Improved Parameters for the Prediction of RNA Hairpin Stability†

Martin J. Serra,* Thomas W. Barnes, Kelly Betschart, Mathew J. Gutierrez, Kimberly J. Sprouse, Cheryl K. Riley, Lora Stewart, and Ryan E. Temel

Department of Chemistry, Allegheny College, Meadville, Pennsylvania 16335

Received October 17, 1996; Revised Manuscript Received February 17, 1997⊗

ABSTRACT: Thermodynamic parameters are reported for hairpin formation in 1 M NaCl by RNA sequences of the type GGXAN$_m$AYCC, where XY is the set of four Watson−Crick base pairs and the underlined loop sequences are three to nine nucleotides. A nearest neighbor analysis of the data indicates the free energy of loop formation at 37 °C is dependent upon loop size and closing base pair. The model previously developed to predict the stability for RNA hairpin loops ($n > 3$) includes contributions from the size of the loop, the identity of the closing base pair, the free energy increment ($\Delta G°_{37mm}$) for the interaction of the closing base pair with the first mismatch and an additional stabilization term for GA and UU first mismatches [Serra, M. J., Axenson, T. J., & Turner, D. H. (1994) *Biochemistry 33*, 14289]. The results presented here allow improvements in the parameters used to predict RNA hairpin stability. For hairpin loops of $n = 4-9$, $\Delta G°_{37iL}(n)$ is 4.9, 5.0, 5.0, 5.0, 4.9, and 5.5 kcal/mol, respectively, and the penalty for hairpin closure by AU or UA is +0.6 kcal/mol. $\Delta G°_{37iL}(n)$ is the free energy for initiating a loop of n nucleotides. The model for predicting hairpin loop stability for loops larger than three becomes $\Delta G°_{37L}(n) = \Delta G°_{37iL}(n) + \Delta G°_{37mm} + 0.6$(if closed by AU or UA) -0.7(if first mismatch is GA or UU). Hairpin loops of three are modeled as independent of loop sequence with $\Delta G°_{37iL}(3) = 4.8$ and the penalty for AU closure of +0.6 kcal/mol. Thermodynamic parameters for hairpin formation in 1 M NaCl for 11 naturally occurring RNA hairpin sequences are reported. The model provides good agreement with the measured values for both T_M (within 10 °C of the measured value) and $\Delta G°_{37}$ (within 0.8 kcal/mol of the measured value) for hairpin formation. In general, the nearest neighbor model allows prediction of RNA hairpin stability to within 5−10% of the experimentally measured values.

Determination of the three-dimensional structure of a biological molecule from its sequence is a major goal of biochemistry. The rules governing the folding of biological molecules into their three-dimensional structures are just beginning to be revealed. The first step toward predicting the three-dimensional structure is predicting secondary structure. RNA, with a limited molecular vocabulary of four nucleotides, and strong local intermolecular forces, is an ideal candidate to elucidate the rules governing secondary structure formation.

Several methods for determining RNA structure are currently used. Phylogenetic analysis uses sequence comparison of RNAs with similar functions to determine common structural features (Woese & Pace, 1993; Gutell et al., 1993). Chemical and enzymatic probes are used to map RNA structure (Parker, 1989; Ehresmann et al., 1987). Recently, NMR analysis provided the structure of several RNA molecules (Heus & Pardi, 1991; Borer et al., 1995; Fountain et al., 1996; Huang et al., 1996). X-ray diffraction analysis has elucidated the structure of tRNAs (Woo et al., 1980; Susman et al., 1978; Robertus et al., 1974) and a hammerhead ribozyme (Pley et al., 1994; Scott et al., 1995). Thermodynamic analysis is used to predict the most stable secondary structure of RNA (Jaeger et al., 1990, 1995). Each of these methods has its advantages in the determination of RNA structure. Combining and improving these methods will make RNA structure determination more reliable.

Several methods for predicting RNA secondary structure based upon free energy minimization techniques are in use (Martinez, 1990; Zuker, 1989). These methods can predict many of the structural features of RNAs, but none can yet predict secondary structure with better than 70% reliability (Jaeger et al., 1990). Further refinement of these methods requires improved thermodynamic parameters for secondary structural motifs.

