Thermodynamics of Three-Way Multibranch Loops in RNA†

Joshua M. Diamond, Douglas H. Turner,* and David H. Mathews

Department of Chemistry, University of Rochester, RC Box 270216, Rochester, New York 14627-0216

Received December 28, 2000; Revised Manuscript Received April 13, 2001

ABSTRACT: RNA multibranch loops (junctions) are loops from which three or more helices exit. They are nearly ubiquitous in RNA secondary structures determined by comparative sequence analysis. In this study, systems in which two strands combine to form three-way junctions were used to measure the stabilities of RNA multibranch loops by UV optical melting and isothermal titration calorimetry (ITC). These data were used to calculate the free energy increment for initiation of a three-way junction on the basis of a nearest neighbor model for secondary structure stability. Imino proton NMR spectra were also measured for two systems and are consistent with the hypothesized helical structures. Incorporation of the experimental data into the *mfold* and RNAstructure computer programs has contributed to an improvement in prediction of RNA secondary structure from sequence.

Multibranch loops (junctions) are nucleic acid loops from which three or more helices exit. For example, an important multibranch loop is the four-helix Holliday junction in DNA (*1*−*4*). Several studies have focused on the factors that affect the stabilities of DNA multibranch loops. For example, it has been shown that two or more unpaired nucleotides between helices in a DNA three-way junction stabilize the structure (*5*). Relative stabilities of three-, four-, and five-way DNA junctions have been defined on the basis of UV optical melting and competitive binding experiments (*5*, *6*). The stability of a DNA three-way junction was quantified with isothermal titration calorimetry (*7*). Studies of both DNA and RNA have also provided information on the salt dependence of multibranch loop formation and conformation (*8*−*12*). Transient electric birefringence (TEB), fluorescence resonance energy transfer (FRET), X-ray crystallography, and NMR[1] have revealed the relative positioning of junction helices (*3*, *9*−*20*). The stability of the RNA multibranch loop at the catalytic core of the hammerhead ribozyme was determined by measuring cleavage kinetics (*21*). The factors determining stability of RNA junctions, however, have not been elucidated, even though such junctions are nearly ubiquitous in RNA secondary structures determined by comparative sequence analysis (*22*−*31*).

In this study, optical melting and isothermal titration calorimetry were used to determine the stabilities of three-way RNA junctions. The design of the three-way junction was based upon the structure and sequence of 5S rRNA (*27*). While most studies of DNA multibranch loops utilized a three-stranded system (*7*, *8*), two strands were employed for this study to simplify the analysis (*32*) and limit the number of possible competing structures. The data demonstrate that prior models for predicting the free energies of RNA multibranch loops are inadequate. Therefore, preliminary data from this study were used as a starting point for choosing new multibranch loop free energy parameters for the prediction of RNA secondary structure by free energy minimization (*33*). The results indicate that insights into the determinants of junction stability improve predictions of secondary structure.

MATERIALS AND METHODS

Oligonucleotide Synthesis and Purification. RNA oligonucleotides were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer with phosphoramidites from Glen Research. The 2′-hydroxyl was protected as a *tert*-butyl dimethylsilyl ether and the 5′-hydroxyl as a dimethoxytrityl (*34*−*37*). Upon completion of synthesis, oligomers were cleaved from the solid support and partially deprotected by addition of 1.5 mL of concentrated ammonia and 0.5 mL of 95% ethanol followed by incubation at 55 °C for 12 h (*38*). The CPG solid support was removed with a spin filter (Quik-Sep Columns, Perkin-Elmer Wallac Inc., Norton, OH), and the vial and sample were washed once with 2 mL of water, once with 4 mL of 1:1 water:acteonitrile, and a final time with 2 mL of acetonitrile to recover the maximum amount of oligonucleotide. The 2′ protecting group was cleaved by incubation at 55 °C for 48 h in 1 M triethylamine−hydrogen fluoride (TEAHF) in pyridine.

For oligomers shorter than 14 nucleotides, organic salts were removed by four extractions with 5 mL of ether each. Waters C-18 Sep-pak cartridges were used to remove inorganic salts from the oligomers. The samples were then purified on a Baker Si500f thin-layer chromatography plate with a solvent of 1-propanol:ammonia:water (55:35:10) (*39*). The product band was visualized under UV light and scraped

† This work was supported by NIH Grant GM22939.

* To whom correspondence should be addressed. Phone: (716) 275-3207. Fax: (716) 473-6889. E-mail: turner@chem.rochester.edu.

[1] Abbreviations: CPG, controlled pore glass; eu, entropic units (calories per kelvin per mole); EDTA, ethylenediaminetetraacetic acid; HPLC, high-pressure liquid chromatography; T_m, melting temperature in degrees Celsius; T_M, melting temperature in kelvin; NMR, nuclear magnetic resonance; TLC, thin-layer chromatography; C_T, total strand concentration.

10.1021/bi0029548 CCC: $20.00
Published on Web 05/18/2001

from the plate. Oligonucleotides were eluted from the silica gel by washing with 4 mL of sterile water four times, for a total of 16 mL, and again desalted by Sep-pak.

For RNA oligomers longer than 14 nucleotides, the samples were loaded onto a 20% polyacrylamide denaturing gel to separate the product from shorter strands. For each sample, the least mobile band was visualized by UV shadowing, excised, and placed in water. The polyacrylamide and water were then shaken at room temperature for 48 h. The samples were spin filtered in Quik-Sep columns (Perkin-Elmer Wallac) to separate the gel from the water. The samples were separated into 2.5 mL fractions and passed twice through prepacked G-25M Sephadex PD-10 columns (Pharmacia Biotech) to remove salts and urea.

The purity of the oligonucleotides of less than 14 nucleotides was found to be greater than 95% by HPLC (Hewlett-Packard Series 1100) on a C-18 column (Hamilton) with a triethylamine acetate and acetonitrile buffer system. The purity of all strands longer than 14 nucleotides was determined by gel electrophoresis to be greater than 95%. For these assays, 200 pmol of sample was ^{32}P 5′-end labeled with T4 polynucleotide kinase (Gibco) and run on a 20% polyacrylamide denaturing gel. The purity was quantified with a Molecular Dynamics phosphorimager and ImageQuant NT software (Molecular Dynamics).

T7 RNA Polymerase Purification. T7 RNA polymerase used in all transcription reactions was purified by a 6-His tagged preparation from DH5α *Escherichia coli* cells transformed with pT7-911Q (*40*). A 2 L culture (0.1 mg/mL ampicillin from Sigma) of transformed *E. coli* cells with A^{600} of 0.5 was incubated with 5 mM isopropyl β-D-thiogalactopyranoside (IPTG from Sigma) for 3 h. Cells were harvested by centrifugation and lysed by sonication with a Branson 1210 sonifier. The 6-His tagged T7 RNA polymerase was purified from the lysate by batch methods on a Ni−NTA−agarose resin (Qiagen). The tagged protein was bound to the resin with gentle rocking for 30 min, and four washes of the resin were performed, each with a 1 mM imidazole (Aldrich) buffer and then with a 10 mM imidazole buffer. Finally, the tagged protein was eluted from the resin with 25 mL of 100 mM imidazole, and the T7 RNA polymerase was concentrated in a Centriprep-30 concentrator (Centricon). The stock polymerase was stored in a 25% glycerol solution at −20 °C.

DNA Template Synthesis. DNA template for T7 RNA polymerase transcription was synthesized on a 1 μmol scale on an Applied Biosystems 392 DNA/RNA synthesizer. The oligomers were synthesized with the dimethoxytrityl (DMT) protecting group remaining on the 5′ end, thus allowing for purification by Poly Pak II (Glen Research) (*41*). The newly synthesized oligomers were cleaved from the solid support and deprotected by incubation in ammonium hydroxide at 55 °C for 12−24 h. This mixture was then passed through Quik-Sep spin filters (Perkin-Elmer Wallac) to remove the CPG beads. Excess ammonia was removed by drying for 5 min in a speed vac. DNA was purified and DMT deblocked on Poly Pak II columns (Glen Research). After elution with 1.5 mL of 20% acetonitrile in water, the acetonitrile was removed by drying until the volume was reduced by half.

T7 RNA Polymerase Transcription. The transcription mixture contained 0.04 M Tris buffer, pH 8, 300 nM DNA template, 6 mM NTP mixture at pH 8, 5 mM DTT, 5 mM spermidine, 7.9 mL of the T7 RNA polymerase preparation, 0.3 μg/μL bovine serum albumin (Gibco), and 20 mM $MgCl_2$ in a final volume of 82.8 mL (*42*, *43*). The mixtures were incubated at 37 °C for 1 h. Reactions were stopped with 10 mL of 0.5 M EDTA. Proteins were removed by phenol:chloroform extraction. The samples were then ethanol precipitated overnight, centrifuged for 15 min at ∼6000*g*, resuspended, and purified on a 20% polyacrylamide denaturing gel as described above for chemically synthesized RNA strands.

RNA Concentration Determination. RNA strand concentrations were determined by UV absorbance using extinction coefficients determined with a nearest neighbor model (*44*, *45*). Absorbances were recorded at temperatures greater than 80 °C. Accurate oligomer concentrations are especially important for ITC experiments; therefore, the concentration for ITC experiments was taken as the average of concentrations determined at 260 and 280 nm, which differed from each other by an average of 15%.

