleotides between helix 1 and helix 2 (*2*). A large and relatively conserved asymmetry in the location of nucleotides is also found in tRNA (*1*). Structures determined by X-ray crystallography demonstrate that these nucleotides often participate in noncanonical pairs that stabilize these multibranch loops in unique conformations.

Free-energy minimization algorithms that predict secondary structure typically use nearest-neighbor parameters to assign free energies (*6*−*10*). For multibranch loops, the free energy is often approximated by two contributions (*6*)

$$\Delta G^\circ_{\mathrm{MBL}} = \Delta G^\circ_{\mathrm{MBL\ init}} + \Delta G^\circ_{\mathrm{MBL\ stacking}} \qquad (1)$$

where $\Delta G^\circ_{\mathrm{MBL\ stacking}}$ is the optimal sum of free energies for the stacking of coaxial helices, for noncanonical pairs at helix termini, and for dangling unpaired nucleotides at helix termini. $\Delta G^\circ_{\mathrm{MBL\ init}}$ is a penalty term for closing the loop. It is often assumed to be of the form (*6*, *11*, *12*)

$$\Delta G^\circ_{\mathrm{MBL\ init}} = a + bn + ch \qquad (2)$$

where n is the number of unpaired nucleotides, h is the number of branching helices, and $a-c$ are parameters that were found by maximizing the accuracy of secondary structure prediction (*6*, *11*). In the absence of experimental data, the linear form of $\Delta G^\circ_{\mathrm{MBL\ init}}$ was chosen as a computationally efficient formula for dynamic programming algorithms (*11*, *13*, *14*).

Diamond et al. determined the stabilities of 12 three-way multibranch loops by optical melting (*15*), establishing an approximate value for $\Delta G^\circ_{\mathrm{MBL\ init}}$ and demonstrating the feasibility of using optical melting to determine multibranch loop stability. The approximate $\Delta G^\circ_{\mathrm{MBL\ init}}$ for three-way multibranch loops was used as a starting point for the optimization of $a-c$ in eq 2 (*6*).

In this study, thermodynamic parameters for 62 three- and four-way multibranch loops were determined by optical melting. These loops vary in the number and placement of unpaired adenines in the loop and in the identity of the closing base pairs. The free energies of these loops at 37 °C (ΔG°_{37}) indicate that the asymmetry in the distribution of unpaired nucleotides is a more important determinant of stability than the absolute number of unpaired nucleotides. Furthermore, three-way multibranch loops with either zero or one unpaired nucleotide are less stable than predicted on the basis of eq 2. Therefore, a new equation is proposed that reflects the observed variation in $\Delta G^\circ_{\mathrm{MBL\ init}}$ as a function of the asymmetry in the distribution of unpaired nucleotides and of the strain for multibranch loops with three branching helices and less than two unpaired nucleotides. The parameters for this new model are found by linear regression on the experimental data.

MATERIALS AND METHODS

RNA Synthesis. RNA strands were synthesized by standard phosphoramidite methods on an Applied Biosystems 392 DNA/RNA synthesizer with phosphoramidites from Glen Research (Sterling, VA). Strands were deblocked and purified as previously described (*15*).

The short strands were greater than 95% pure on the basis of HPLC[1] with a C-18 column (Hamilton, Reno, NV) on a

* To whom correspondence should be addressed.

† This work was supported by NIH Grant GM 22939. D.H.M. is a trainee in the Medical Scientist Training Program (NIH Grant 5T32 GM07356).

10.1021/bi011441d CCC: $22.00
Published on Web 12/19/2001

Hewlett−Packard 1100 series HPLC. The buffer system was triethylamine acetate and acetonitrile starting with 0% acetonitrile and increasing the percent acetonitrile at 1%/min for 40 min. The long strands were greater than 95% pure on the basis of polyacrylamide gel electrophoresis of 5′ ^{32}P-labeled strands. The bands were imaged on a Molecular Dynamics Phosphorimager and quantified with ImageQuant NT software (Molecular Dynamics, Cleveland, OH).

Strand concentrations were determined from UV absorbance at 90 °C using nearest-neighbor parameters (*16*, *17*). The average of concentrations determined from absorbances at 260 and 280 nm was used.

Optical Melting. The buffer for optical melting experiments was 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA adjusted to pH 7 with NaOH. Previous experiments showed no difference in stability for three-way multibranch loops melted in this buffer and in a buffer system with 10 mM $MgCl_2$, 20 mM sodium cacodylate, and 150 mM KCl adjusted to pH 7 with KOH (*15*). Melting curves were measured at 280 nm with a Beckman Coulter model DU 640 spectrophotometer with a temperature controller and a six-sample transport accessory. At least five strand concentrations were melted for each system. Melting curves were transferred to a personal computer for analysis by the program Meltwin as described by McDowell and Turner (*18*). Thermodynamic parameters for duplex formation were derived by two methods: the average of individual curve fits and the linear fit of $1/T_M$ versus $\ln(C_T/4)$, where C_T is the total strand concentration.

The concentration range that could be analyzed for each system was limited by two factors. The lowest concentration was specified by the lowest absorbance that could be accurately read by the spectophotometer for a sample in a 1 cm path length cell. The highest concentration was limited by the ability to adequately separate the low- and high-temperature transitions when fitting the data. As strand concentration increases, the low-temperature (bimolecular) transition is shifted to higher temperatures, but the higher-temperature (unimolecular) transition is not shifted.

Melting curves were obtained for each strand by itself to check for any self-complementarity. By design, each long strand had only a high-temperature transition associated with the melting of the stem−loop regions. These transitions start at about 65 °C and continue above 90 °C. Each short strand had no optical melting transition in the absence of a long strand.

Change in Heat Capacity. The difference in heat capacity between the duplex and single strands, $\Delta C_p°$, was calculated by linear fits of enthalpy and entropy changes as a function of melting temperature (*19*). The temperature dependences of $\Delta H°$ and $\Delta S°$ are given by

$$d\Delta H°/dT = \Delta C_p° \quad (3)$$

$$d\Delta S°/d(\ln T) = \Delta C_p° \quad (4)$$

where ln indicates the natural logarithm. If one assumes that the enthalpy and entropy derived from each curve fit are the values at T_M, then $\Delta C_p°$ can be derived by a linear fit of the following (*19*):

$$\Delta H°(T_M) = \Delta H°(T_0) + \Delta C_p°(T_M - T_0) \quad (5)$$

$$\Delta S°(T_M) = \Delta S°(T_0) + \Delta C_p° \ln(T_M/T_0) \quad (6)$$

where T_0 is a chosen reference temperature.

Linear Regression. Microsoft Excel, version 97, was used for the linear regression analysis of the database.

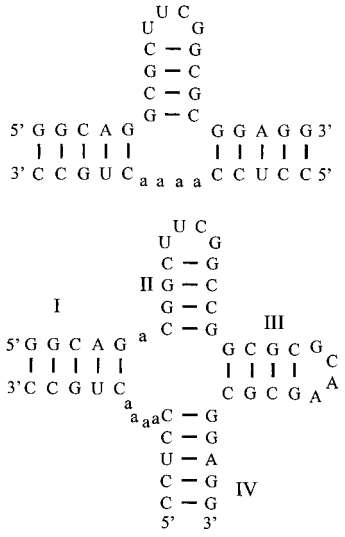

FIGURE 1: Sequence Design. (Top) The design of the three-way multibranch loops is shown. (Bottom) The design of the four-way multibranch loops is shown. The three-way loop is the system G_GC_G/Ca$_4$C, and the four-way loop is GaCG_GC_G/Ca$_4$C (see Table 1). The helices in the four-way junction are labeled with Roman numerals.

RESULTS

Optical Melting Experiments. Figure 1 and Tables 1 and 2 demonstrate the design features of the duplexes forming multibranch loops. Each system is composed of two strands so that the denaturation of the multibranch loop is bimolecular, allowing for quantification of the data with well-established methods (*18*, *19*).

Both the three- and four-way multibranch loop systems include a short strand and a long strand. The long strands incorporate helices terminated with stable tetraloops (*20*, *21*), and the short strands are designed to form base pairs only in association with the long strands. Compared to a previous study of multibranch loop stability (*15*), the bimolecular helices are each extended by one base pair to make the transitions more cooperative. The three- and four-way loops differ in the number of helices incorporated into the long strand. The base-pairing regions that bind the two strands together are invariant, but the number of unpaired nucleotides in the loop and the identity of the terminal base pair in the intramolecular helices are varied. This allows for the formation of a large number of loops by combining a smaller number of long and short sequences. The systems were designed with mostly adenosines in the multibranch loops to minimize the potential for alternative structure formation. The helical regions are composed of predominantly G−C base pairs; therefore, A is the nucleotide least likely to compete with the intended canonical base pairs.Moreover,

[1] Abbreviations: eu, entropic units, cal K^{-1} mol^{-1}; EDTA, ethylenediaminetetraacetic acid; HPLC, high-pressure liquid chromatography; T_m, melting temperature, °C; T_M, melting temperature, K; C_T, total strand concentration.

Table 1: Nomenclature for Three-Way Multibranch Loop Systems Studied

long strand[a]	short strand[a]					
	CCUCC_CUGCC	CCUCCaCUGCC	CCUCCaaCUGCC	CCUCCaaaCUGCC	CCUCCauaCUGCC	CCUCCaaaaCUGCC
GGCAG_GCGCUUCGGCGC_GGAGG	G_GC_G/C_C	G_GC_G/CaC	G_GC_G/Ca_2C	G_GC_G/Ca_3C	G_GC_G/CauaC	G_GC_G/Ca_4C
GGCAGaGCGCUUCGGCGC_GGAGG	GaGC_G/C_C	GaGC_G/CaC	GaGC_G/Ca_2C	GaGC_G/Ca_3C	GaGC_G/CauaC	GaGC_G/Ca_4C
GGCAGaaGCGCUUCGGCGC_GGAGG	Ga_2GC_G/C_C	Ga_2GC_G/CaC	Ga_2GC_G/Ca_2C	Ga_2GC_G/Ca_3C	Ga_2GC_G/CauaC	Ga_2GC_G/Ca_4C
GGCAGaaGCGCUUCGGCGCaGGAGG	Ga_2GCaG/C_C	Ga_2GCaG/CaC	Ga_2GCaG/Ca_2C	Ga_2GCaG/Ca_3C	Ga_2GCaG/CauaC	Ga_2GCaG/Ca_4C
GGCAGaaGCGCUUCGGCGCaaGGAGG	Ga_2GCa_2G/C_C	Ga_2GCa_2G/CaC	Ga_2GCa_2G/Ca_2C	Ga_2GCa_2G/Ca_3C	Ga_2GCa_2G/CauaC	Ga_2GCa_2G/Ca_4C
GGCAG_UCGCUUCGGCGA_GGAGG	G_UA_G/C_C	G_UA_G/CaC	G_UA_G/Ca_2C	G_UA_G/Ca_3C	G_UA_G/CauaC	G_UA_G/Ca_4C
GGCAGaUCGCUUCGGCGA_GGAGG	GaUA_G/C_C	GaUA_G/CaC	GaUA_G/Ca_2C	GaUA_G/Ca_3C	GaUA_G/CauaC	GaUA_G/Ca_4C

[a] Sequences are written in the 5′ to 3′ direction. Nucleotides that are unpaired in the junction are in lowercase. An underscore indicates a helix−helix interface that is without unpaired nucleotides.