Hairpins are an important structural motif in RNA. For example, nearly 70% of *Escherichia coli* 16S rRNA nucleotides are found in small hairpin structures. Hairpins likely provide nucleation sites for the overall three-dimensional folding of RNA. Additionally, hairpins are involved in a number of important tertiary interactions with either protein (Wu & Uhlenbeck, 1987; Lazinski et al., 1989) or RNA (Murphy et al., 1995; Marino et al., 1995).

We have previously presented a model to predict RNA hairpin stability based upon a study of RNA hairpin loops of six nucleotides (Serra et al., 1994). We report here improved thermodynamic parameters for hairpin formation with loops of three to nine nucleotides. Combined with previous studies, these results permit improved prediction of RNA hairpin stability.

MATERIALS AND METHODS

RNA Synthesis and Purification. Oligomers were synthesized on solid support using the phosphoramidite approach (Wu et al., 1989; Usman et al., 1987). After ammonia and fluoride deprotection, the crude oligomer was purified by

† This work was supported by an award of Research Corporation, NIH Grant GM49429, and Merck/AAAS Undergraduate Science Research Program.

* Author to whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0006-2960(96)02608-6 CCC: $14.00

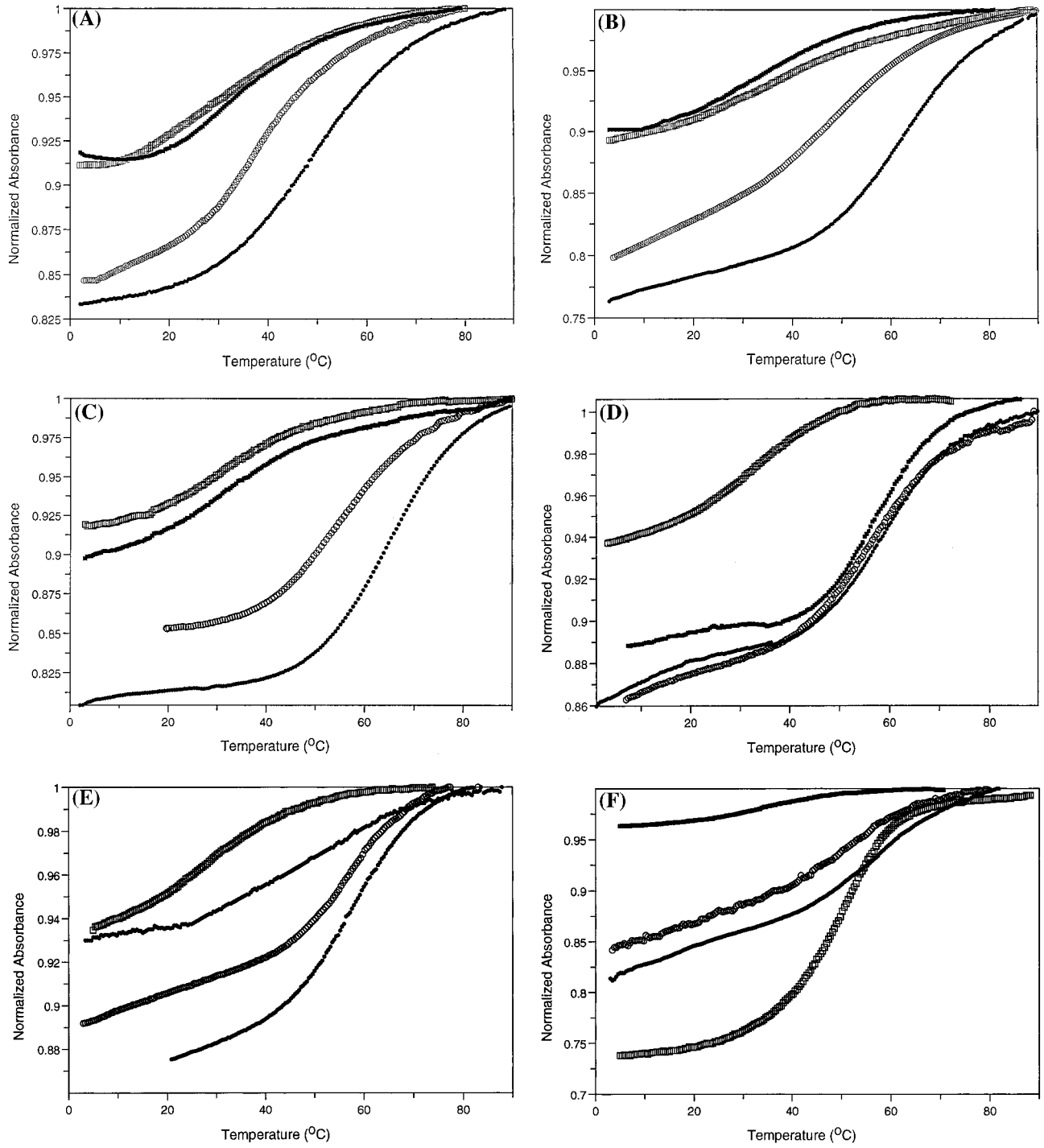