Preparation of Samples for UV Melting Experiments in 1 M Na^+. For each system, equal concentrations of the two strands were mixed. They were then evaporated to dryness and reconstituted in 1 M Na^+ buffer consisting of 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA adjusted to pH 7 with NaOH. Melts of the hairpin strands alone were performed in a buffer of 20 mM sodium cacodylate and 0.5 mM EDTA adjusted to pH 7 with NaOH. To prevent the formation of kinetic traps, all samples were heated to greater than 80 °C before melting experiments were run.

Preparation of Samples for UV Melting Experiments in 10 mM Mg^{2+}. A 10× stock of 0.1 M $MgCl_2$ solution was prepared. Similarly, a 10× stock buffer was prepared of cacodylic acid (0.2 M) and potassium chloride (1.5 M), adjusted to pH 7 with concentrated KOH. For each sample melted, equal concentrations of the two strands were mixed. The mixture was then evaporated to 80% of the final melt volume. Stock cacodylic acid and KCl buffer equaling 10% of the final melt volume was added to the mixture. For annealing, the mixture was placed in a 90 °C heat block for 10 min, and the heat block and sample were allowed to slowly cool for approximately 45 min. When the sample reached 30 °C, an aliquot of 0.1 M $MgCl_2$ equaling 10% of the final melt volume was added to the mixture.

UV Optical Melting. Samples were melted over a 9−88-fold range of strand concentration. Data were obtained with a Gilford Instruments 250 spectrophotometer at a wavelength of 280 nm at a heating rate of 1 °C/min controlled by a Gilford Instruments thermoprogrammer, 2527. Data were analyzed by fitting the lower temperature transition to a two-state model with sloping baselines using a nonlinear least squares program and by T_M^{-1} vs $\ln(C_T/4)$ plots (*32*, *46*):

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

where R is the gas constant and C_T is total strand concentration.

To account for the effects of heat capacity changes, ΔC_p°, the fitted values of ΔH° and ΔS° from each curve were plotted as a function of T_M and $\ln(T_M)$, respectively. The slopes of these plots provide values of ΔC_p° and allow

calculation of $\Delta H°$ and $\Delta S°$ at any temperature according to

$$\Delta H°(T_M) = \Delta H°_T + \Delta C_p°(T_M - T) \quad (2)$$

$$\Delta S°(T_M) = \Delta S°_T + \Delta C_p° \ln(T_M/T) \quad (3)$$

Here T is a reference temperature in kelvin.

Isothermal Titration Calorimetry. Experiments were performed on a MicroCal OMEGA ultrasensitive isothermal titration calorimeter (MicroCal, Inc.) (*47*). The oligomers were in 1 M Na^+ melt buffer. A 1.4 mL volume of sample was injected into the sample cell, and a mixing speed of 300 rpm was used. Data were analyzed with ORIGIN software, version 5 (Microcal, Inc.). All ITC experiments were performed with the longer strand initially in the cell and the shorter, complement strand titrated in. As a control to account for heats of dilution and any temperature difference between the injected strand and the cell, the short strand was also titrated into 1 M Na^+ melt buffer. The control data were subtracted from the experimental data after integration of the peaks.

To determine the conditions for isothermal titration calorimetry, two competing requirements are considered. First, there must be enough of each strand to minimize the error in the experimental data, and, second, the temperature of the experiment must be chosen so that the amount of each strand is low enough to make synthesis manageable and affordable. The optimum total heat, determined by trial and error, is 4×10^{-3} cal, and therefore the concentration of strand placed in the cell, M_{tot}, is

$$M_{tot} \sim \frac{4 \times 10^{-3}\ \text{cal}}{\Delta H°(1.4 \times 10^{-3}\ \text{L})} \quad (4)$$

Here $\Delta H°$ is the enthalpy change derived from the T_M^{-1} vs $\ln(C_T/4)$ analysis of optical melting experiments, and 1.4×10^{-3} L is the sample volume. The concentration of oligomer in the syringe was about 20 times that of the oligomer in the sample cell. Once M_{tot} is determined, the temperature is chosen such that the equilibrium constant can be fit to the data. According to Wiseman et al. (*47*), the curve fits are most accurate when $50 \leq C \leq 500$, for C determined by

$$C = K_{eq} M_{tot} \quad (5)$$

where K_{eq} is the equilibrium constant. The best results were obtained with $C = 500$.

The temperature is then chosen on the basis of the enthalpy and entropy changes determined by optical melting and a rearrangement of $-RT \ln(K_{eq}) = \Delta H° - T\Delta S°$:

$$T = \Delta H°/[\Delta S° - R \ln(K_{eq})] \quad (6)$$

The ITC was calibrated by Y axis deflection, an internal calibration procedure in which the size of the deflection from baseline is measured. The calibration was checked by measuring the heat of dilution of 0.2 M NaCl (*48*).

Cleaning of the Isothermal Titration Calorimeter. The cell and all syringes were thoroughly cleaned to remove residual proteins and contaminants, specifically RNase. The sample cell was first rinsed by allowing roughly 400 mL of a 60 °C, 5% solution of Contrad-70 (Fisher Scientific) to flow through the system. This was followed by a wash with greater than 1 L of sterile water. The cell was filled with RNase Zap (Ambion) and then flushed with more than 3 L of distilled water. All syringes were flushed with RNase Zap followed by three rinses with distilled water.

NMR Spectroscopy. After purification, 250 nmol of the longer strand were dried and resuspended in 250 μL of buffer containing 100 mM NaCl, 10 mM sodium phosphate (pH 7.5), and 0.1 mM EDTA in 90% H_2O and 10% D_2O. Spectra were obtained, and then 250 nmol of the shorter strand was added in 50 μL of water. Spectra were recorded on a Varian INOVA 500 MHz spectrometer using a 1:3:3:1 binomial pulse sequence to suppress water (*49*). The frequency offset was set to maximize intensity at about 13 ppm with first nodes at 21 and 5 ppm. A total of 12 000 points were collected over a sweep width of 12 kHz. Free induction decays were multiplied with a 4 Hz line-broadening function and Fourier transformed with Varian VNMR software. The spectra were referenced by setting the residual water peak to its known chemical shift.

RESULTS

Design of the System. The two-strand system used in this study was designed to form a junction in an intermolecular reaction with a melting temperature much lower than that associated with hairpin formation in the longer strand (Figure 1). To provide a hairpin component that melts at a much higher temperature than the junction, a stable tetraloop, 5′UUCG3′ (*50*), was closed by stems that ensure the hairpin melts at temperatures greater than 70 °C. This allows the intermolecular transition to be studied independently of the melting of the hairpin (Figure 1). Design of the sequences was greatly facilitated by secondary structure prediction using RNAstructure (*33*) to identify sequences forming the desired structures without having any predicted suboptimal structure within 2 kcal/mol at 37 °C.

The system was also designed to allow measurements on many sequence variations (Figure 2). First, with the exception of system C, all of the top and bottom strands are interchangeable because they rely on formation of the same base pairs to form a junction. The interchangeability allows numerous perturbations in junction sequence while minimizing synthesis. For example, system F in Figure 2 is composed of the long strand from system D (D1) and the short strand from system B (B2). The design thus allows study of the effects of changing the number and type of unpaired nucleotides in the junction.

The 5′ end of the top strands begins with GGA, a sequence that is transcribed with high efficiency by T7 RNA polymerase (*43*). This allows production of large amounts of the long top strands for isothermal titration calorimetry and NMR experiments.

A search of the three-way junctions in a 5S rRNA database (*27*) revealed the most prevalent unpaired nucleotides and closing base pairs. These data were used to design structure D in Figure 2, which mimics the consensus structure determined by comparative sequence analysis. Previous comparative analysis of the 5S rRNA sequences in the Berlin RNA Database revealed that the sequence 5′CAUA3′ contained in structure D is highly conserved and likely plays a prominent role in 5S rRNA function (*15*, *51*). The exact

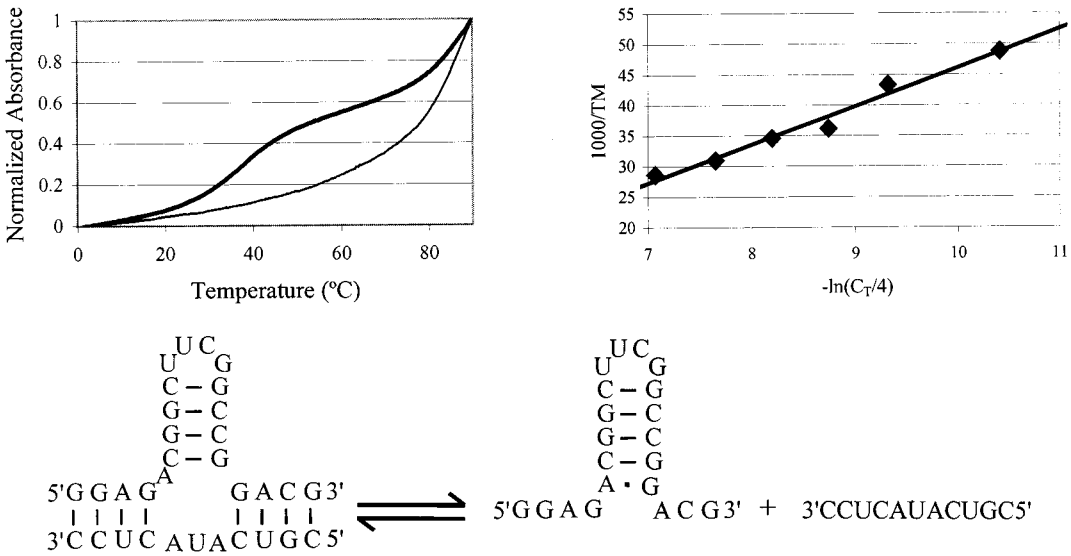

FIGURE 1: (Top left) Typical melting curve for system B. The heavy line is the melting curve of the multibranch loop system while the thin line is the melting curve of the longer strand (B1) alone. (Top right) Linear fit of T_m^{-1} vs $-\ln(C_T/4)$ for system B. (Bottom) Lower temperature equilibrium for system B. Reported thermodynamics are for the reverse reaction.