Table 2: Nomenclature for Four-Way Multibranch Loop Systems Studied

long strand[a]	short strand[a]					
	CCUCC_CUGCC	CCUCCaCUGCC	CCUCCaaCUGCC	CCUCCaaaCUGCC	CCUCCauaCUGCC	CCUCCaaaaCUGCC
GGCAG_CGGCUUCGGCCG_GCGCGCAAGCGC_GGAGG	G_CG_GC_G/C_C	G_CG_GC_G/CaC	G_CG_GC_G/Ca_2C	G_CG_GC_G/Ca_3C		G_CG_GC_G/Ca_4C
GGCAGaCGGCUUCGGCCG_GCGCGCAAGCGC_GGAGG	GaCG_GC_G/C_C	GaCG_GC_G/CaC	GaCG_GC_G/Ca_2C	GaCG_GC_G/Ca_3C	GaCG_GC_G/CauaC	GaCG_GC_G/Ca_4C
GGCAGaCGGCUUCGGCCG_GCGCGCAAGCGCaGGAGG	GaCG_GCaG/C_C	GaCG_GCaG/CaC	GaCG_GCaG/Ca_2C	GaCG_GCaG/Ca_3C	GaCG_GCaG/CauaC	GaCG_GCaG/Ca_4C
GGCAGaaCGGCUUCGGCCGaa GCGCGCAAGCGCaaGGAGG	Ga_2CGa_2GCa_2G/C_C	Ga_2CGa_2GCa_2G/CaC	Ga_2CGa_2GCa_2G/Ca_2C			

[a] Sequences are written in the 5′ to 3′ direction. Nucleotides that are unpaired in the junction are in lowercase. An underscore indicates a helix−helix interface that is without unpaired nucleotides.

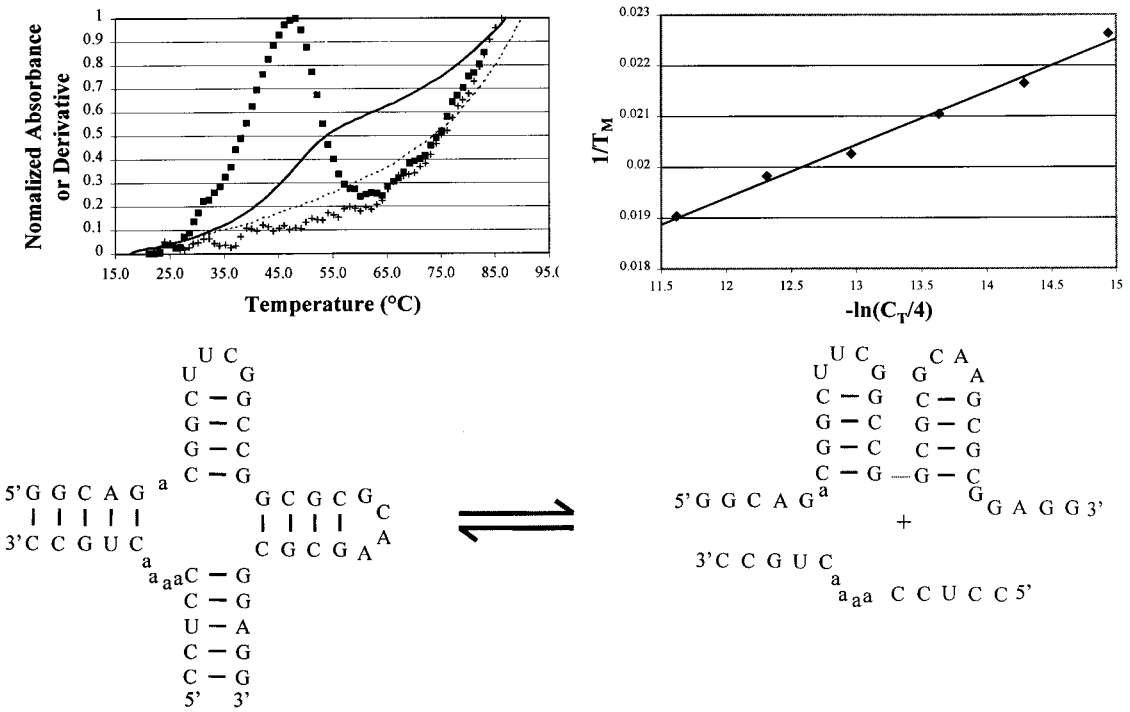

FIGURE 2: Melting of a four-way multibranch loop. (Top left) Optical melting curves for the multibranch loop GaCG_GC_G/Ca$_4$C (solid line) and the strand GaCG_GC_G only (dashed line) are shown, with the absorbances normalized to appear on the same graph. The GaCG_GC_G/Ca$_4$C total strand concentration is 2.5 μM, and the total strand concentration of GaCG_GC_G is 26.0 μM. The normalized derivatives of the multibranch loop and the long strand only are shown with squares and plus symbols, respectively. Derivatives were calculated with a Savitsky−Golay function with a third-order polynomial and nine points (*69*). (Top right) The $1/T_M$ dependence as a function of the natural log of the total strand concentration and its linear fit to $1/T_M = (R/\Delta H°)\ln(C_T/4) + \Delta S°/\Delta H°$. (Bottom) The equilibrium of the lower temperature transition is shown, with the low-temperature species on the left. Note that the thermodynamic parameters reported are for the equilibrium written in the opposite direction.

A is the most common nucleotide not in canonical pairs in small and large rRNAs (*22*).

Tables 1 and 2 show the sequence of each three and four-way system studied, respectively. These tables also provide the nomenclature used to describe the closing pairs and unpaired nucleotides in a loop. For example, the four-way multibranch loop in Figure 1, formed by the association of 5′GGCAGaCGGCUUCGGCCG_GCGCGCAAGCGC_G-GAGG3′ and 5′CCUCCaaaaCUGCC3′, is referred to as GaCG_GC_G/Ca$_4$C. The loop sequence is indicated from 5′ to 3′ with uppercase letters indicating the canonically paired nucleotides that close helices and with lowercase letters indicating unpaired nucleotides in the loop. Helices that have no intervening unpaired nucleotides are separated by an underscore, and a slash delineates the two strands.

A typical melting curve and its derivative with respect to temperature are shown in Figure 2. The two maxima in the derivative indicate that two transitions occur during the melting of the complete system. The lower-temperature transition is the bimolecular transition shown in Figure 2, corresponding to the denaturation of the multibranch loop. The higher-temperature transition is the denaturation of the stable tetraloops in the long strand, as illustrated by the fact that the melting curve of the long strand alone shows the higher-temperature transition only (Figure 2). The bimolecular and unimolecular transitions are separated by a trough in the derivative of the melting curve. Thus, the T_M, $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$ for the bimolecular transition can be determined by fitting the data for temperatures below the trough (*12*, *15*, *23*). Figure 2 shows the resulting linear fit of the T_M^{-1} versus $\ln(C_T/4)$ plot for the system GaCG_G-C_G/Ca$_4$C.

Tables 3 and 4 list the thermodynamic parameters derived from experiments on three- and four-way loops, respectively. In most cases, the difference in $\Delta H°$ as determined by the two fitting methods is less than 15%, a value consistent with the transition being two-state (*19*, *24*). A comparison of multibranch loop thermodynamics measured by optical melting and isothermal titration calorimetry indicated that optical melting provides reasonable values of $\Delta G°_{37}$, even when the $\Delta H°$ values derived from curve fitting and from T_M^{-1} plots differ by more than 15% (*15*). Therefore, systems with enthalpy differences larger than 15% are included in the nearest-neighbor parameter derivation in this work.

Determination of Loop Stability and Loop Initiation. In a nearest-neighbor model, the bimolecular free energy, $\Delta G°_{bimol}$, for the formation of the duplex with a multibranch loop is

$$\Delta G°_{bimol} = \Delta G°_{helix\ 1} + \Delta G°_{helix\ 2} + \Delta G°_{bimol\ init} + \Delta G°_{MBL} - \Delta G°_{product\ mm} \quad (7)$$

where helix 1 and helix 2 are the intermolecular paired helices with $\Delta G°$ predicted from nearest-neighbor parameters as determined by Xia et al. for Watson−Crick pairs (*25*). The helical free-energy terms are calculated without including bimolecular initiation so that bimolecular initiation, $\Delta G°_{bimol\ init}$, appears only once in eq 7. The $\Delta G°_{product\ mm}$ is a term that accounts for the stacking free-energy increment of the nucleotides that can stack on the hairpin loop stems to form a modified motif after the two strands have dissociated