FIGURE 1: Normalized melting curves. Loop sizes are (A) three, (B) four, (C) five, (D) seven, (E) eight, and (F) nine. Base closures are AU (■), UA (□), GC (○), and CG (●). The loop of seven with the AU closure has the stem $\overset{\text{GCGA}}{\underset{\text{CGCU}}{\bullet\bullet\bullet\bullet}}$, the loop of nine with the UA closure has the stem $\overset{\text{GCGU}}{\underset{\text{CGCA}}{\bullet\bullet\bullet\bullet}}$, and all other hairpins have three base pair stems. Buffer is 1 M NaCl, 10 mM sodium cacodylate, and 0.5 mM EDTA, pH 7.0.

preparative TLC (*n*-propanol:ammonium hydroxide:water, 55:35:10) and Sep-Pak C18 (Waters) chromatography. Purities were checked by either analytical C8 HPLC chromatography or analytical TLC and were greater than 95%.

Melting Curves and Data Analysis. The buffer for the melting studies was 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7, unless indicated otherwise. Single-stranded extinction coefficients were calculated from the extinction coefficients for dinucleotide monophosphate and nucleosides, as described previously (Borer, 1975; Richards, 1975). Strand concentrations were determined from high-temperature absorbance at 280 nm. Absorbance versus temperature curves were measured at 280 nm with a heating or cooling rate of 1.0 °C min^{-1}, on a Perkin Elmer Lambda 2S spectrophotometer as described previously (Serra et al., 1994). Oligomer concentrations were varied over at least a 40-fold range between 1 mM and 10 μM.

Absorbance versus temperature profiles were fit to a two-state model with sloping base lines by using a nonlinear least-squares program (Petersheim & Turner, 1983; Freier et al., 1983), adapted for unimolecular hairpin transitions. Thermodynamic parameters for hairpin formation were obtained