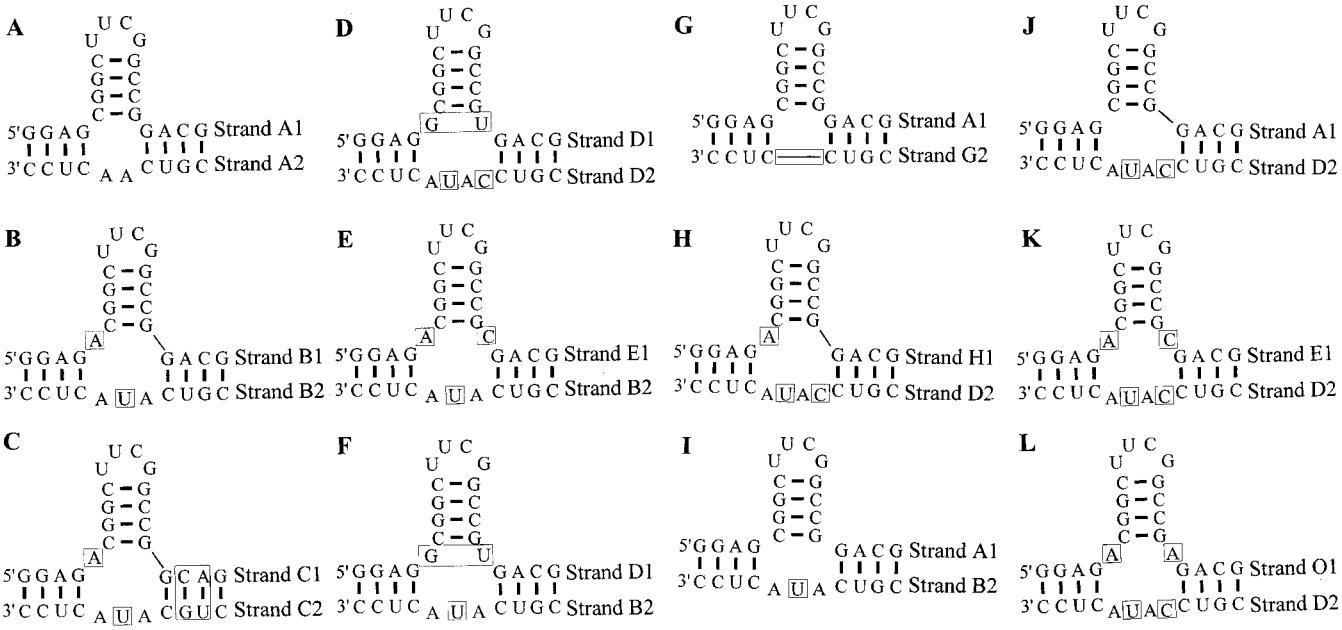

FIGURE 2: The multibranch loop systems studied. Differences from system A are boxed. System D is the structure most closely related to 5S rRNA. The interchangeability between the top and bottom strands allowed for many combinations. For example, system F is composed of the top strand of system D and the bottom strand of system B.

junction sequence in structure D, including closing base pairs, is found in the 5S rRNA of several important organisms, including *Homo sapiens*, *Pneumocystis carinii*, and *Trypanosoma brucei* (*27*).

Thermodynamics from Optical Melting. Single strand melts of the hairpin-containing strands, A1 and B1, revealed no self-structure, other than the hairpin which melted at temperatures above 70 °C (Figure 1). Thus the lower temperature transition only reflects formation of the duplex and therefore the multibranch loop stability. At temperatures below the beginning of the hairpin transition, the absorbance vs temperature data were fit to a two-state model to provide thermodynamic parameters for the intermolecular association forming the junction. For all systems, the transitions are well enough separated in temperature to allow this analysis. The results are listed in Table 1.

While the values in Table 1 are derived on the basis of a two-state model, most sequences are clearly not two state. Deviations from the two-state model are of different types. First, the difference in enthalpy determined by the average of the melt curve fits and linear plots of T_M^{-1} vs $\ln(C_T/4)$ is often much greater than 15% (Table 1), which is considered a signature of non-two-state behavior (*52*). Second, first derivative plots of the individual melt curves occasionally indicate two transitions occurring over a small temperature range as is the case for system H (Figure 3). System D also revealed two close transitions. Imino proton spectra of system D over a range of temperatures support the hypothesis that the two intermolecular stems are melting at different temperatures (see below).

The analysis that provides the values in Table 1 assumes that $\Delta H°$ and $\Delta S°$ are independent of temperature, i.e., $\Delta C_p°$

Table 1: Thermodynamic Parameters Determined by Optical Melting of Three-Way Junctions[a]

	$1/T_M$ vs $\ln(C_T/4)$				average of melt curve fits				
system	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_m[b] (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_m[b] (°C)	$\Delta H°$[c] % diff
A	5.42 ± 0.02	40.43 ± 0.88	112.88 ± 2.87	28.7	5.45 ± 0.18	49.08 ± 9.51	140.68 ± 30.47	30.3	19.3
B	4.05 ± 0.11	40.16 ± 3.15	116.42 ± 10.45	19.0	3.94 ± 0.09	46.28 ± 5.88	136.51 ± 19.09	20.5	13.2
C	5.85 ± 0.08	61.70 ± 4.55	180.08 ± 14.39	33.6	5.87 ± 0.14	58.64 ± 4.04	170.15 ± 12.84	33.6	5.0
D	6.00 ± 0.01	52.62 ± 0.97	150.29 ± 3.09	33.9	6.47 ± 0.69	88.93 ± 9.54	265.86 ± 28.72	36.0	39.2
E	4.79 ± 0.15	36.92 ± 3.53	103.62 ± 11.77	23.0	4.51 ± 0.16	47.65 ± 8.69	139.08 ± 28.33	24.4	22.5
F	6.18 ± 0.01	52.66 ± 1.15	149.88 ± 3.68	34.9	6.50 ± 0.54	74.51 ± 10.38	219.31 ± 31.84	36.9	25.3
G	5.72 ± 0.14	54.10 ± 5.99	156.01 ± 19.10	32.4	5.77 ± 0.15	69.09 ± 10.43	204.13 ± 33.77	33.6	21.7
H	4.58 ± 0.11	26.92 ± 1.88	72.01 ± 6.33	16.1	4.31 ± 0.29	35.74 ± 6.44	101.32 ± 20.30	18.9	24.7
I	5.43 ± 0.01	41.58 ± 1.12	116.57 ± 3.61	29.0	5.50 ± 0.50	67.60 ± 8.70	200.22 ± 26.59	32.3	38.5
J	5.92 ± 0.20	39.65 ± 4.97	108.75 ± 15.88	32.3	6.55 ± 1.46	76.40 ± 19.91	225.24 ± 60.33	37.1	48.1
K	5.25 ± 0.04	27.87 ± 1.16	72.93 ± 3.82	23.4	4.87 ± 1.94	79.53 ± 20.44	240.72 ± 59.73	30.7	65.0
L	5.91 ± 0.004	53.62 ± 0.80	153.82 ± 2.56	33.5	6.05 ± 0.70	90.99 ± 3.70	273.87 ± 11.18	35.4	41.1

[a] Buffer was 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. [b] T_m at 10^{-4} M total strand concentration. [c] This column lists the percentage difference between $\Delta H°$ obtained from $1/T_M$ plots and from averaging values from fits of melting curves.