Table 3: Thermodynamic Data from Optical Melting of Three-Way Multibranch Loops

	analysis of T_m dependence				average of curve fits						
system	$-\Delta H°$ (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)	$\Delta H°$ % diff	$\Delta G°_{37}$ % diff	conc range
G_GC_G/C_C	71.22 ± 5.49	202.6 ± 17.5	8.40 ± 0.12	45.4	65.33 ± 5.88	183.6 ± 18.7	8.38 ± 0.14	46.0	8.6	0.2	57
G_GC_G/CaC	85.19 ± 4.40	241.1 ± 13.8	10.40 ± 0.13	51.8	80.49 ± 2.86	226.4 ± 9.1	10.29 ± 0.09	52.2	5.7	1.1	28
G_GC_G/Ca_2C	86.44 ± 8.56	241.6 ± 26.5	11.50 ± 0.37	55.9	87.78 ± 6.62	245.7 ± 20.6	11.59 ± 0.26	56.0	1.5	0.8	14
G_GC_G/Ca_3C	76.86 ± 5.08	210.8 ± 15.7	11.49 ± 0.23	58.4	91.30 ± 6.77	255.2 ± 20.5	12.15 ± 0.42	57.3	17.2	5.6	14
G_GC_G/CauaC	103.1 ± 8.1	290.9 ± 24.9	12.86 ± 0.40	57.3	112.3 ± 5.5	319.3 ± 16.7	13.31± 0.37	56.9	8.6	3.4	27
G_GC_G/Ca_4C	99.65 ± 12.67	281.1 ± 39.1	12.47 ± 0.61	56.7	94.24 ± 5.01	264.3 ± 15.6	12.26 ± 0.19	57.1	5.6	1.7	15
GaGC_G/C_C	52.19 ± 2.42	139.9 ± 7.7	8.81 ± 0.04	51.2	66.30 ± 7.85	184.6 ± 24.3	9.06 ± 0.33	49.3	23.8	2.8	14
GaGC_G/CaC	72.29 ± 2.85	199.1 ± 8.9	10.53 ± 0.10	55.2	78.52 ± 5.34	218.4 ± 16.5	10.77 ± 0.24	54.7	8.3	2.3	14
GaGC_G/Ca_2C	79.64 ± 2.47	220.9 ± 7.6	11.14 ± 0.10	56.1	88.10 ± 9.88	247.0 ± 30.3	11.50 ± 0.47	55.6	10.1	3.2	27
GaGC_G/Ca_3C	71.94 ± 1.20	197.5 ± 3.7	10.70 ± 0.04	56.1	87.08 ± 9.08	244.4 ± 27.8	11.27 ± 0.46	54.9	19.0	5.2	15
GaGC_G/CauaC	88.95 ± 6.66	250.2 ± 20.6	11.35 ± 0.27	54.8	101.2 ± 11.8	287.8 ± 36.3	11.88 ± 0.54	54.3	12.8	4.6	10
GaGC_G/Ca_4C	79.71 ± 6.01	221.6 ± 18.7	10.97 ± 0.24	55.3	87.56 ± 6.11	245.9 ± 18.7	11.29 ± 0.31	54.8	9.4	2.9	28
Ga_2GC_G/C_C	57.95 ± 1.57	155.5 ± 4.9	9.72 ± 0.04	55.1	76.05 ± 5.09	212.1 ± 15.3	10.25 ± 0.33	52.9	27.0	5.3	14
Ga_2GC_G/CaC	83.98 ± 7.53	233.4 ± 23.3	11.61 ± 0.33	57.0	92.74 ± 4.16	260.4 ± 12.7	11.98 ± 0.25	56.4	9.9	3.1	14
Ga_2GC_G/Ca_2C	94.27 ± 7.44	265.7 ± 23.0	11.86 ± 0.32	55.6	97.07 ± 3.31	274.4 ± 10.5	11.96 ± 0.11	55.4	2.9	0.8	28
Ga_2GC_G/Ca_3C	86.22 ± 7.69	241.7 ± 23.9	11.25 ± 0.30	55.0	90.86 ± 3.59	256.1 ± 10.9	11.43 ± 0.20	54.7	5.2	1.6	14
Ga_2GC_G/CauaC	89.08 ± 6.95	250.2 ± 21.6	11.49 ± 0.27	55.3	102.9 ± 8.1	292.9 ± 25.0	12.04 ± 0.38	54.5	14.4	4.7	14
Ga_2GC_G/Ca_4C	84.63 ± 6.71	237.5 ± 20.9	10.96 ± 0.24	54.1	89.24 ± 5.12	251.8 ± 15.9	11.14 ± 0.20	53.9	5.3	1.6	14
Ga_2GCaG/C_C	85.91 ± 3.28	239.4 ± 10.1	11.65 ± 0.14	56.6	87.42 ± 3.01	244.1 ± 9.4	11.73 ± 0.13	56.6	1.7	0.7	29
Ga_2GCaG/CaC	96.91 ± 14.70	270.9 ± 45.0	12.88 ± 0.81	58.7	95.26 ± 3.75	265.8 ± 11.6	12.82 ± 0.23	58.9	1.7	0.5	15
Ga_2GCaG/Ca_2C	102.4 ± 5.2	288.1 ± 16.0	13.03 ± 0.25	58.0	90.16 ± 2.77	250.5 ± 8.8	12.46 ± 0.08	58.8	12.7	4.5	18
Ga_2GCaG/Ca_3C	112.8 ± 5.5	321.3 ± 17.1	13.19 ± 0.26	56.5	93.83 ± 2.78	262.8 ± 8.8	12.34 ± 0.13	57.5	18.4	6.7	29
Ga_2GCaG/CauaC	104.3 ± 4.8	294.6 ± 14.7	12.89 ± 0.24	57.2	110.4 ± 3.9	313.4 ± 11.9	13.19 ± 0.24	56.9	5.7	2.3	14
Ga_2GCaG/Ca_4C	92.33 ± 7.24	257.8 ± 22.3	12.36 ± 0.34	57.9	93.78 ± 6.73	262.2 ± 20.8	12.45 ± 0.28	57.9	1.6	0.7	14
Ga_2GCa_2G/C_C	103.5 ± 2.9	293.4 ± 8.9	12.51 ± 0.13	56.0	88.31 ± 3.32	246.5 ± 10.4	11.86 ± 0.12	56.9	15.9	5.3	28
Ga_2GCa_2G/CaC	100.3 ± 4.6	281.0 ± 14.2	13.09 ± 0.24	58.7	94.63 ± 2.81	263.8 ± 8.6	12.82 ± 0.17	59.1	5.8	2.1	28
Ga_2GCa_2G/Ca_2C	93.24 ± 7.03	259.1 ± 21.5	12.88 ± 0.38	59.7	95.75 ± 4.89	266.7 ± 15.0	13.03 ± 0.26	59.6	2.7	1.2	10
Ga_2GCa_2G/Ca_3C	94.05 ± 10.02	262.5 ± 30.8	12.65 ± 0.51	58.6	93.22 ± 6.74	259.8 ± 20.8	12.63 ± 0.29	58.7	0.9	0.2	14
Ga_2GCa_2G/CauaC	109.9 ± 7.1	311.7 ± 21.7	13.20 ± 0.34	57.0	105.6 ± 4.1	298.6 ± 12.6	13.02 ± 0.19	57.3	3.9	1.4	28
Ga_2GCa_2G/Ca_4C	102.8 ± 4.5	289.7 ± 14.0	12.90 ± 0.22	57.5	93.02 ± 5.69	259.7 ± 17.7	12.47 ± 0.21	58.1	10.0	3.4	28
G_UA_G/C_C	63.71 ± 3.60	175.9 ± 11.4	9.16 ± 0.09	50.4	61.48 ± 5.13	169.8 ± 16.0	9.16 ± 0.20	50.8	3.6	0.0	28
G_UA_G/CaC	99.01 ± 10.50	281.3 ± 32.5	11.75 ± 0.46	54.3	74.08 ± 6.63	204.2 ± 20.7	10.76 ± 0.31	55.8	28.8	8.8	14
G_UA_G/Ca_2C	111.3 ± 7.5	316.3 ± 23.1	13.15 ± 0.37	56.6	88.08 ± 5.64	245.1 ± 17.8	12.05 ± 0.12	57.8	23.3	8.7	15
G_UA_G/Ca_3C	111.8 ± 7.1	316.8 ± 21.8	13.50 ± 0.37	57.6	93.45 ± 12.09	260.8 ± 36.9	12.58 ± 0.66	58.5	17.8	7.1	15
G_UA_G/CauaC	112.7 ± 6.5	319.2 ± 19.9	13.65 ± 0.33	57.9	99.56 ± 12.73	279.0 ± 39.0	13.04 ± 0.65	58.7	12.3	4.6	28
G_UA_G/Ca_4C	87.06 ± 4.51	241.4 ± 13.8	12.20 ± 0.22	58.6	90.70 ± 6.31	252.4 ± 19.4	12.41 ± 0.32	58.5	4.1	1.7	29
GaUA_G/C_C	73.67 ± 2.79	205.8 ± 8.8	9.85 ± 0.07	51.6	72.64 ± 2.30	202.5 ± 7.2	9.84 ± 0.08	51.8	1.4	0.1	28
GaUA_G/CaC	92.96 ± 13.73	261.7 ± 42.5	11.79 ± 0.64	55.6	92.10 ± 8.02	259.0 ± 25.1	11.78 ± 0.28	55.7	0.9	0.9	14
GaUA_G/Ca_2C	98.17 ± 10.94	276.6 ± 33.6	12.39 ± 0.54	56.7	91.35 ± 11.53	255.4 ± 35.7	12.16 ± 0.48	57.4	7.2	1.9	28
GaUA_G/Ca_3C	91.18 ± 3.33	255.3 ± 10.3	12.00 ± 0.15	56.8	92.18 ± 9.16	258.3 ± 28.4	12.08 ± 0.38	56.9	1.1	0.7	29
GaUA_G/CauaC	91.21 ± 4.02	255.2 ± 12.4	12.05 ± 0.17	57.0	91.43 ± 6.24	255.8 ± 19.3	12.09 ± 0.27	57.1	0.2	0.3	14
GaUA_G/Ca_4C	89.57 ± 6.20	250.4 ± 19.1	11.91 ± 0.28	56.8	92.59 ± 9.07	259.5 ± 27.9	12.10 ± 0.43	56.8	3.3	1.6	28

[a] T_m is calculated for a total strand concentration of 0.1 mM.

(Figure 2, bottom). This is the more favorable of the coaxial stacking of helices (*6, 12, 23*) or of the stacking of unpaired nucleotides (*19, 26*). The experimental $\Delta G°_{bimol}$ value is taken from T_M^{-1} versus $\ln(C_T/4)$ plots. Rearrangement of eq 7 allows for the calculation of $\Delta G°_{MBL}$, the free-energy increment for the formation of the multibranch loop. Values at 37 °C for three- and four-way loops are listed in Tables 5 and 6, respectively.

In a simple nearest-neighbor model, $\Delta G°_{MBL}$ is a sum of favorable stacking interactions, noncanonical pairs, and initiation ($\Delta G°_{37\ MBL\ init}$). Initiation is unfavorable because it includes an entropic penalty for constraining the nucleotides in a loop and tethering the helices. Equation 1 shows the simple approximation for multibranch loop stability, where $\Delta G°_{MBL\ stacking}$ is the maximum stability imparted by terminal mismatches, dangling ends, and coaxial stacking of helices (*6, 12, 15, 23, 26*). The calculated $\Delta G°_{MBL\ init}$ at 37 °C for three- and four-way loops are in Tables 5 and 6, respectively.

For example, consider system GaCG_GC_G/Ca_4C as drawn in Figure 2 (bottom). The overall free energy of the lower-temperature transition is −11.18 kcal/mol at 37 °C, as determined by the linear fit of T_M^{-1} versus $\ln(C_T/4)$. At 37 °C, eq 7 gives

$$\Delta G°_{37\ MBL} = \Delta G°_{37\ bimol} - \Delta G°_{37\ helix\ 1} - \Delta G°_{37\ helix\ 2} - \Delta G°_{37\ bimol\ init} + \Delta G°_{37\ prod\ mm} \quad (8)$$

$$= -11.18 + 10.87 + 10.95 - 4.09 - 3.26 = 3.29\ \text{kcal/mol}$$

The $\Delta G°_{37\ prod\ mm}$ term is the free energy of coaxial stacking of the two helices in the long strand (*12, 25*). Note that, for cases in which the terminal intramolecular base pair in the long strand is AU, the terminal AU penalty (*25*) is not applied in the calculation because this pair is always present on both sides of the equilibrium in Figure 2.