Table 2: Change in Heat Capacity Determined from the Temperature Dependence of $\Delta H°$ and $\Delta S°$

	$-\Delta C_p°$ [kcal/(mol K)]		$-\Delta H°_{37}$		$-\Delta G°_{37}$
system	from $\Delta H°$	from $\Delta S°$	(kcal/mol)	$-\Delta S°_{37}$ (eu)	(kcal/mol)
A	1.41 ± 0.09	1.38 ± 0.09	56.91 ± 0.74	165.8 ± 2.4	5.47 ± 0.01
B	0.86 ± 0.33	0.83 ± 0.32	54.08 ± 3.42	161.5 ± 11.0	4.00 ± 0.05
C	0.27 ± 0.37	0.26 ± 0.36	58.95 ± 1.53	171.1 ± 4.9	5.88 ± 0.06
D	1.60 ± 0.47	1.47 ± 0.47	83.77 ± 2.65	250.8 ± 8.4	5.98 ± 0.09
E	1.99 ± 0.13	1.95 ± 0.12	68.48 ± 1.50	206.1 ± 4.5	4.56 ± 0.09
F	1.59 ± 0.28	1.52 ± 0.27	70.72 ± 1.68	207.9 ± 5.3	6.23 ± 0.05
G	−2.40 ± 0.63	−2.46 ± 0.62	76.83 ± 3.00	229.5 ± 9.4	5.65 ± 0.09
H	0.34 ± 0.20	0.32 ± 0.20	40.06 ± 3.18	114.7 ± 10.4	4.49 ± 0.16
I	1.31 ± 0.19	1.23 ± 0.18	68.62 ± 1.11	203.6 ± 3.5	5.49 ± 0.03
J	1.93 ± 0.51	1.80 ± 0.51	72.48 ± 4.32	214.1 ± 13.6	6.08 ± 0.21
K	1.82 ± 0.28	1.62 ± 0.26	89.13 ± 3.14	269.3 ± 9.7	5.62 ± 0.15
L	0.30 ± 0.23	0.17 ± 0.23	90.69 ± 1.20	273.4 ± 3.9	5.91 ± 0.22

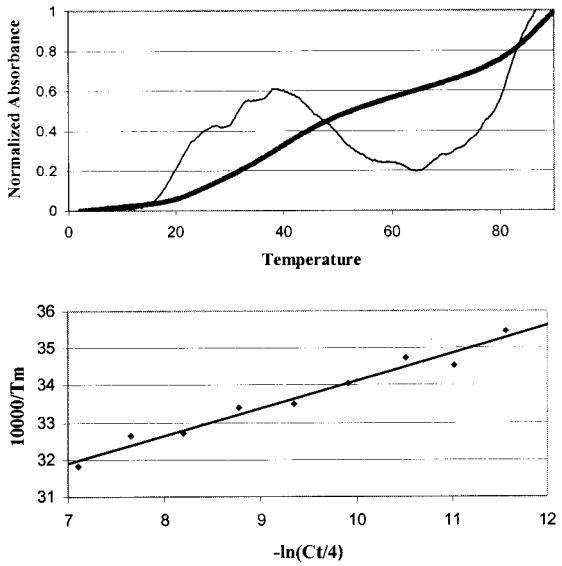

FIGURE 3: (Top) A typical melting curve for system H. The dark line is the melting curve, while the thin line is the first derivative plot of the melting curve. (Bottom) Linear fit of T_m^{-1} vs $-\ln(C_T/4)$ for system H.

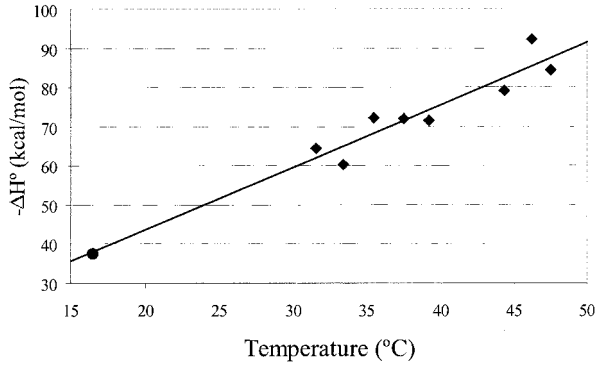

FIGURE 4: Plot of $-\Delta H°$ vs temperature based on optical melting data for system F (diamonds). The slope is $\Delta C_p°$. The $\Delta H°$ determined by isothermal titration calorimetry is also shown (circle) and is in close agreement with the optical melting data.

is zero. To allow comparisons with isothermal calorimetry, the dependencies on T_M of $\Delta H°$ and $\Delta S°$ obtained from fits of individual melting curves were used to estimate $\Delta C_p°$ (*46*). In particular, the slopes of the linear fits to plots of $\Delta H°$ vs T_M and of $\Delta S°$ vs $\ln(T_M)$ gave separate estimates of $\Delta C_p°$ (Figure 4 and Table 2). Except for systems K and L, both values of $\Delta C_p°$ were within 8% of each other. The error limits on $\Delta C_p°$ are large. Nevertheless, most values are similar to expectations from $\Delta C_p°$'s for base-paired duplexes (*46*, *53*, *54*). The clearest exception is system G, which is the only system with a positive $\Delta C_p°$. System G is also the only multibranch loop without unpaired nucleotides. Values for $\Delta H°_{37}$ and $\Delta S°_{37}$ were also obtained from the plots. Values of $\Delta G°_{37}$ calculated from $\Delta H°_{37}$ and $\Delta S°_{37}$ are similar to those derived assuming $\Delta C_p°$ is zero (see Tables 1 and 2).

Effect of Mg^{2+}. Previous investigations of the salt dependence of DNA multibranch loops have shown that magnesium has a profound effect on the conformation of the helices

Table 3: Thermodynamic Parameters from Isothermal Titration Calorimetry and from Optical Melting[a]

			ITC results[b]				optical melting: fitted curves, $\Delta C_p^\circ \neq 0$			optical melting: T_M^{-1} vs ln C_T, $\Delta C_p^\circ = 0$	
system	T[c] (°C)	N	$-\Delta S^\circ_T$ (eu)	$-\Delta H^\circ_T$ (kcal/mol)	$-\Delta G^\circ_T$ (kcal/mol)	$-\Delta G^\circ_{37}$ (kcal/mol)	$-\Delta H^\circ_T$ (kcal/mol)	$-\Delta G^\circ_T$ (kcal/mol)	$-\Delta G^\circ_{37}$ (kcal/mol)	$-\Delta G^\circ_T$ (kcal/mol)	$-\Delta G^\circ_{37}$ (kcal/mol)
B	11.5	1.02 ± 0.03	95.1 ± 4.2	33.8 ± 1.2	6.74 ± 0.11	4.30 ± 1.78	32.4 ± 9.9	6.63 ± 0.20	4.00 ± 0.05	7.25 ± 0.18	4.05 ± 0.11
D	15.0	0.99 ± 0.01	101 ± 2.1	37.0 ± 0.6	7.88 ± 0.10	5.67 ± 0.89	48.5 ± 20.4	7.48 ± 0.42	5.98 ± 0.09	9.61 ± 0.08	6.00 ± 0.01
F	16.5	1.08 ± 0.02	102 ± 2.5	37.6 ± 0.7	8.15 ± 0.17	5.96 ± 1.04	38.0 ± 11.9	7.95 ± 0.17	6.23 ± 0.05	9.54 ± 0.09	6.18 ± 0.01

[a] Buffer was 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7. [b] Oligonucleotide concentrations (mM) in cell and syringe, respectively, were B (0.074, 1.48), D (0.054, 1.08), and F (0.052, 0.96). [c] T is the temperature of the ITC experiments and of all parameters labeled with a subscript T.

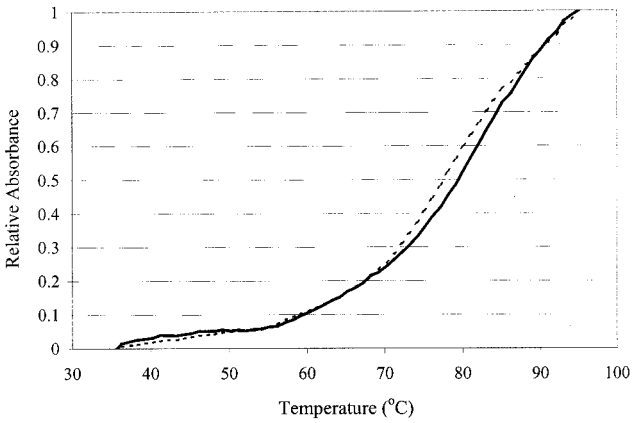

FIGURE 5: Melting curves of strand D1 (dotted line) and 5′GGCGGCUUCGGCCGUG3′ (solid line), the hairpin and stem plus the first mismatch on the stem from strand D1. The buffer for these experiments contained 20 mM sodium cacodylate, adjusted to pH 7 with NaOH, and 0.5 mM Na_2EDTA. The low salt buffer was used so that the transition occurs in an accessible range of temperatures.

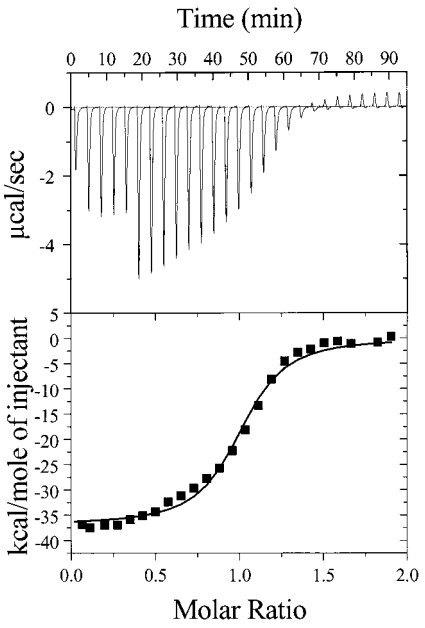

FIGURE 6: Isothermal titration data for system D.

(*8*−*12*). The dependence of stability on cations was tested in melts of systems A, D, and F (Figure 2) in 10 mM $MgCl_2$, 20 mM cacodylic acid, and 150 mM KCl. The Mg^{2+} and K^+ concentrations were chosen to approximate intracellular concentrations of these cations. Changing the salt conditions resulted in less than a 10% change in free energy at 37 °C. Previous studies on short duplexes (*32*, *55*) also showed essentially no difference for stabilities measured in 1 M NaCl and in 10 mM Mg^{2+}.

Effect of 5′-Triphosphate. The hairpin strands used in all ITC experiments were synthesized by T7 RNA polymerase transcription and therefore have a 5′-terminal triphosphate. To test the effect of a 5′-triphosphate on stability, system B with a 5′-triphosphate was melted in 1 M NaCl. The measured ΔG°_{37}'s with and without a 5′-triphosphate are −4.11 and −4.05 kcal/mol, respectively. Thus the 5′-triphosphate has a negligible effect on stability, as expected from previous studies of the effects of 5′-monophosphates on duplex stability (*56*).

Effect of Dangling Ends on the Folding of the Hairpin Strand. To test the effects of the long unpaired dangling ends on the stability of the hairpin strands alone, the sequences 5′ACGGCUUCGGCCGG3′ and 5′GGCGGCUUCGGCCGUG3′, where underlined nucleotides are unpaired, were melted in 20 mM sodium cacodylate buffer to compare with melts for strands B1 and D1 in the same buffer (Figure 5). These experiments were designed to determine whether the dangling ends beyond the first mismatch add stability to the hairpin intermediate after denaturation of the multibranch loop. If so, the free energy of such interactions would have to be subtracted from the measured free energy in calculating free energy increments for junctions. The experiments revealed minimal differences in the stabilities of the hairpin strands with and without the additional dangling nucleotides. In both cases, the strand containing only the first mismatch was minimally more stable (i.e., melted at a slightly higher temperature) than the strand containing the longer dangling ends. The difference in T_M in 20 mM sodium cacodylate was so slight that quantitation was not possible.

Isothermal Titration Calorimetry. Only systems A, B, and C melt in a manner very close to that expected for a two-state system (Table 1). Thus isothermal titration calorimetry (ITC) was used to provide another measure of the thermodynamics of junction formation for systems B, D, and F (Figure 6).

ITC permits measurement at a single temperature of ΔH°, ΔG°, and N, the molecularity of the reaction. The theoretical value for N is 1 because it is assumed that for every RNA molecule titrated into the cell only one molecule of complementary strand will bind. As shown in Table 3, this stoichiometry was confirmed for all three systems; the average value of N was 1.03 ± 0.02.

Free energy changes measured by UV melting are most accurate near the melting temperature, where half of the

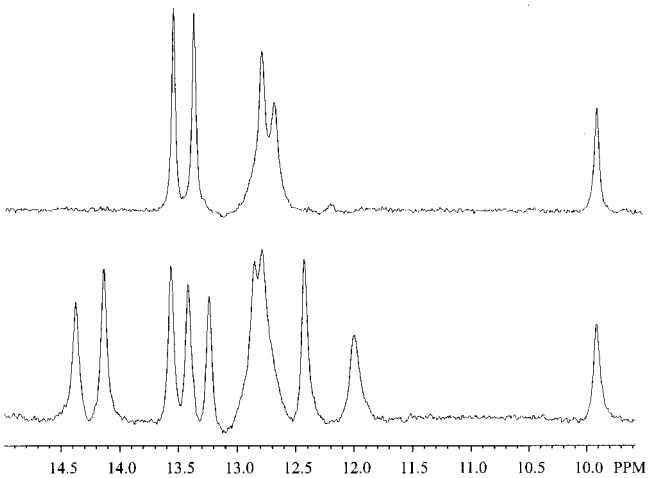

FIGURE 7: (Top) Imino proton spectrum of strand B1 at 0 °C. (Bottom) Imino proton spectrum of system B at 0 °C. The resonances at 14.1 and 14.4 ppm indicate the presence of two A-U pairs as predicted on the basis of free energy minimization.

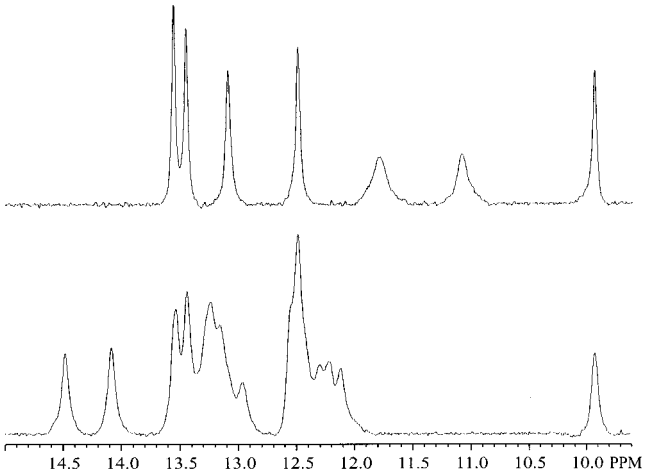

FIGURE 8: (Top) Imino proton spectrum of strand D1 at 10 °C. The two resonances between 11 and 12 ppm are consistent with formation of a G·U pair at the base of the hairpin stem. (Bottom) Imino proton spectrum of system D at 10 °C. The resonances at 14.1 and 14.5 ppm indicate the presence of two A-U pairs as predicted on the basis of free energy minimization. The absence of resonances between 11 and 12 ppm suggests no G·U pair is formed.

strands are in duplex. Thus the sequences in Figure 2 were chosen to allow melting temperatures near 37 °C, the most common temperature for predictions of RNA secondary structure. ITC measurements, however, must be done at a single temperature where all of the strands are in duplex at the end of the experiment for a given strand concentration. Thus the ITC experiments were done at temperatures ranging from 11.5 to 16.5 °C. Comparison of results (Table 3) shows that the $\Delta G°$ at 11.5 °C measured by ITC for system B, which is two state, agrees within 7% with the value calculated from optical melting results analyzed with a T_M^{-1} vs $\ln(C_T/4)$ plot. For the non-two-state systems, D and F, however, the $\Delta G_T°$'s measured by ITC are less favorable by 18% and 15%, respectively, than those calculated from T_M^{-1} vs $\ln(C_T/4)$ plots. Nevertheless, extrapolation of the ITC results to 37 °C, with the assumption of $\Delta C_p°$ equal to zero, yields $\Delta G°$'s that are less favorable by only 6% and 4%, respectively, than those calculated from T_M^{-1} vs $\ln(C_T/4)$ plots. Moreover, the difference between the $\Delta G°_{37}$ values when calculated from optical melting and ITC results for D and F are only 0.3 and 0.2 kcal/mol, respectively. Evidently, $\Delta G°$ values determined from T_M^{-1} vs $\ln(C_T/4)$ plots are reasonably accurate near the experimental melting temperatures, even when melting is not strictly a two-state process. When $\Delta G°$ values are determined from values of $\Delta H°$ and $\Delta S°$ obtained from fitting individual melting curves and accounting for $\Delta C_p°$, they agree well with ITC values at the temperature of the ITC experiment and extrapolated to 37 °C (Table 3).

The enthalpy changes determined by optical melting can also be compared with those measured at low temperature by ITC when $\Delta C_p°$ is taken into account. As shown in Table 3, the $\Delta H°$'s from optical melting are within experimental error of those measured by ITC, but the experimental error is large.

NMR Spectra. Imino proton NMR spectra of strand B1 and system B (Figure 7) are consistent with the hypothesized structures (Figure 2). The resonance at 9.9 ppm in both spectra is assigned to G13, consistent with the chemical shift found by Varani et al. for the imino proton of the G in a UUCG tetraloop closed by a C-G pair (*57*). The emergence of the peaks at 14.1 and 14.4 ppm in the bimolecular structure is consistent with formation of the two A-U pairs in the stems closing the multibranch loop.

Imino proton NMR spectra were also obtained for strand D1 and system D (Figure 8). Upon titration of strand D2 into strand D1, peaks at 14.1 and 14.5 ppm appeared, consistent with the expected A-U pairs in the stems forming the junction (Figure 2). Predicted coaxial stacking for system D suggests that the multibranch loop is more stable when the G and U at the base of the top stem are not base paired with each other (*33*). The imino proton spectrum for strand D1 revealed peaks at 11.1 and 11.9 ppm, consistent with a G and U, respectively, in a G·U pair (*32*, *58*, *59*). Upon titration with strand D2, both of these peaks disappeared. The lack of resonances consistent with a G·U pair upon multibranch loop formation supports the predicted stacking conformation and is also consistent with expectations from comparisons with natural sequences (*27*).

Optical melts of system D indicated that it does not melt in a two-state manner, suggesting that one intermolecular stem melts before the other. To test this hypothesis, imino proton spectra were taken at several temperatures. One peak assigned to an A-U pair disappears before the other as the temperature is increased from 20 to 35 °C (Figure 9). This is consistent with the suggestion that the non-two-state melting reflects unequal stabilities of the two helical arms.

Free Energy Increments for Multibranch Loop Initiation. Optical melting measures the overall free energy change for association of the two strands forming the multibranch loop (Figure 1). These data can be analyzed with a nearest neighbor model to find $\Delta G°_{37,MBL}$, the free energy increment for the multibranch loop. The measured free energy change for the bimolecular association, $\Delta G°_{37,bimol}$, is the sum of the free energy increments of the multibranch loop, $\Delta G°_{37,MBL}$, the base pairs made in the two helices by duplex formation, $\Delta G°_{37,helix}$, the free energy penalty for bimolecular initiation, $\Delta G°_{37,bimol\ init}$, minus any possible stacking of mismatches formed on the helix in the product hairpin strand, $\Delta G°_{37,st\ in\ prod}$, that stabilize the hairpin product as shown in Figure 1.