To determine the optimal stacking configuration, all possibilities must be examined. The optimal configuration is that with the lowest sum of free-energy increments. For system GaCG_GC_G/Ca_4C (Figure 1, bottom), helix I could stack on helix II with an intervening AA mismatch, helix II could stack flush on helix III, or helix III could stack

Table 4: Thermodynamic Data from Optical Melting of Four-Way Multibranch Loops

	analysis of T_M dependence				average of curve fits						
system	$-\Delta H°$ (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)	$\Delta H°$ % diff	$\Delta G°_{37}$ % diff	conc range
G_CG_GC_G/C_C	120.6 ± 14.9	335.7 ± 44.7	16.45 ± 1.07	64.8	114.7 ± 11.7	317.9 ± 34.9	16.08 ± 0.91	65.2	5.0	2.3	21
G_CG_GC_G/CaC	112.4 ± 16.0	317.6 ± 49.0	13.91 ± 0.91	58.8	88.92 ± 9.72	245.6 ± 30.0	12.76 ± 0.45	60.4	23.3	8.6	21
G_CG_GC_G/Ca_2C	102.9 ± 7.2	291.8 ± 22.3	12.34 ± 0.31	55.6	93.5 ± 20.7	262.7 ± 64.0	12.02 ± 0.89	56.4	9.5	2.6	31
G_CG_GC_G/Ca_3C	83.11 ± 3.79	231.9 ± 11.8	11.17 ± 0.14	55.4	86.66 ± 13.70	242.8 ± 42.6	11.36 ± 0.50	55.3	4.2	1.7	31
G_CG_GC_G/Ca_4C	76.43 ± 10.28	212.7 ± 32.1	10.48 ± 0.42	53.9	87.7 ± 10.1	247.8 ± 32.6	10.84 ± 0.38	53	13.7	3.4	9
GaCG_GC_G/C_C	107.9 ± 6.3	305.6 ± 19.4	13.15 ± 0.29	57.3	87.15 ± 7.28	241.5 ± 22.8	12.24 ± 0.24	58.7	21.3	7.2	12
GaCG_GC_G/CaC	129.5 ± 13.5	367.9 ± 41.2	15.37 ± 0.79	59.7	99.05 ± 15.06	275.2 ± 46.4	13.7 ± 0.69	61.2	26.6	11.5	31
GaCG_GC_G/Ca_2C	125.3 ± 9.3	360.4 ± 28.7	13.53 ± 0.41	55.4	88.06 ± 1.97	245.5 ± 5.8	11.92 ± 0.31	57.2	34.9	12.7	30
GaCG_GC_G/Ca_3C	95.85 ± 10.00	270.8 ± 31.0	11.85 ± 0.41	55.2	80.66 ± 3.74	223.6 ± 11.9	11.31 ± 0.10	56.5	17.2	4.7	15
GaCG_GC_G/CauaC	103.8 ± 4.2	295.0 ± 14.6	12.28 ± 0.19	55.2	96.04 ± 8.15	270.9 ± 25.3	12.02 ± 0.31	55.8	7.7	2.1	28
GaCG_GC_G/Ca_4C	84.59 ± 3.11	236.7 ± 9.7	11.18 ± 0.11	55	84.05 ± 6.91	234.9 ± 21.4	11.21 ± 0.28	55.3	0.6	0.3	28
GaCG_GCaG/C_C	111.0 ± 7.1	314.1 ± 21.6	13.6 ± 0.37	58.1	97.08 ± 10.60	271.4 ± 32.3	12.9 ± 0.59	58.8	13.4	5.3	18
GaCG_GCaG/CaC	102.4 ± 13.0	287.1 ± 39.6	13.39 ± 0.76	59.3	93.39 ± 5.69	259.5 ± 17.4	12.91 ± 0.33	59.7	9.2	3.7	18
GaCG_GCaG/Ca_2C	133.3 ± 12.0	382.2 ± 36.7	14.78 ± 0.62	57.5	94.19 ± 5.61	262.2 ± 17.7	12.87 ± 0.24	59.4	34.4	13.8	28
GaCG_GCaG/Ca_3C	101.9 ± 6.1	286.6 ± 18.7	12.96 ± 0.29	57.9	98.38 ± 5.87	275.9 ± 18.0	12.81 ± 0.30	58.2	3.5	1.2	28
GaCG_GCaG/CauaC	92.21 ± 6.22	256.7 ± 19.1	12.6 ± 0.31	58.9	104.3 ± 13.1	293.7 ± 40.2	13.25 ± 0.69	58.3	12.3	5.0	28
GaCG_GCaG/Ca_4C	99.69 ± 8.85	280.6 ± 27.3	12.66 ± 0.40	57.3	90.02 ± 4.43	250.7 ± 13.7	12.27 ± 0.19	58.1	10.2	3.1	14
Ga_2CGa_2GCa_2G/C_C	98.46 ± 0.17	279.5 ± 0.5	11.76 ± 0.01	54.4	83.29 ± 6.05	232.4 ± 18.7	11.22 ± 0.27	55.5	16.7	4.7	23
Ga_2CGa_2GCa_2G/CaC	111.4 ± 8.3	318.4 ± 25.7	12.65 ± 0.35	55	84.55 ± 3.09	235.4 ± 9.5	11.56 ± 0.22	56.6	27.4	9.0	11
Ga_2CGa_2GCa_2G/Ca_2C	112.7 ± 1.9	322.2 ± 6.0	12.8 ± 0.08	55.3	81.73 ± 8.11	226.3 ± 25.4	11.53 ± 0.26	57.2	31.9	10.4	10

[a] T_m is calculated for a total strand concentration of 0.1 mM.

flush on helix IV. Note that helix IV cannot coaxially stack onto helix I with the current nearest-neighbor rules because the helices are separated by more than one unpaired nucleotide (*6*). The optimal calculated configuration for GaCG_GC_G/Ca_4C is the coaxial stacking of helix I onto helix II, with an intervening AA mismatch, and helix III onto helix IV. The total stability of these stacks is

$$\Delta G°_{37\ \text{stacking}} = \begin{matrix} 5'\text{GaC}3' \\ 3'\text{CaG}5' \\ \uparrow\downarrow \\ 5'3' \end{matrix} + \begin{matrix} 5'\ \text{C}\ \text{G} \\ 3'\ \text{G}\ \text{C} \end{matrix} \quad (9)$$

$$= -1.1 - 2.1 - 2.36 = -5.56\ \text{kcal/mol}$$

Therefore, from eq 1, the multibranch loop initiation free energy is

$$\Delta G°_{37\ \text{MBL init}} = \Delta G°_{37\ \text{MBL}} - \Delta G°_{37\ \text{MB stacking}} \quad (10)$$

$$= 3.29 + 5.56 = 8.9\ \text{kcal/mol}$$

Parametrizing the Loop Initiation Free Energies. Several secondary structure prediction algorithms (*6*, *8*, *9*) use eq 2 for predicting $\Delta G°_{37\ \text{MBL init}}$, but eq 2 does not adequately predict the stabilities observed experimentally for multibranch loops. For example, consider the multibranch loops GaCG_G/CauaC, G_CG_G/CcauaC, GaGC_G/Ca_3C, GaGC_G/CauaC, Ga_2GC_G/Ca_2C, Ga_2GCaG/CaC, Ga_2GCa_2G/C_C, G_UA_G/Ca_4C, GaUA_G/Ca_3C, and GaUA_G/CauaC in Table 5. Equation 2 predicts that all have the same initiation free energy because each has three branching helices and four unpaired nucleotides. The initiation free energies range from 7.5 to 10.5 kcal/mol, however. The coefficient of determination, R^2, of the linear regression fit of the data to eq 2 is 0.26, and the root-mean-square deviation is 1.23 kcal/mol, even with multibranch loops with outlying stabilities removed from the analysis.

The initiation free energies at 37 °C correlate better with asymmetry in the distribution of unpaired nucleotides than with the actual number of unpaired nucleotides. Also, three-way loops with less than two unpaired nucleotides are less stable than other loops. These results are approximated by eq 11

$$\Delta G°_{37\ \text{MBL init}} = a + b(\text{average asymmetry}) + ch + \Delta G°_{37\ \text{strain}}(\text{three-way loops with less than two unpaired nucleotides}) \quad (11)$$

where a−c are parameters. $\Delta G°_{37\ \text{strain}}$ is a strain free energy that applies only to three-way multibranch loops with zero or one unpaired nucleotide, and average asymmetry is a term that reflects the distribution of unpaired nucleotides in the loop as defined by

$$\text{average asymmetry} = \min\left[2.0, \frac{\left(\sum_{1}^{h}|\text{unpaired nucleotides } 5' - \text{unpaired nucleotides } 3'|\right)}{h}\right] \quad (12)$$

where h is the number of branching helices. The data suggest that the penalty term for loop asymmetry reaches a maximum at an average asymmetry of about 2.0; therefore, the average asymmetry is limited to 2.0 by eq 12.

Equation 12 was chosen empirically as a measure of multibranch loop asymmetry. Its use can be illustrated with system GaCG_GC_G/Ca_4C, drawn in Figure 1 (bottom). The sum in the numerator is over each helix, and the absolute value of the difference in unpaired nucleotides between the 5′ side and 3′ side of helices I−IV are 3, 1, 0, and 4, respectively. The mean (sum divided by number of helices) is therefore 2. The average asymmetry is tabulated for each system in Tables 5 and 6.

Comparisons of loop stabilities illustrate the dependence of $\Delta G°_{37\ \text{MBL init}}$ on asymmetry rather than on the number of unpaired nucleotides. Systems GaCG_G/CauaC and Ga_2GCaG/Ca_4C have the same average asymmetry (2.00) but differ in the number of unpaired nucleotides at 4 and 7, respectively. They have similar $\Delta G°_{37\ \text{MBL init}}$ values of 8.7 and 8.6 kcal/mol, respectively. Ga_2GCaG/Ca_4C and Ga_2-

Table 5: Loop Free Energies for Three-Way Multibranch Loops[a]

system	$\Delta G^\circ_{37\ \mathrm{MBL}}$[b]	$\Delta G^\circ_{37\ \mathrm{MBL\ stacking}}$	$\Delta G^\circ_{37\ \mathrm{MBL\ init}}$	unpaired	average asymmetry	$\Delta G^\circ_{37\ \mathrm{MBL\ init\ predicted}}$	$\Delta G^\circ_{37\ \mathrm{MBL\ init}} - \Delta G^\circ_{37\ \mathrm{MBL\ init\ predicted}}$
G_CG_G/Ca_2C[c]	3.64 ± 0.61	−5.1	8.8	2	1.33	8.6 ± 1.0	0.2
GaCG_G/CauaC[c]	4.90 ± 0.56	−4.9	9.8	4	2.00	9.2 ± 1.0	0.6
GaCG_G/CauaC[d]	3.76 ± 0.55	−4.9	8.7	4	2.00	9.2 ± 1.0	−0.5
GgCGuG/CcauaC[c,e]	1.26 ± 0.55	−4.5	5.8	6	1.33	8.6 ± 1.0	−2.8
GaCGcG/CauaC[c]	4.66 ± 0.57	−4.9	9.6	5	1.33	8.6 ± 1.0	1.0
GgCGuG/CauaC[c,e]	1.06 ± 0.55	−5.4	6.5	5	1.00	8.3 ± 1.0	−1.8
G_CG_G/C_C[c,e]	3.43 ± 0.59	−3.4	6.9	0	0.00	10.5 ± 1.1	−3.6
GaCG_G/CauacC[c]	4.37 ± 0.56	−4.4	8.7	5	2.67[f]	9.2 ± 1.0	−0.5
G_CG_G/CauaC[c]	3.72 ± 0.61	−5.1	8.8	3	2.00	9.2 ± 1.0	−0.4
G_CG_G/CcauaC[c]	3.23 ± 0.64	−4.2	7.5	4	2.67[f]	9.2 ± 1.0	−1.7
GaCGcG/CcauaC[c]	4.20 ± 0.55	−4.3	8.5	6	2.00	9.2 ± 1.0	−0.7
GaCGaG/CcauaC[c]	3.54 ± 0.54	−4.2	7.7	6	2.00	9.2 ± 1.0	−1.5
G_GC_G/C_C	7.73 ± 0.59	−3.3	11.0	0	0.00	10.5 ± 1.1	0.5
G_GC_G/CaC	5.73 ± 0.62	−5.0	10.7	1	0.67	11.1 ± 1.1	−0.4
G_GC_G/Ca_2C	4.63 ± 0.71	−5.0	9.6	2	1.33	8.6 ± 1.0	1.0
G_GC_G/Ca_3C	4.64 ± 0.65	−5.0	9.6	3	2.00	9.2 ± 1.0	0.4
G_GC_G/CauaC	3.27 ± 0.73	−5.0	8.3	3	2.00	9.2 ± 1.0	−0.9
G_GC_G/Ca_4C	3.66 ± 0.86	−5.0	8.6	4	1.33	8.6 ± 1.0	0.0
GaGC_G/C_C	7.52 ± 0.61	−3.5	11.0	1	0.67	11.1 ± 1.1	−0.1
GaGC_G/CaC	5.80 ± 0.62	−3.5	9.3	2	1.33	8.6 ± 1.0	0.7
GaGC_G/Ca_2C	5.19 ± 0.55	−4.9	10.1	3	1.33	8.6 ± 1.0	1.5
GaGC_G/Ca_3C	5.63 ± 0.55	−4.9	10.5	4	2.00	9.2 ± 1.0	1.3
GaGC_G/CauaC	4.58 ± 0.61	−4.9	9.5	4	2.00	9.2 ± 1.0	0.3
GaGC_G/Ca_4C	5.36 ± 0.60	−4.9	10.3	5	2.67[f]	9.2 ± 1.0	1.1
Ga_2GC_G/C_C	6.61 ± 0.61	−3.5	10.1	2	1.33	8.6 ± 1.0	1.5
Ga_2GC_G/CaC	4.72 ± 0.81	−3.5	8.2	3	1.33	8.6 ± 1.0	−0.4
Ga_2GC_G/Ca_2C	4.47 ± 0.69	−3.5	7.9	4	1.33	8.6 ± 1.0	−0.7
Ga_2GC_G/Ca_3C	5.08 ± 0.68	−3.5	8.5	5	2.00	9.2 ± 1.0	−0.7
Ga_2GC_G/CauaC	4.84 ± 0.67	−3.5	8.3	5	2.00	9.2 ± 1.0	−0.9
Ga_2GC_G/Ca_4C	5.37 ± 0.65	−3.5	8.8	6	2.67[f]	9.2 ± 1.0	−0.4
Ga_2GCaG/C_C	4.58 ± 0.62	−4.8	9.3	3	1.33	8.6 ± 1.0	0.7
Ga_2GCaG/CaC	3.35 ± 0.98	−4.7	8.1	4	0.67	8.0 ± 1.0	0.1
Ga_2GCaG/Ca_2C	2.80 ± 0.60	−4.7	7.5	5	0.67	8.0 ± 1.0	−0.5
Ga_2GCaG/Ca_3C	3.04 ± 0.60	−4.7	7.7	6	1.33	8.6 ± 1.0	−0.9
Ga_2GCaG/CauaC	3.34 ± 0.60	−4.7	8.0	6	1.33	8.6 ± 1.0	−0.6
Ga_2GCaG/Ca_4C	3.87 ± 0.64	−4.7	8.6	7	2.00	9.2 ± 1.0	−0.6
Ga_2GCa_2G/C_C	3.72 ± 0.62	−4.8	8.5	4	1.33	8.6 ± 1.0	0.1
Ga_2GCa_2G/CaC	3.14 ± 0.60	−5.1	8.2	5	0.67	8.0 ± 1.0	0.2
Ga_2GCa_2G/Ca_2C	2.95 ± 0.55	−4.1	7.1	6	0.00	7.4 ± 1.0	−0.3
Ga_2GCa_2G/Ca_3C	3.58 ± 0.65	−4.1	7.7	7	0.67	8.0 ± 1.0	−0.3
Ga_2GCa_2G/CauaC	3.03 ± 0.52	−4.1	7.1	7	0.67	8.0 ± 1.0	−0.9
Ga_2GCa_2G/Ca_4C	3.33 ± 0.45	−4.1	7.4	8	1.33	8.6 ± 1.0	−1.2
G_UA_G/C_C	7.77 ± 0.58	−3.3	11.0	0	0.00	10.5 ± 1.1	0.5
G_UA_G/CaC[e]	5.18 ± 0.68	−3.9	9.1	1	0.67	11.1 ± 1.1	−2.0
G_UA_G/Ca_2C[e]	3.78 ± 0.71	−3.9	7.7	2	1.33	8.6 ± 1.0	−0.9
G_UA_G/Ca_3C[e]	3.43 ± 0.71	−3.9	7.4	3	2.00	9.2 ± 1.0	−1.8
G_UA_G/CauaC[e]	3.28 ± 0.69	−3.9	7.2	3	2.00	9.2 ± 1.0	−2.0
G_UA_G/Ca_4C[e]	4.73 ± 0.65	−3.9	8.7	4	2.67[f]	9.2 ± 1.0	−0.5
GaUA_G/C_C	7.08 ± 0.61	−3.6	10.6	1	0.67	11.1 ± 1.1	−0.5
GaUA_G/CaC	5.14 ± 0.82	−3.2	8.3	2	0.67	8.0 ± 1.0	0.3
GaUA_G/Ca_2C	4.54 ± 0.77	−4.9	9.4	3	1.33	8.6 ± 1.0	0.8
GaUA_G/Ca_3C	4.93 ± 0.57	−4.9	9.8	4	2.00	9.2 ± 1.0	0.6
GaUA_G/CauaC	4.48 ± 0.57	−4.9	9.4	4	2.00	9.2 ± 1.0	0.2
GaUA_G/Ca_4C	5.02 ± 0.61	−4.9	9.9	5	2.67[f]	9.2 ± 1.0	0.7

[a] Values are given in kcal/mol at 37 °C. [b] The error in $\Delta G^\circ_{37\ \mathrm{MBL}}$ was calculated with the assumption that the error in free energy for dangling ends and terminal mismatches is 0.2 kcal/mol (*68*). The errors for helical parameters are from Xia et al. (*25*), and those for coaxial stacking are from Mathews et al. (*6*). [c] Systems from Diamond et al. (*15*) were constructed with the sequences 5′GGAGxCGGCUUCGGCCGyGACG3′ and 5′CGUCzCUCC3′. [d] This system, from Diamond et al. (*15*), is for the sequences 5′GGAGaCGGCUUCGGCCG_GCAG3′ and 5′CUGCAUACUCC3′. [e] These systems were excluded from the determination of the multibranch loop parameters by linear regression. [f] The average asymmetry is set to 2.0 for predicting $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$.

GCa_2G/CauaC, however, have the same number of unpaired nucleotides, seven, but a different average asymmetry and different $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$ values of 8.6 and 7.1 kcal/mol, respectively.

Linear regression was used to derive the three parameters for eq 11 (Table 7). The coefficient of determination, R^2, was 0.58, and the p value of the F test was 1.9×10^{-9}, indicating a very low probability that the correlation between $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$ and eq 11 is due to random chance. The root-mean-square deviation for sequences included in the fit is 0.79 kcal/mol, which suggests that this model predicts the loop initiation free energies within about 1 kcal/mol on

Table 6: Loop Free Energies for Four-Way Multibranch Loops[a]

system	$\Delta G^\circ_{37\ \mathrm{MBL}}$[b]	$\Delta G^\circ_{37\ \mathrm{MBL\ stacking}}$	$\Delta G^\circ_{37\ \mathrm{MBL\ init}}$	unpaired	average asymmetry	$\Delta G^\circ_{37\ \mathrm{MBL\ init\ predicted}}$	$\Delta G^\circ_{37\ \mathrm{MBL\ init}} - \Delta G^\circ_{37\ \mathrm{MBL\ init\ predicted}}$
G_CG_GC_G/C_C[c]	−4.08 ± 1.33	−6.5	2.4	0	0.00	6.7 ± 1.0	−4.3
G_CG_GC_G/CaC	0.16 ± 1.20	−5.8	5.9	1	0.50	7.2 ± 1.0	−1.3
G_CG_GC_G/Ca_2C	1.73 ± 0.85	−5.8	7.5	2	1.00	7.6 ± 1.0	−0.1
G_CG_GC_G/Ca_3C	2.90 ± 0.80	−5.8	8.7	3	1.50	8.1 ± 1.1	0.6
G_CG_GC_G/Ca_4C	3.59 ± 0.89	−5.8	9.4	4	2.00	8.5 ± 1.1	0.9
GaCG_GC_G/C_C	0.92 ± 0.84	−6.5	7.4	1	0.50	7.2 ± 1.0	0.2
GaCG_GC_G/CaC	−1.30 ± 1.21	−6.9	5.6	2	0.50	7.2 ± 1.0	−1.6
GaCG_GC_G/Ca_2C	0.54 ± 0.85	−5.6	6.1	3	1.00	7.6 ± 1.0	−1.5
GaCG_GC_G/Ca_3C	2.22 ± 0.85	−5.6	7.8	4	1.50	8.1 ± 1.1	−0.3
GaCG_GC_G/CauaC	2.19 ± 0.76	−5.6	7.8	4	1.50	8.1 ± 1.1	−0.3
GaCG_GC_G/Ca_4C	3.29 ± 0.56	−5.6	8.9	5	2.00	8.5 ± 1.1	0.4
GaCG_GCaG/C_C	0.87 ± 0.87	−6.5	7.4	2	1.00	7.6 ± 1.0	−0.2
GaCG_GCaG/CaC	1.08 ± 0.99	−6.9	7.9	3	0.50	7.2 ± 1.1	0.7
GaCG_GCaG/Ca_2C	−0.31 ± 0.85	−6.8	6.5	4	1.00	7.6 ± 1.0	−1.1
GaCG_GCaG/Ca_3C	1.51 ± 0.65	−6.8	8.3	5	1.50	8.1 ± 1.1	0.2
GaCG_GCaG/CauaC	1.87 ± 0.66	−6.8	8.7	5	1.50	8.1 ± 1.1	0.6
GaCG_GCaG/Ca_4C	1.81 ± 0.70	−6.8	8.6	6	2.00	8.5 ± 1.1	0.1
$Ga_2CGa_2GCa_2$G/C_C	2.97 ± 0.64	−5.9	8.8	6	1.00	7.6 ± 1.0	1.2
$Ga_2CGa_2GCa_2$G/CaC	2.08 ± 0.68	−6.2	8.3	7	0.50	7.2 ± 1.0	1.1
$Ga_2CGa_2GCa_2$G/Ca_2C	1.93 ± 0.45	−5.2	7.1	8	0.00	6.7 ± 1.0	0.4

[a] Values are given in kcal/mol at 37 °C. [b] The error in $\Delta G^\circ_{37\ \mathrm{MBL}}$ was calculated with the assumption that the error in free energy for dangling ends and terminal mismatches is 0.2 kcal/mol (*68*). The errors for helical parameters are from Xia et al. (*25*), and those for coaxial stacking are from Mathews et al. (*6*). [c] This system was excluded from the determination of the multibranch loop parameters by linear regression.