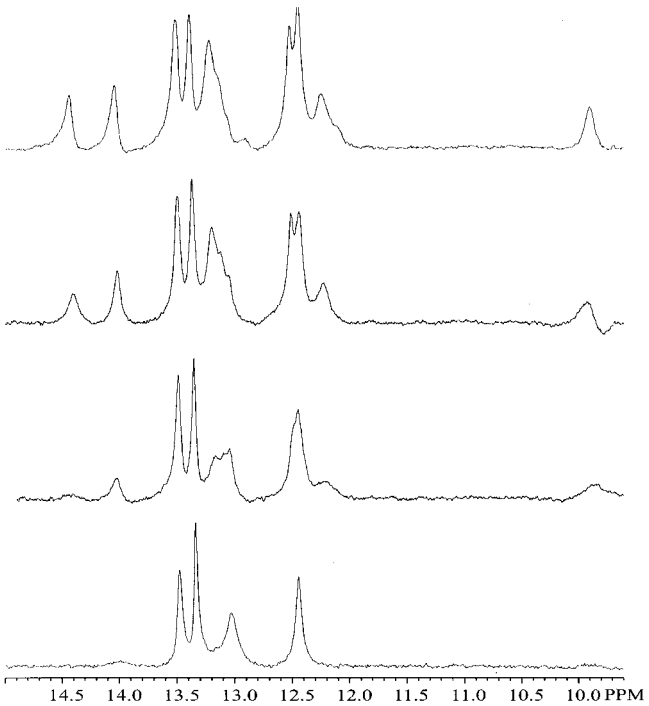

FIGURE 9: Imino proton spectra of system D at temperatures ranging from 20 °C (top) to 35 °C (bottom) in 5 °C increments. Notice that the resonance at 14.5 ppm disappears at a lower temperature than the resonance at 14.1 ppm, indicating that the helices are denaturing at slightly different temperatures.

Solving for $\Delta G^\circ_{37,\mathrm{MBL}}$:

$$\Delta G^\circ_{37,\mathrm{MBL}} = \Delta G^\circ_{37,\mathrm{bimol}} - \Delta G^\circ_{37,\mathrm{helix\ L}} - \Delta G^\circ_{37,\mathrm{helix\ R}} - \Delta G^\circ_{37,\mathrm{bimol\ init}} + \Delta G^\circ_{37,\mathrm{st\ in\ prod}} \quad (7)$$

The contributions from the helices and mismatches can be approximated with their nearest neighbor parameters (*33*, *60*, *61*). Values of $\Delta G^\circ_{37,\mathrm{MBL}}$ are given in Table 4.

Analysis of system B on the basis of the ΔG°_{37} from T_M^{-1} plots illustrates the process for determining the above parameters. The free energy of the helices can be determined by summing the free energies of the individual nearest neighbors (*60*). For the bimolecular helix to the left in Figure 2:

$$\Delta G^\circ_{37,\mathrm{helix\ L}} = \Delta G^\circ_{37}{}^{5'GG3'}_{3'CC5'} + \Delta G^\circ_{37}{}^{5'GA3'}_{3'CU5'} + \Delta G^\circ_{37}{}^{5'AG3'}_{3'UC5'} = -3.26 - 2.35 - 2.08 = -7.69 \text{ kcal/mol} \quad (8)$$

For the bimolecular helix to the right:

$$\Delta G^\circ_{37,\mathrm{helix\ R}} = \Delta G^\circ{}^{5'GA3'}_{3'CU5'} + \Delta G^\circ{}^{5'AC3'}_{3'UG5'} + \Delta G^\circ{}^{5'CG3'}_{3'GC5'} = -2.35 - 2.24 - 2.36 = -6.95 \text{ kcal/mol} \quad (9)$$

At 37 °C, intermolecular association has an unfavorable free energy increment of 4.09 kcal/mol (*60*). After separation of the two strands, there is the possibility of forming a $^{5'GG3'}_{3'CA5'}$ terminal mismatch, $\Delta G^\circ_{37} = -1.6$ kcal/mol (*61*). From Table 1, $\Delta G^\circ_{37,\mathrm{bimol}} = -4.05$ kcal/mol. From eq 7:

$$\Delta G^\circ_{37,\mathrm{MBL}} = -4.05 + 7.69 + 6.95 - 4.09 - 1.6 = 4.9 \text{ kcal/mol} \quad (10)$$

The total free energy of the multibranch loop is the sum of the favorable interactions in the multibranch loop and the initiation penalty for multibranch loop formation, $\Delta G^\circ_{37,\mathrm{MBLinit}}$.

Table 4: $\Delta G^\circ_{37,\mathrm{loop}}$, $\Delta G^\circ_{37,\mathrm{stacking}}$, and $\Delta G^\circ_{37,\mathrm{init}}$ for Multibranch Loops

system	$\Delta G^\circ_{37,\mathrm{MBL}}$ (kcal/mol) exptl[a]	$\Delta G^\circ_{37,\mathrm{MBLstacking}}$ (kcal/mol)[b]	$\Delta G^\circ_{37,\mathrm{MBLinit}}$ (kcal/mol)[c]
A	3.73	−5.12	8.85
B	4.90	−4.9	9.80
C	3.76	−4.9	8.66
D	1.26	−4.5	5.76
E	4.66	−4.9	9.56
F	1.06	−5.4	6.46
	[3.17][d]	[−4.3][d]	[7.47][d]
G	3.43	−3.42	6.85
H	4.37	−4.36	8.73
I	3.72	−5.12	8.84
J	3.23	−4.22	7.45
K	4.20	−4.3	8.50
L	3.54	−4.2	7.74

[a] Experimental values calculated from free energy changes determined from T_M^{-1} vs $\ln(C_T/4)$ plots of the optical melting data. For systems D and F, $\Delta G^\circ_{37,\mathrm{st\ in\ prod}}$ in eq 7 includes terms for formation of a G-U pair and a G·G mismatch at the terminus of the hairpin product. [b] $\Delta G^\circ_{37,\mathrm{MBLstacking}}$ is the sum of the free energy contributions for stacking of dangling ends, terminal mismatches, and helixes. The free energy change of the most stable stacking conformation is listed here. [c] $\Delta G^\circ_{37,\mathrm{MBLinit}} = \Delta G^\circ_{37,\mathrm{MBL}} - \Delta G^\circ_{37,\mathrm{MBLstacking}}$ (see eq 11). [d] $\Delta G^\circ_{37,\mathrm{MBL}}$, $\Delta G^\circ_{37,\mathrm{MBLstacking}}$, and $\Delta G^\circ_{37,\mathrm{MBLinit}}$ are presented for an alternate structural conformation for system F in which the unpaired U in strand D1 shown in Figure 2 is base paired with the 5′ single-stranded A in strand B2.

The favorable interactions include stacking of dangling ends, mismatches, and helixes. Solving for $\Delta G^\circ_{37,\mathrm{MBLinit}}$:

$$\Delta G^\circ_{37,\mathrm{MBLinit}} = \Delta G^\circ_{37,\mathrm{MBL}} - \Delta G^\circ_{37,\mathrm{dangling\ ends}} - \Delta G^\circ_{37,\mathrm{MBLmismatches}} - \Delta G^\circ_{37,\mathrm{coaxial\ stacking}} = \Delta G^\circ_{37,\mathrm{MBL}} - \Delta G^\circ_{37,\mathrm{MBLstacking}} \quad (11)$$

Subtracting the free energy contributions due to 5′ and/or 3′ dangling ends, terminal mismatches, and coaxial stacking of helixes for the conformation inferred to be most stable allows for calculation of $\Delta G^\circ_{37,\mathrm{MBLinit}}$ for multibranch loops (*33*, *60*−*64*).

Calculation of $\Delta G^\circ_{37,\mathrm{MBLinit}}$ is complex because of the number of possible stacking arrangements in a junction. For the example of system B (Figure 2), one possibility is the helix on the left side stacking on the top helix. For this conformation, the coaxial stacking interaction can be considered as $^{5'GA\ C3'}_{3'CA/G5'}$ which is decomposed into two nearest neighbors (*33*, *61*, *64*):

$$\Delta G^\circ_{37}{}^{5'GA\ C3'}_{3'CA/G5'} = \Delta G^\circ_{37}{}^{5'GA3'}_{3'CA5'} - \Delta G^\circ_{37}$$

(for the coaxial stack with the intervening mismatch) =

$$-1.1 - 2.1 = -3.2 \text{ kcal/mol} \quad (12)$$

The resulting 3′ dangling end on the helix on the right side, $\Delta G^\circ_{37} = -1.7$ kcal/mol (*61*), must also be considered. The total free energy contribution of the mismatches, dangling ends, and coaxial stack is −4.9 kcal/mol at 37 °C.

The other possibility for system B is that the top helix stacks on the right-hand helix. If this were so, the stacking interaction would be $^{5'GG3'}_{3'CC5'}$, $\Delta G^\circ_{37} = -3.26$ kcal/mol (*60*, *62*). Furthermore, there would be a terminal mismatch on the left-hand helix, $^{5'GA3'}_{3'CA5'}$, with a free energy of −1.1 kcal/mol (*61*). The sum of the contributions due to stacking and mismatches would be $\Delta G^\circ_{37} = -4.36$ kcal/mol, and therefore

the stacking conformation found in the first case results in a more stable structure (−4.9 vs −4.36 kcal/mol), and so the $\Delta G^\circ_{37,\mathrm{MBLinit}}$ for system B, from eq 11, is

$$\Delta G^\circ_{37,\mathrm{MBLinit}} = 4.9 + 4.9 = 9.8 \text{ kcal/mol} \qquad (13)$$

The calculations for $\Delta G^\circ_{37,\mathrm{MBLinit}}$ for the other systems presented here are done in a similar manner, always determining which stacking conformation is most stable using nearest neighbor values from the literature (*33*, *60*, *61*). The results are shown in Table 4. Note that determinants of stacking conformations in junctions are poorly understood, so that values in Table 4 should be considered first approximations (*9*−*14*).

DISCUSSION

RNA multibranch loops are poorly characterized thermodynamically, even though they are prevalent in nature. A more complete understanding of the stabilities of multibranch loops is integral to an improvement in the ability of computer algorithms to accurately predict RNA secondary structure, on the basis of either free energy minimization (*33*, *65*−*67*) or folding kinetics (*68*). Accurately predicting RNA secondary structure is becoming essential as the number of known RNA sequences increases. Knowledge of secondary structure can assist in understanding function and determining sites available for hybridization of antisense or probe oligonucleotides (*69*).

In previous studies aimed at determining the stabilities of various RNA structural motifs (*32*, *50*, *53*, *54*, *59*−*64*, *70*−*82*), such as internal loops, hairpins, and Watson−Crick base pairs, UV optical melting utilizing short oligonucleotides has been successful. The results of these experiments usually fit the two-state model and provide excellent approximations of stability. One of the major difficulties in studying multibranch loops is the large numbers of nucleotides needed to form such a structure. As the number of nucleotides increases, so does the number of possible secondary structures, and the more difficult it becomes to find a sequence that will produce the desired structure and melt in a two-state manner.

Due to the large number of multibranch loops that failed to fit the two-state model, ITC was used as an alternative method for determining the thermodynamic parameters of three systems (Table 3). ITC experiments provide a way to check the accuracy of thermodynamics determined by optical melting. Agreement between the free energy changes derived from optical melting and ITC experiments for the systems that were not two state indicates that optical melting is a reasonable method for measuring free energy changes for multibranch loops, even when the system is not two state. Optical methods permit wider surveys of sequence dependence because less material is required and data collection is faster than with calorimetry. We note, however, that the thermodynamics reported here cannot be compared with those obtained from optical melting by Klostermeier and Millar (*83*) because the analysis of the latter neglects the effect of strand stoichiometry on equilibrium constant.

Experimental determination of the stabilities of junctions allows testing of the functional form and parameters used to predict these stabilities. In current versions of the *mfold* and RNAstructure programs for predicting RNA secondary structure (*33*), the stability of a multibranch loop with fewer than six unpaired nucleotides is given by

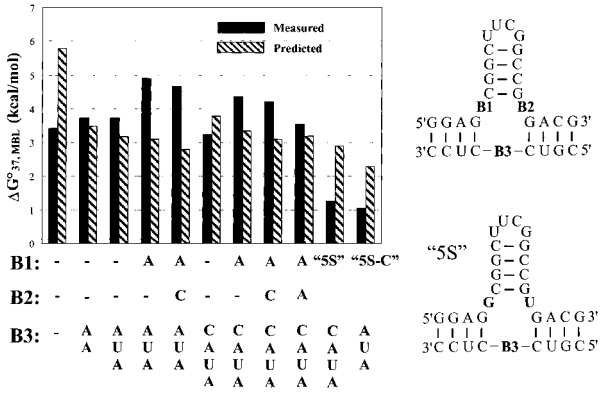

FIGURE 10: Comparison of measured (black bars) and predicted (striped bars) free energy increments for multibranch loops at 37 °C, $\Delta G^\circ_{37,\mathrm{MBL}}$. The systems with identical closing helices are illustrated with B1, B2, and B3 representing the bridging regions of variable unpaired nucleotides. System D, the mimic of the 5S rRNA sequence, is labeled "5S" and system F is labeled "5S-C".

$$\Delta G^\circ_{\mathrm{MBL}} = a + bn + ch + \Delta G^\circ_{\mathrm{MBLstacking}} \qquad (14)$$

Here *a*, *b*, and *c* are parameters, *n* is the number of unpaired nucleotides, *h* is the number of helixes forming the junction, and $\Delta G^\circ_{\mathrm{MBLstacking}}$ is the sum of the free energies for stacking of dangling ends, mismatches, and helixes in the multibranch loop. The value of $\Delta G^\circ_{\mathrm{MBLstacking}}$ is calculated from free energies of stacking interactions in model systems by a recursive algorithm that finds the most favorable combination of interactions. The values for *a*, *b*, and *c* were chosen by a genetic algorithm that varied *a*, *b*, and *c* to optimize predictions of known secondary structures containing roughly 140 000 nucleotides (*33*). The starting point for the genetic algorithm was based on preliminary experimental results for systems A, B, C, and I in Figure 2. There is, however, no experimental support for eq 14 or for any other functional form for the stabilities of junctions.

Figure 10 shows a comparison between measured free energies of multibranch loop formation and those calculated with eq 14 substituted with the parameters optimized to predict known secondary structures (*33*). The average measured free energy increment is 3.5 kcal/mol, and the average difference between measured and predicted increments is 1.1 kcal/mol. For an earlier set of parameters derived by trial and error for a roughly 10-fold smaller set of secondary structures (*63*), the average difference between measured and predicted increments is 1.9 kcal/mol. Thus the larger data set, genetic algorithm, and preliminary experimental results have improved predictions of junction stability.

While predictions of junction stability have improved, most of the multibranch loops in Table 4 are less stable than predicted (Figure 10). Because the parameters for multibranch loop initiation were chosen to optimize prediction of known secondary structures (*33*), the comparison suggests that natural multibranch loops are on average more stable than those with sequences not found in known structures. The results for system D are consistent with this suggestion. System D is based on comparative sequence analysis of the multibranch loop from 5S rRNA (*15*, *27*, *51*), and the free energy increment for loop initiation, $\Delta G^\circ_{37,\mathrm{MBLinit}}$, is roughly

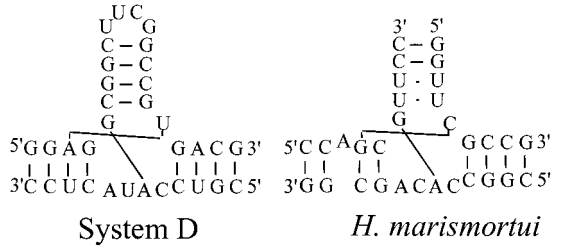

FIGURE 11: Comparison of system D and *H. marismortui* 5S rRNA (*84*) multibranch loop structures. System D, the most stable multibranch loop studied, was patterned after the consensus 5S rRNA multibranch loop structure. As illustrated, it can form two of the tertiary contacts that are observed in the crystal structure of the *H. marismortui* 5S rRNA (*84*).

2.5 kcal/mol more stable than the average of the other loops in this study and 1.6 kcal/mol more stable than predicted. This specific sequence may have evolved to be highly stable to serve its role in the ribosome. The sequence dependence of interactions providing this enhanced stability is apparently not adequately represented by eq 14.

The crystal structure of the 5S rRNA in the large ribosomal subunit from *Haloarcula marismortui* (*84*) suggests some factors that may stabilize the junction in system D. The *H. marismortui* multibranch loop involves a base triple, coaxial stacking, and a single-stranded region of four nucleotides that help to stabilize the change in direction of the phosphodiester backbone for the helix that is not stacked. These features are also possible in system D. System D differs in some aspects (see Figure 11): The hairpin-containing helix in system D is closed by two C-G pairs that are U·U pairs in *H. marismortui*. The adjacent G and U in system D, which on the basis of imino proton NMR spectra are not paired, are an unpaired G and C, respectively, in *H. marismortui*. In *H. marismortui*, this C is involved in a C-G-C base triple. In system D, an isosteric U-A-U triple can form with the A-U pair in the left helix. A universally conserved bulged A that stacks on this base triple, however, is absent in system D. In *H. marismortui*, the G that does not form a G-C pair is instead paired with the 5′ most A in a single-stranded 5′CACA sequence within the loop. This same interaction is possible with the single-stranded 5′CAUA sequence in system D. The C to U substitution in the single-stranded region does not interfere with the formation of an analogous structure because the C in the *H. marismortui* structure is positioned outside the loop to facilitate a sharp change in the direction of the phosphodiester backbone and is not involved with any loop contacts. These comparisons and the thermodynamics of system D suggest that base triples and noncanonical pairs involving multibranch loop nucleotides can contribute at least 2 kcal/mol to loop stability. Such interactions are ignored by eq 14.

The crystal structure of a fragment of the ribosome (*85*) provides further evidence that eq 14 is oversimplified. In particular, coaxial stacking of helixes predicted by eq 14 using current rules for coaxial stacking (*33*, *62*−*64*) is not observed in the structure. Evidently, certain sequence motifs will be even more stable than predicted by eq 14. These sequence motifs may extend beyond the junction region. For example, systems B and C have the same junction, but their stabilities differ by 1.1 kcal/mol. While it is possible that this reflects experimental error, it may also reflect a real difference. For example, as discussed above, base triples can form between junction nucleotides and Watson−Crick base pairs in adjacent helixes. Future studies of structure and energetics will be required to provide better approximations of junction stability from sequence.

This study shows that optical melting and ITC experiments can be used to measure the stabilities of RNA multibranch loops. The thermodynamic parameters measured have improved the accuracy of RNA secondary structure prediction and also serve as a starting point for further studies of the sequence dependence of stability.