Table 7: Nearest-Neighbor Parameters for Multibranch Loop Initiation[a]

	$\Delta C_p^\circ = 0$		$\Delta C_p^\circ \neq 0$	
parameter	value (kcal/mol)	error (kcal/mol)	value (kcal/mol)	error (kcal/mol)
a	9.25	0.91	9.34	1.01
b	0.91	0.19	0.77	0.21
c	−0.63	0.24	−0.43	0.26
$\Delta G^\circ_{37\ \mathrm{strain}}$	3.14	0.44	2.60	0.48

[a] These parameters are used to approximate the free energy of multibranch loop initiation at 37 °C from eqs 11 and 12.

average. While this simple model fits the data reasonably well, it does not eliminate the possibility that other models may also fit the data.

Systems GgCGuG/CcauaC (*15*), GgCGuG/CauaC (*15*), G_CG_G/C_C (*15*), G_UA_G/CaC, G_UA_G/Ca_2C, G_UA_G/Ca_3C, G_UA_G/CauaC, G_UA_G/Ca_4C, and G_CG_GC_G/C_C were excluded from the regression analysis. System GgCGuG/CcauaC, a mimic of the 5S rRNA multibranch loop, was excluded because it was much more stable than the other systems, probably because of tertiary interactions (*15*). GgCGuG/CauaC and G_CG_G/C_C were excluded because the optical melting curves showed multiple transitions and their stabilities cannot be considered reliable (*15*). The systems with the strand G_UA_G were excluded because alternative conformations involving the U in the terminal UA base pair were possible with the adenines in the loop. G_CG_GC_G/C_C was excluded because it was significantly more stable than other similar multibranch loops. It was the only four-way loop studied without unpaired nucleotides. Interestingly, in a database of known secondary structures (*6*), the only multibranch loops that appear without unpaired nucleotides are four-way loops. This suggests that four-way loops without unpaired nucleotides may have enhanced stability.

Heat Capacity. The thermodynamic parameters shown in Tables 3 and 4 are derived assuming that enthalpy and entropy changes are constant with temperature. As illustrated by Figure 3, however, the enthalpy and entropy changes from the individual fits of melting curves vary with T_M, as would be expected if the heat capacity difference between the single strands and the duplex, ΔC_p°, is nonzero. Tables 8 and 9 give the ΔC_p° value calculated for each system by a linear fit of ΔH° versus T_M and ΔS° versus $\ln(T_M/273.15)$ and the average of the two fits. The two calculations of ΔC_p° for each multibranch loop system agree well, with percent differences most often below 5%. The average ΔC_p° is −1.74 kcal mol^{-1} K^{-1}, indicating a reduced magnitude of enthalpy and entropy changes at lower temperatures. Per broken base pair, this is −0.174 kcal mol^{-1} K^{-1}, similar to values obtained

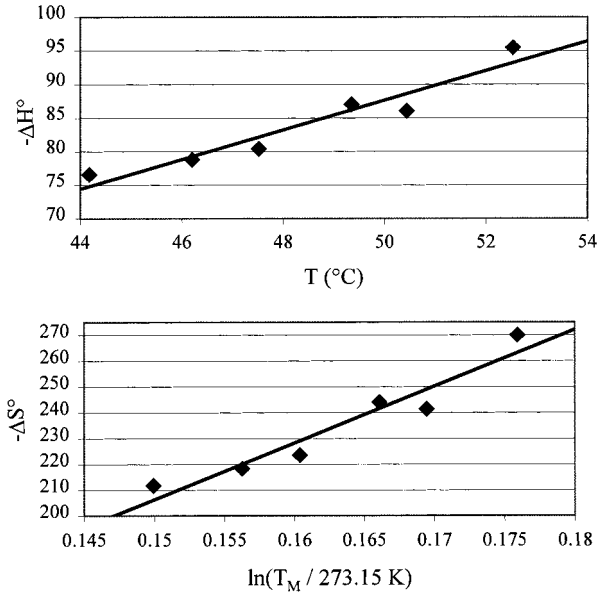

FIGURE 3: Fits of enthalpy change as a function of temperature (top) and entropy change as a function of the natural logarithm of absolute temperature divided by the reference temperature of 273.15 K (bottom), for the system GaCG_GC_G/Ca_4C. ΔC_p° is the slope of each fitted line, 2.20 and 2.19 kcal mol^{-1} K^{-1} for the top and bottom graphs, respectively.

Table 8: ΔC_p° for Three-Way Multibranch Loops

system	$-\Delta C_p^\circ$ from ΔH° (kcal mol^{-1} K^{-1})	$-\Delta C_p^\circ$ from ΔS° (kcal mol^{-1} K^{-1})	average $-\Delta C_p^\circ$ (kcal mol^{-1} K^{-1})	$-\Delta H^\circ_{37}$ (kcal/mol)	$-\Delta S^\circ_{37}$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)
G_GC_G/C_C	1.16 ± 0.33	1.17 ± 0.32	1.17	60.4 ± 1.9	168 ± 6	8.4 ± 0.1
G_GC_G/CaC	0.78 ± 0.28	0.80 ± 0.28	0.79	73.4 ± 2.7	203 ± 8	10.3 ± 0.1
G_GC_G/Ca$_2$C	2.30 ± 0.70	2.29 ± 0.71	2.30	59.3 ± 8.8	156 ± 28	10.7 ± 0.1
G_GC_G/Ca$_3$C	2.29 ± 0.36	2.24 ± 0.37	2.27	59.4 ± 5.1	157 ± 16	10.7 ± 0.1
G_GC_G/CauaC	0.97 ± 0.97	0.95 ± 0.95	0.96	97.4 ± 15.1	274 ± 47	12.5 ± 0.7
G_GC_G/Ca$_4$C	2.06 ± 0.44	2.06 ± 0.47	2.06	65.7 ± 6.1	174 ± 21	11.7 ± 0.3
GaGC_G/C_C	1.90 ± 0.32	1.84 ± 0.31	1.87	59.1 ± 1.7	162 ± 5	8.8 ± 0.1
GaGC_G/CaC	1.69 ± 0.36	1.66 ± 0.36	1.68	61.0 ± 3.8	164 ± 12	10.2 ± 0.1
GaGC_G/Ca$_2$C	1.12 ± 1.43	1.11 ± 1.41	1.12	73.8 ± 18.7	202 ± 58	11.2 ± 0.6
GaGC_G/Ca$_3$C	2.29 ± 1.11	2.22 ± 1.11	2.26	61.6 ± 12.7	166 ± 40	10.1 ± 0.3
GaGC_G/CauaC	5.02 ± 1.43	4.96 ± 1.44	4.99	38.2 ± 18.1	91.3 ± 57.8	9.9 ± 0.3
GaGC_G/Ca$_4$C	1.08 ± 0.76	1.05 ± 0.76	1.07	74.9 ± 9.3	207 ± 29	10.7 ± 0.3
Ga$_2$GC_G/C_C	1.37 ± 0.15	1.31 ± 0.15	1.34	64.5 ± 1.3	177 ± 4	9.6 ± 0.1
Ga$_2$GC_G/CaC	1.32 ± 0.52	1.29 ± 0.52	1.31	75.4 ± 7.0	207 ± 22	11.2 ± 0.2
Ga$_2$GC_G/Ca$_2$C	0.46 ± 0.54	0.47 ± 0.56	0.47	103.1 ± 7.3	294 ± 24	11.9 ± 0.2
Ga$_2$GC_G/Ca$_3$C	1.20 ± 0.45	1.19 ± 0.43	1.20	77.3 ± 5.2	214 ± 16	11.0 ± 0.4
Ga$_2$GC_G/CauaC	2.95 ± 0.87	2.90 ± 0.88	2.93	68.0 ± 10.5	184 ± 34	10.8 ± 0.1
Ga$_2$GC_G/Ca$_4$C	1.89 ± 0.40	1.88 ± 0.41	1.89	69.7 ± 4.2	190 ± 14	10.6 ± 0.1
Ga$_2$GCaG/C_C	0.65 ± 0.37	0.64 ± 0.38	0.65	78.4 ± 5.2	216 ± 17	11.4 ± 0.1
Ga$_2$GCaG/CaC	1.51 ± 0.25	1.50 ± 0.28	1.51	71.6 ± 3.9	192 ± 14	12.0 ± 0.4
Ga$_2$GCaG/Ca$_2$C	1.13 ± 0.22	1.17 ± 0.22	1.15	72.5 ± 3.4	193 ± 11	12.6 ± 0.1
Ga$_2$GCaG/Ca$_3$C	0.77 ± 0.45	0.84 ± 0.45	0.81	82.3 ± 6.7	224 ± 21	13.0 ± 0.3
Ga$_2$GCaG/CauaC	0.28 ± 1.05	0.29 ± 1.04	0.29	114.8 ± 16.8	328 ± 52	13.2 ± 0.7
Ga$_2$GCaG/Ca$_4$C	1.93 ± 1.16	1.92 ± 1.17	1.93	65.9 ± 17.0	175 ± 54	11.6 ± 0.3
Ga$_2$GCa$_2$G/C_C	0.27 ± 0.64	0.31 ± 0.65	0.29	84.5 ± 9.2	233 ± 29	12.3 ± 0.3
Ga$_2$GCa$_2$G/CaC	0.54 ± 0.46	0.55 ± 0.46	0.55	85.8 ± 7.8	235 ± 24	12.9 ± 0.4
Ga$_2$GCa$_2$G/Ca$_2$C	1.83 ± 0.93	1.81 ± 0.94	1.82	64.2 ± 16.1	168 ± 51	11.8 ± 0.3
Ga$_2$GCa$_2$G/Ca$_3$C	1.81 ± 1.24	1.81 ± 1.25	1.81	65.6 ± 19.0	173 ± 60	11.9 ± 0.4
Ga$_2$GCa$_2$G/CauaC	1.30 ± 0.55	1.32 ± 0.55	1.31	85.8 ± 8.4	235 ± 26	12.8 ± 0.3
Ga$_2$GCa$_2$G/Ca$_4$C	1.91 ± 0.57	1.93 ± 0.58	1.92	63.7 ± 8.9	166 ± 28	12.1 ± 0.2
G_UA_G/C_C	0.58 ± 0.51	0.57 ± 0.51	0.58	57.9 ± 4.0	158 ± 12	9.1 ± 0.1
G_UA_G/CaC	1.44 ± 1.48	1.52 ± 1.47	1.48	55.3 ± 19.6	142 ± 61	11.4 ± 0.9
G_UA_G/Ca$_2$C	2.46 ± 0.75	2.53 ± 0.78	2.50	49.3 ± 11.9	120 ± 39	12.2 ± 0.1
G_UA_G/Ca$_3$C	0.92 ± 3.41	1.00 ± 3.39	0.96	78.1 ± 57.3	208 ± 179	13.4 ± 1.9
G_UA_G/CauaC	3.31 ± 2.13	3.31 ± 2.14	2.90	45.1 ± 35.4	108 ± 111	11.7 ± 1.0
G_UA_G/Ca$_4$C	1.53 ± 0.69	1.51 ± 0.71	1.52	66.6 ± 11.1	178 ± 35	11.4 ± 0.2
GaUA_G/C_C	0.65 ± 0.08	0.65 ± 0.08	0.65	67.3 ± 0.7	185 ± 2	9.8 ± 0.1
GaUA_G/CaC	1.55 ± 1.70	1.53 ± 1.72	1.54	72.7 ± 21.7	198 ± 69	11.3 ± 0.4
GaUA_G/Ca$_2$C	4.00 ± 0.68	4.02 ± 0.68	4.01	33.2 ± 10.1	71.3 ± 31.4	11.1 ± 0.4
GaUA_G/Ca$_3$C	1.62 ± 1.37	1.60 ± 1.38	1.61	69.1 ± 19.8	186 ± 63	11.2 ± 0.4
GaUA_G/CauaC	1.94 ± 1.10	1.94 ± 1.10	1.94	64.5 ± 15.4	170 ± 48	11.6 ± 0.4
GaUA_G/Ca$_4$C	2.72 ± 0.73	2.70 ± 0.74	2.71	54.1 ± 10.5	139 ± 34	10.9 ± 0.2

for other duplexes (*15*, *19*, *27*−*34*). Tables 8 and 9 also give ΔH°, ΔS°, and ΔG° extrapolated to 37 °C from the linear fits. Table 7 shows the derived values of $a-c$ and $\Delta G^\circ_{37\ \text{strain}}$ when heat capacity is considered.

DISCUSSION

Multibranch loops are common in RNA secondary structures (*1*−*4*, *35*−*42*), but little is known about the factors that affect their stabilities. The database of stabilities determined for RNA multibranch loops in this work and in the work of Diamond et al. (*15*) provides an opportunity to search for factors that affect multibranch loop stability.

Coaxial Stacking in the Context of a Multibranch Loop. The substitution of a UA pair for the terminal intramolecular GC pair in the long strand provides an opportunity to compare multibranch loops with an identical configuration except for the closing base pair. If the initiation free energy is sequence-independent as hypothesized and if the parameters for coaxial stacking stability are reasonable, then loops that only differ by the closing pair should have the same initiation free energy. There are seven such pairs for this comparison: G_GC_G/C_C and G_UA_G/C_C, GaGC_G/C_C and GaUA_G/C_C, GaGC_G/CaC and GaUA_G/CaC, GaGC_G/Ca$_2$C and GaUA_G/Ca$_2$C, GaGC_G/Ca$_3$C and GaUA_G/Ca$_3$C, GaGC_G/CauaC and GaUA_G/CauaC, and GaGC_G/Ca$_4$C and GaUA_G/Ca$_4$C. For these pairs, the $\Delta G^\circ_{37\ \text{MBL init}}$ difference is 1.0 kcal/mol at most, and the root-mean-square deviation is 0.6 kcal/mol. This is a reasonable deviation from expected behavior, considering the experimental errors (*6*). Any deviation is always such that the loop initiation penalty determined for the UA closed multibranch loop is smaller than that determined for the GC closed multibranch loop, with an average difference in $\Delta G^\circ_{37\ \text{MBL init}}$ of 0.5 kcal/mol, suggesting that the stability of coaxial stacking with UA pairs in multibranch loops may be underestimated (*6*, *12*, *23*). Perhaps UA pairs are more flexible than GC pairs and are therefore more able to adapt to local structural contexts. This flexibility may involve non-Watson−Crick as well as Watson−Crick pairing of UA.

Thermodynamic Parameters for RNA Multibranch Loop Stability. Linear regression analysis was used on the database of experimentally determined stabilities to test functional forms and to determine parameters for approximating multibranch loop stabilities. The results show that the commonly

Table 9: ΔC_p° for Four-Way Multibranch Loops

system	$-\Delta C_p^\circ$ from ΔH° (kcal mol^{-1} K^{-1})	$-\Delta C_p^\circ$ from ΔS° (kcal mol^{-1} K^{-1})	average $-\Delta C_p^\circ$ (kcal mol^{-1} K^{-1})	$-\Delta H^\circ_{37}$ (kcal/mol)	$-\Delta S^\circ_{37}$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)
G_CG_GC_G/C_C	3.04 ± 2.26	3.03 ± 2.24	3.04	45.6 ± 51.7	104 ± 159	13.4 ± 2.5
G_CG_GC_G/CaC	3.11 ± 1.52	3.15 ± 1.53	3.13	37.4 ± 25.3	82.4 ± 79.9	11.9 ± 0.7
G_CG_GC_G/Ca$_2$C	3.58 ± 3.56	3.62 ± 3.54	3.60	46.3 ± 47.7	112 ± 149	11.5 ± 1.4
G_CG_GC_G/Ca$_3$C	2.06 ± 1.89	2.07 ± 1.89	2.07	62.5 ± 22.8	166 ± 72.1	10.9 ± 0.5
G_CG_GC_G/Ca$_4$C	2.12 ± 2.21	2.09 ± 2.20	2.11	67.6 ± 21.4	185 ± 68	10.3 ± 0.5
GaCG_GC_G/C_C	2.68 ± 0.88	2.74 ± 0.88	2.71	47.4 ± 13.1	113 ± 41	12.2 ± 0.3
GaCG_GC_G/CaC	2.77 ± 3.20	2.88 ± 3.20	2.83	48.3 ± 58.9	110 ± 186	14.1 ± 1.5
GaCG_GC_G/Ca$_2$C	−0.25 ± 0.46	−0.14 ± 0.45	−0.20	91.6 ± 6.5	252 ± 20	13.4 ± 0.9
GaCG_GC_G/Ca$_3$C	0.99 ± 0.72	1.02 ± 0.74	1.01	69.0 ± 8.6	186 ± 28	11.5 ± 0.2
GaCG_GC_G/CauaC	3.07 ± 0.54	3.08 ± 0.54	3.08	56.9 ± 7.0	147 ± 22	11.4 ± 0.2
GaCG_GC_G/Ca$_4$C	2.20 ± 0.32	2.19 ± 0.31	2.20	59.1 ± 3.8	156 ± 12	10.7 ± 0.2
GaCG_GCaG/C_C	1.12 ± 2.50	1.17 ± 2.48	1.15	78.2 ± 42.3	209 ± 132	13.2 ± 1.5
GaCG_GCaG/CaC	0.90 ± 1.29	0.90 ± 1.29	0.90	77.2 ± 23.4	209 ± 73	12.4 ± 0.9
GaCG_GCaG/Ca$_2$C	2.49 ± 0.63	2.60 ± 0.61	2.55	54.0 ± 10.3	130.2 ± 31	13.6 ± 0.8
GaCG_GCaG/Ca$_3$C	1.22 ± 0.96	1.22 ± 0.96	1.22	79.8 ± 14.9	217 ± 47	12.4 ± 0.5
GaCG_GCaG/CauaC	3.49 ± 1.50	3.44 ± 1.52	3.46	50.0 ± 23.8	125 ± 75	11.1 ± 0.5
GaCG_GCaG/Ca$_4$C	1.80 ± 0.51	1.83 ± 0.49	1.82	65.0 ± 7.1	171 ± 22	12.0 ± 0.4
Ga$_2$CGa$_2$GCa$_2$G/C_C	−0.47 ± 1.28	−0.42 ± 1.28	−0.45	88.6 ± 14.6	247 ± 46	11.9 ± 0.4
Ga$_2$CGa$_2$GCa$_2$G/CaC	0.05 ± 0.99	0.12 ± 0.98	0.08	83.9 ± 13.2	230 ± 41	12.5 ± 0.8
Ga$_2$CGa$_2$GCa$_2$G/Ca$_2$C	3.71 ± 1.75	3.80 ± 1.75	3.76	31.7 ± 23.7	65.1 ± 74.7	11.6 ± 0.6

employed eq 2 does not reproduce experimental trends. The relatively simple eq 11, however, does provide a statistically significant fit for the available data. The most significant difference between eq 2 and eq 11 is that the asymmetry in the distribution of unpaired nucleotides in the loop is more important for predicting loop stability than the number of unpaired nucleotides. This is similar to the asymmetry penalty of 0.48 kcal/mol that is important for predicting internal-loop free-energy increments (*6*, *43*−*45*). Symmetric distributions of unpaired nucleotides may facilitate noncanonical base pair formation. It is also possible that a symmetric arrangement of unpaired nucleotides in the loops is more able to accommodate a favorable stacking of nucleotides into the loop.

The Jacobson and Stockmayer theory (*46*) suggests that closing an unstructured loop is associated with an entropic penalty that should increase with a logarithmic dependence on the number of unpaired nucleotides. The experimental stability data show no dependence on the number of nucleotides. Perhaps this is because the loops are structured with both stacking and noncanonical pairs. It is likely that the increase in entropy with logarithmic dependence on unpaired nucleotides will be important for predicting the melting behavior of RNA because, as the RNA denatures, the multibranch loops are likely to become unstructured and larger.

Equation 11 contains a term that depends on the number of branching helices so that multibranch loops with more helices are more favorable (Table 7). Because this parameter was derived by considering only three- and four-way branching loops, there are three possible interpretations of this parameter. It may reflect a true decrease in $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$ with an increasing number of branches. This is possible if increased branching allows for more loop flexibility for adopting favorable coaxial stacking or if there is an excluded volume effect that reduces the entropic cost of constraining nucleotides in the loop. Alternatively, this term may reflect a difference between three- and four-way loop stability that may ultimately need to be measured for each possible number of branching helices in a multibranch loop. Finally, this term may instead reflect a difference in stability between multibranch loops with an odd number or an even number of helical branches. This is conceivable because an even number of branches allows for pairwise coaxial stacking that involves each helix. Multibranch loops with an odd number of branching helices will have at least one helix that cannot coaxially stack. Thus, eq 11 may not be general for all types of multibranch loops.

Comparisons to Studies of DNA Multibranch Loop Stability. The stabilities determined for RNA multibranch loops in this study and in the work of Diamond et al. (*15*) are consistent with qualitative conclusions drawn previously for DNA loop stability (*47*, *48*). In both RNA and DNA, unpaired nucleotides stabilize a three-way multibranch loop, and unpaired nucleotides in a four-way multibranch loop can be destabilizing relative to a loop without unpaired nucleotides. Similarly constructed four-way multibranch loops without unpaired nucleotides are more favorable than three-way multibranch loops without unpaired nucleotides.