ACKNOWLEDGMENT

The authors thank T. W. Barnes, M. Burkard, R. Kierzek, S. Schroeder, and T. Xia for discussions on synthesis and thermodynamic parameters. The authors also thank S. Testa and J. Sabina for assistance with T7 RNA transcription, T. Shrader for making available the clone of the 6-His tagged T7 RNA polymerase, and the group of Professor K. Breslauer for helpful suggestions about ITC.

REFERENCES

1. Meselson, M. S., and Radding, C. H. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 358−361.
2. Sigel, N., and Alberts, B. (1972) *J. Mol. Biol. 71*, 789−793.
3. Eichman, B. F., Vargason, J. M., Mooers, B. H. M., and Shing Ho, P. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 3971−3976.
4. Holliday, R. (1964) *Genet. Res. 5*, 282−304.
5. Kadrmas, J. L., Ravin, A. J., and Leontis, N. B. (1995) *Nucleic Acids Res. 23*, 2212−2222.
6. Zhong, M., Rashes, M. S., Leontis, N. B., and Kallenbach, N. R. (1994) *Biochemistry 33*, 3660−3667.
7. Ladbury, J. E., Sturtevant, J. M., and Leontis, N. B. (1994) *Biochemistry 33*, 6828−6833.
8. Leontis, N. B., Kwok, W., and Newman, J. S. (1991) *Nucleic Acids Res. 19*, 759−766.
9. Leontis, N. B., Hills, M. T., Piotto, M., Malhotra, A., Nussbaum, J., and Gorenstein, D. G. (1993) *J. Biomol. Struct. Dyn. 11*, 215−223.
10. Orr, J. W., Hagerman, P. J., and Williamson, J. R. (1998) *J. Mol. Biol. 275*, 453−464.
11. Rosen, M. A., and Patel, D. J. (1993) *Biochemistry 32*, 6563−6575.
12. Welch, J. B., Walter, F., and Lilley, D. M. J. (1995) *J. Mol. Biol. 251*, 507−519.
13. Hagerman, P. J. (1996) *Cur. Opin. Struct. Biol. 6*, 643−649.
14. Overmars, F. J. J., Pikkemaat, J. A., van den Elst, H., van Boom, J. H., and Altona, C. (1996) *J. Mol. Biol. 255*, 702−713.
15. Shen, Z., and Hagerman, P. J. (1994) *J. Mol. Biol. 241*, 415−430.
16. Bassi, G. S., Murchie, A. I. H., Walter, F., Clegg, R. M., and Lilley, D. M. J. (1997) *EMBO J. 16*, 7481−7489.
17. Walter, F. W., Murchie, A. I. H., Duckett, D. R., and Lilley, D. M. J. (1998) *RNA 4*, 719−728.
18. Scott, W. G., Finch, J. T., and Klug, A. (1995) *Cell 81*, 991−1002.
19. Scott, W. G., Murray, J. B., Arnold, J., Stoddard, B. L., and Klug, A. (1996) *Science 274*, 2065−2069.
20. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) *Nature 372*, 111−114.
21. Hertel, K. J., Stage-Zimmermann, T. K., Ammons, G., and Uhlenbeck, O. C. (1998) *Biochemistry 37*, 16983−16988.
22. Brown, J. W. (1998) *Nucleic Acids Res. 26*, 351−352.
23. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.

24. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
25. Michel, F., Umesono, K., and Ozeki, H. (1989) *Gene 82*, 5−30.
26. Larsen, N., Samuelsson, T., and Zwieb, C. (1998) *Nucleic Acids Res. 26*, 177−178.
27. Szymanski, M., Specht, T., Barciszewska, M. Z., Barciszewski, J., and Erdmann, V. A. (1998) *Nucleic Acids Res. 26*, 156−159.
28. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) *J. Mol. Biol. 256*, 701−719.
29. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) *Nucleic Acids Res. 26*, 148−153.
30. Mathews, D. H., Banerjee, A. R., Luan, D. D., Eickbush, T. H., and Turner, D. H. (1997) *RNA 3*, 1−16.
31. Williams, K. P., and Bartel, D. P. (1996) *RNA 2*, 1306−1310.
32. McDowell, J. A., and Turner, D. H. (1996) *Biochemistry 35*, 14077−14089.
33. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
34. McBride, L. J., and Caruthers, M. H. (1983) *Tetrahedron Lett. 24*, 245−249.
35. Beaucage, S. L., and Caruthers, M. H. (1981) *Tetrahedron Lett. 22*, 1859−1862.
36. Usman, N., Ogilvie, K. K., Jiang, M.-V., and Cedergren, R. (1987) *J. Am. Chem. Soc. 109*, 7845−7854.
37. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) *Nucleic Acids Res. 23*, 2677−2684.
38. Stawinski, J., Stromberg, R., Thelin, M., and Westman, E. (1988) *Nucleosides Nucleotides 7*, 779−782.
39. Chou, S.-H., Flynn, P., and Reid, B. (1989) *Biochemistry 28*, 2422−2435.
40. Ichetovkin, I., Abramochkin, G., and Shrader, T. E. (1997) *J. Biol. Chem. 272*, 33009−33014.
41. McBride, L. J., McCollum, C., Davidson, S., Efcavitch, J. W., Andrus, A., and Lombardi, S. J. (1988) *BioTechniques 6*, 362−367.
42. Testa, S. M., Haidaris, C. G., Gigliotti, F., and Turner, D. H. (1997) *Biochemistry 36*, 15303−15314.
43. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783−8798.
44. Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 589, CRC Press, Cleveland, OH.
45. Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, G. D., Ed.) p 197, CRC Press, Cleveland, OH.
46. Petersheim, M., and Turner, D. H. (1983) *Biochemistry 22*, 256−268.
47. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L.-N. (1989) *Anal. Biochem. 179*, 131−137.
48. Gulbransen, E. A., and Robinson, A. L. (1934) *J. Am. Chem. Soc. 56*, 2637−2641.
49. Hore, P. J. (1983) *J. Magn. Reson. 55*, 283−300.
50. Antao, V. P., Lai, S. Y., and Tinoco, I., Jr. (1991) *Nucleic Acids Res. 19*, 5901−5905.
51. Specht, T. (1991) *Nucleic Acids Res. 19*, 2189−2191.
52. 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.
53. Holbrook, J. A., Capp, M. W., Saecker, R. M., and Record, M. T., Jr. (1999) *Biochemistry 38*, 8409−8422.
54. Chalikian, T. V., Volker, J., Plum, G. E., and Breslauer, K. J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 7853−7858.
55. Xia, T., McDowell, J. A., and Turner, D. H. (1997) *Biochemistry 36*, 12486−12487.
56. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., and Turner, D. H. (1985) *Biochemistry 24*, 4533−4539.
57. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
58. Johnston, P. D., and Redfield, A. D. (1981) *Biochemistry 20*, 1147−1156.
59. He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., and Turner, D. H. (1991) *Biochemistry 30*, 11124−11132.
60. 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.
61. Turner, D. H. (2000) 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.
62. Walter, A. E., and Turner, D. H. (1994) *Biochemistry 33*, 12715−12719.
63. 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.
64. Kim, J., Walter, A. E., and Turner, D. H. (1996) *Biochemistry 35*, 13753−13761.
65. McCaskill, J. S. (1990) *Biopolymers 29*, 1105−1119.
66. Rivas, E., and Eddy, S. R. (1999) *J. Mol. Biol. 285*, 2053−2068.
67. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994) *Monatsh. Chem. 125*, 167−168.
68. Gultyaev, A. P., van Batenburg, F. H. D., and Pleij, C. W. A. (1995) *J. Mol. Biol. 250*, 37−51.
69. Mathews, D. H., Burkard, M. E., Freier, S. M., Wyatt, J. R., and Turner, D. H. (1999) *RNA 5*, 1458−1469.
70. Antao, V. P., and Tinoco, I., Jr. (1992) *Nucleic Acids Res. 20*, 819−824.
71. Freier, S. M., Burger, B. J., Alkema, D., Neilson, T., and Turner, D. H. (1983) *Biochemistry 22*, 6198−6206.
72. Giese, M. R., Betschart, K., Dale, T., Riley, C. K., Rowan, C., Sprouse, K. J., and Serra, M. J. (1998) *Biochemistry 37*, 1094−1100.
73. Groebe, D. R., and Uhlenbeck, O. C. (1988) *Nucleic Acids Res. 16*, 11725−11735.
74. Groebe, D. R., and Uhlenbeck, O. C. (1989) *Biochemistry 28*, 742−747.
75. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) *Biochemistry 29*, 278−285.
76. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) *Biochemistry 30*, 6428−6436.
77. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) *Biochemistry 30*, 8242−8251.
78. Schroeder, S., and Turner, D. H. (2000) *Biochemistry 39*, 9257−9274.
79. Serra, M. J., Barnes, T. W., Betschart, K., Gutierrez, M. J., Sprouse, K. J., Riley, C. K., Stewart, L., and Temel, R. E. (1997) *Biochemistry 36*, 4844−4851.
80. Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) *Biochemistry 25*, 5755−5759.
81. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
82. Dale, T., Smith, R., and Serra, M. J. (2000) *RNA 6*, 608−615.
83. Klostermeier, D., and Millar, D. P. (2000) *Biochemistry 39*, 12970−12978.
84. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 298*, 905−920.
85. Agalarov, S. C., Prasad, G. S., Funke, P. M., Stout, C. D., and Williamson, J. R. (2000) *Science 288*, 107−112.

BI0029548