Change in Heat Capacity. The change in heat capacity, ΔC_p°, for the multibranch loop systems, Tables 8 and 9, demonstrates a trend of reduced magnitude of enthalpy and entropy changes of association at lower temperatures. It is probable that the cases in which the heat capacity change is positive are misleading and are a result of experimental error that arises from the relatively small temperature range that is considered. The order of magnitude of the heat capacity change per base pair is similar to that found in other studies (*15*, *19*, *27*−*34*), and these data suggest that the heat capacity change must be considered for extrapolations of ΔH° and ΔS° to temperatures away from the melting temperature. The importance of these extrapolations to the temperature of interest is illustrated by the difference in the values of ΔH° and ΔS° between Tables 3 and 4 (temperature independent values) and Tables 8 and 9 (values extrapolated to 37 °C).

The adjustments in ΔH° and ΔS° when extrapolated to 37 °C tend to compensate in ΔG°_{37}. Therefore, the ΔG°_{37} value calculated from temperature independent parameters, Tables 3 and 4, is close to that calculated from the values of ΔH° and ΔS° extrapolated to 37 °C, Tables 8 and 9. On

average, the percent difference in ΔG°_{37} calculated by these two methods is 5%.

The negative heat capacity change is consistent with the hypothesis that the heat capacity change arises from the single strands being more organized at low rather than at high temperature (*19, 28, 49*). The extent of stacking in single strands is known to increase with decreasing temperature (*32, 49*−*51*).

Limitations of the Nearest-Neighbor Model. Tables 5 and 6 show the difference between the predicted $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$ and the measured $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$. For the sequences used to fit the parameters, the root-mean-square deviation between the two values is 0.79 kcal/mol, suggesting that there are important interactions that are not known. For example, unpaired nucleotides not adjacent to helix termini may contribute to loop stability, an interaction neglected by this simple model. Multibranch systems that differ by having short strands CauaC or Ca_3C test this hypothesis. For nine such comparisons, the root-mean-square deviation in $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$ is 0.47 kcal/mol, with the aua-containing strands usually having a more favorable $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$. Systems G_GC_G/$Ca_3$C and G_GC_G/CauaC have the largest deviation, with the latter more favorable by 1.37 kcal/mol. Furthermore, Diamond et al. (*15*) demonstrated that multibranch loop stabilities are not strictly nearest-neighbor because two identical loops closed by helices that differed in positions beyond the closing base pair were different in stability by 0.86 kcal/mol.

Crystallography reveals intricate structures that are not considered in this model for multibranch loops (e.g., base triples between loop nucleotides and adjacent helices (*52*−*56*)). The mimic of the 5S rRNA multibranch loop (system GgCGuG/CcauaC) is 2.8 kcal/mol more stable than the nearest-neighbor model predicts (Table 5), and this may be due to a base triple and an intraloop GA pair (*15*). When the data of Hertel et al. on hammerheads (*57*) are used to calculate $\Delta G^\circ_{37\ \mathrm{MBL\ init}}$, the average value is 1.3 kcal/mol more favorable than expected from eq 11 if $\Delta H^\circ_{\mathrm{MBL\ init}}$ is 0 kcal/mol. Conformational studies of naturally occurring multibranch loops indicate that they often prefer a single three-dimensional orientation of the helical branches (*58*−*64*). These results suggest that the intricate structures of natural multibranch loops may indicate an evolutionary pressure to find stable conformations. The characterization of important canonical motifs in three-dimensional RNA structures suggests that the expanding database may allow for the identification of stabilizing motifs in multibranch loops (*65*). For example, the 5S rRNA motif mentioned previously can also be formed by a three-way multibranch loop found in the secondary structure of an R2 retrotransposon from *Drosophila yakuba* (*66, 67*). Rules for such special motifs could be added to the rules reported here for generic multibranch loops in order to improve the prediction of RNA secondary structure.

Application to RNA Secondary Structure Prediction. Most RNA secondary structure prediction algorithms minimize free energy (*6*−*9*). Previous additions to the nearest-neighbor free-energy parameters have improved the accuracy of secondary structure prediction to about 73% of canonical base pairs in a database of 151 503 nucleotides in 955 known secondary structures (*6*). The equation and parameters for multibranch loops reported here are the first nearest-neighbor parameters for multibranch loop stability with an experimental basis. They represent an important step toward understanding multibranch loop stability and, therefore, should improve the accuracy of secondary structure prediction.

ACKNOWLEDGMENT

The authors thank Thomas W. Barnes III and Joshua M. Diamond for helpful discussions.

REFERENCES

1. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res. 26*, 148−153.
2. Szymanski, M., Specht, T., Barciszewska, M. Z., Barciszewski, J., and Erdmann, V. A. (1998) *Nucleic Acids Res. 26*, 156−159.
3. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
4. Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3502−3507.
5. Walter, N. G., Burke, J. M., and Millar, D. P. (1999) *Nat. Struct. Biol. 6*, 544−549.
6. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
7. Rivas, E., and Eddy, S. R. (1999) *J. Mol. Biol. 285*, 2053−2068.
8. Gultyaev, A. P., van Batenburg, F. H. D., and Pleij, C. W. A. (1995) *J. Mol. Biol. 250*, 37−51.
9. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) *Monatsh. Chem. 125*, 167−168.
10. Ding, Y., and Lawrence, C. E. (1999) *Comput. Chem. 23*, 387−400.
11. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.
12. Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Müller, P., Mathews, D. H., and Zuker, M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.
13. Zuker, M., and Sankoff, D. (1984) *Bull. Math. Biol. 46*, 591−621.
14. Zuker, M. (1989) *Science 244*, 48−52.
15. Diamond, J. M., Turner, D. H., and Mathews, D. H. (2001) *Biochemistry 40*, 6971−6981.
16. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) pp 589, CRC Press, Boca Raton, FL.
17. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) pp 197, CRC Press, Boca Raton, FL.
18. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
19. Petersheim, M., and Turner, D. H. (1983) *Biochemistry 22*, 256−268.
20. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., D'Auberton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr., Brody, E. N., and Gold, L. (1988) *Proc. Natl. Acad. Sci. U.S.A 85*, 1364−1368.
21. Antao, V. P., and Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 819−824.
22. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) pp 233−264, Cold Spring Harbor Laboratory Press, Plainview, NY.
23. Kim, J., Walter, A. E., and Turner, D. H. (1996) *Biochemistry 35*, 13753−13761.
24. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.
25. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
26. Turner, D. H. (2000) in *Nucleic Acids* (Bloomfield, V. A., Crothers, D. M., and Tinoco, I., Jr., Eds.) pp 259−334, University Science Books, Sausalito, CA.

27. Chalikian, T. V., Völker, J., Plum, G. E., and Breslauer, K. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 7853−7858.
28. Holbrook, J. A., Capp, M. W., Saecker, R. M., and Record, M. T., Jr. (1999) *Biochemistry 38*, 8409−8422.
29. Ross, P. D., and Scruggs, R. L. (1965) *Biopolymers 3*, 491−496.
30. Krakauer, H., and Sturtevant, J. M. (1968) *Biopolymers 6*, 491−512.
31. Neumann, E., and Ackerman, T. (1967) *J. Phys. Chem. 73*, 2170−2178.
32. Suurkuusk, J., Alvarez, J., Freire, E., and Biltonen, R. (1977) *Biopolymers 16*, 2641−2652.
33. Filimonov, V. V., and Privalov, P. L. (1978) *J. Mol. Biol. 122*, 465−470.
34. Rawitscher, M. A., Ross, P. D., and Sturtevant, J. M. (1963) *J. Am. Chem. Soc. 85*, 1915−1918.
35. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) *J. Mol. Biol. 256*, 701−719.
36. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
37. Michel, F., Umesono, K., and Ozeki, H. (1989) *Gene 82*, 5−30.
38. Brown, J. W. (1998) *Nucleic Acids Res. 26*, 351−352.
39. Larsen, N., Samuelsson, T., and Zwieb, C. (1998) *Nucleic Acids Res. 26*, 177−178.
40. Chen, J. L., Blasco, M. A., and Greider, C. W. (2000) *Cell 100*, 503−514.
41. Romero, D. P., and Blackburn, E. H. (1991) *Cell 67*, 343−353.
42. Zwieb, C., and Wower, J. (2000) *Nucleic Acids Res. 28*, 169−170.
43. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) *Biochemistry 30*, 6428−6436.
44. Schroeder, S. J., and Turner, D. H. (2000) *Biochemistry 38*, 9257−9274.
45. Schroeder, S. J., Burkard, M. E., and Turner, D. H. (2001) *Biopolymers 52*, 157−167.
46. Jacobson, H., and Stockmayer, W. H. (1950) *J. Chem. Phys. 18*, 1600−1606.
47. Leontis, N. B., Kwok, W., and Newman, J. S. (1991) *Nucleic Acids Res. 19*, 759−766.
48. Kadrmas, J. L., Ravin, A. J., and Leontis, N. B. (1995) *Nucleic Acids Res. 23*, 2212−2222.
49. Pörschke, D., Uhlenbeck, O. C., and Martin, F. H. (1973) *Biopolymers 12*, 1313−1335.
50. Appleby, D. W., and Kallenbach, N. R. (1973) *Biopolymers 12*, 2093−2120.
51. Freier, S. M., Hill, K. D., Dewey, T. G., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) *Biochemistry 20*, 1419−1426.
52. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature 407*, 327−339.
53. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905−920.
54. Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., and Yonath, A. (2000) *Cell 102*, 615.
55. Conn, G. L., Draper, D. E., Lattman, E. E., and Gittis, A. G. (1999) *Science 284*, 1171−1174.
56. Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D., and Williamson, J. R. (2000) *Science 288*, 107−113.
57. Hertel, K. J., Stage-Zimmermann, T. K., Ammons, G., and Uhlenbeck, O. C. (1998) *Biochemistry 37*, 16983−16988.
58. Bassi, G. S., Møllegaard, N. E., Murchie, A. I. H., and Lilley, D. M. J. (1999) *Biochemistry 38*, 3345−3354.
59. Thompson, J. B., and Lilley, D. M. J. (1999) *RNA 5*, 180−187.
60. Zhao, Z., Wilson, T. J., Maxwell, K., and Lilley, D. M. J. (2000) *RNA 6*, 1833−1846.
61. Frazer-Abel, A. A., and Hagerman, P. J. (1999) *J. Mol. Biol. 285*, 581−591.
62. Friederich, M. W., Vacano, E., and Hagerman, P. J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 3572−3577.
63. Amiri, K. M. A., and Hagerman, P. J. (1996) *J. Mol. Biol. 261*, 125−134.
64. Shen, Z., and Hagerman, P. J. (1994) *J. Mol. Biol. 241*, 415−430.
65. Leontis, N. B., and Westhof, E. (1998) *J. Mol. Biol. 283*, 571−583.
66. Mathews, D. H., Banerjee, A. R., Luan, D. D., Eickbush, T. H., and Turner, D. H. (1997) *RNA 3*, 1−16.
67. Xia, T., Mathews, D. H., and Turner, D. H. (1999) in *Prebiotic Chemistry, Molecular Fossils, Nucleosides, and RNA* (Söll, D. G., Nishimura, S., and Moore, a. P. B., Eds.) pp 21−47, Elsevier, New York.
68. Burkard, M. E., Kierzek, R., and Turner, D. H. (1999) *J. Mol. Biol. 290*, 967−982.
69. Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992) *Numerical Recipes in C*, Cambridge University Press, New York.

BI011441D