Table 1: Thermodynamic Parameters for Hairpin Formation in 1.0 M NaCl

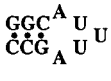
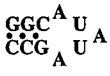
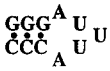
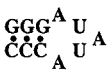
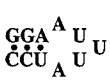
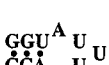
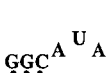
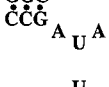
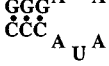
RNA Hairpin	T_M °C (pred)	ΔH° kcal/mol (pred)	ΔS° eu (pred)	$\Delta G°_{37}$ kcal/mol (pred)	$\Delta H°_L$[a] kcal/mol	$\Delta S°_L$[a] eu	$\Delta G°_{37L}$[a] kcal/mol (pred)	$\Delta G°_{37IL}$[b] kcal/mol
	51.0 (56.1)	-21.9±1.9 (-26.4)	-67.5±6.0 (-80.1)	-0.94±0.15 (-1.5)	+4.5	-2.9	+5.4 (+4.8)	+5.4
	38.9 (51.5)	-27.6±1.7 (24.3)	-88.3±5.1 (-74.9)	-0.26±0.22 (-1.1)	-3.3	-28.9	+5.6 (+4.8)	+5.6
	30.5 (34.7)	-27.8±1.4 (-24.8)	-91.7±4.7 (-80.7)	+0.61±0.1 (+0.2)	-2.4	-26.5	+5.8 (+5.4)	+5.8
	24.5 (32.2)	-14.9±1.3 (-21.8)	-49.9±4.8 (-71.4)	+0.63±0.2 (+0.4)	+7.5	+6.0	+5.7 (+5.4)	+5.7
	63.3 (65.0)	-32.2±1.4 (-35.5)	-95.7±4.3 (-104.9)	-2.51±0.09 (-2.9)	-5.8	-31.1	+3.8 (+3.4)	+5.3
	53.3 (60.8)	-26.0±1.9 (-29.5)	-79.7±6.0 (-88.4)	-1.30±0.11 (-2.1)	-1.7	-20.3	+4.6 (+3.8)	+5.7
	37.6 (48.6)	-20.5±1.1 (-29.3)	-65.9±3.7 (-91.2)	-0.05±0.15 (-0.5)	+4.9	-0.7	+5.2 (+4.7)	+6.0
	41.2 (51.3)	-21.7±2.6 (-26.4)	-64.2±8.1 (-81.4)	-0.32±0.12 (-0.6)	+0.7	-8.3	+4.7 (+4.5)	+5.7
	64.7 (64.0)	-31.6±2.6 (-35.5)	-93.3±8.0 (-105.2)	-2.55±0.18 (-2.8)	-5.2	-28.7	+3.8 (+3.5)	+5.3
	66.3 (64.0)	-27.6±2.0 (-35.5)	-81.3±6.5 (-105.2)	-2.73±0.10 (-2.8)	-1.2	-16.7	+3.6 (+3.5)	+5.1
	55.4 (59.6)	-29.3±2.4 (-29.5)	-89.0±7.7 (-88.7)	-1.67±0.08 (-2.0)	-5.0	-29.6	+4.2 (+3.9)	+5.3
	54.3 (59.6)	-26.6±2.2 (-29.5)	-81.3±6.2 (-88.7)	-1.48±0.18 (-2.0)	-2.3	-21.9	+4.4 (+3.9)	+5.5
	39.4 (47.5)	-23.9±2.3 (-29.3)	-78.4±8.5 (-91.5)	-0.21±0.15 (-0.4)	+1.5	-13.2	+5.0 (+4.8)	+5.8
	40.7 (50.0)	-23.2±3.4 (-26.4)	-74.0±10.5 (-81.7)	-0.30±0.29 (-0.5)	+0.8	-18.1	+4.8 (+4.6)	+5.8
	59.2 (64.0)	-32.6±2.6 (-35.5)	-98.3±2.6 (-105.2)	-2.2±0.2 (-2.8)	-6.2	-33.7	+4.1 (+3.5)	+5.6
	57.3 (59.6)	-34.0±1.6 (-29.5)	-102.8±5.0 (-88.7)	-2.1±0.1 (-2.0)	-9.7	-43.4	+3.8 (+3.9)	+4.9
	57.5 (66.3)	-39.6±2.6 (-39.4)	-119.9±8.3 (-116.1)	-2.5±0.2 (-2.9)	-4.1	-30.1	+5.2 (+4.8)	+6.0
	42.4 (50.0)	-21.9±1.2 (-26.4)	-69.3±3.8 (-81.7)	-0.38±0.2 (-0.5)	+0.5	-13.4	+4.7 (+4.6)	+5.7

Table 1 (Continued)

RNA Hairpin	T_M °C (pred)	$\Delta H°$ kcal/mol (pred)	$\Delta S°$ eu (pred)	$\Delta G°_{37}$ kcal/mol (pred)	$\Delta H°_{L}$[a] kcal/mol	$\Delta S°_{L}$[a] eu	$\Delta G°_{37L}$[a] kcal/mol (pred)	$\Delta G°_{37iL}$[b] kcal/mol
GGC A U A A CCG A U A A	56.5 (65.0)	-36.4±2.3 (-35.5)	-110.4±737 (-104.9)	-2.1±0.1 (-2.9)	-10.0	-45.8	+4.2 (+3.4)	+5.7
GGG A U A A CCC A U A A	58.1 (60.8)	-36.0±1.3 (-29.5)	-108.5±3.9 (-88.4)	-2.3±0.2 (-2.1)	-11.7	-49.1	+3.6 (+3.8)	+4.7
GGA A U A A CCU A U A A	60.8 (48.6)	-32.0±6.0 (-29.3)	-96.1±18.8 (-91.2)	-2.2±0.4 (-0.5)	-6.6	-30.9	+3.0 (+4.8)	+3.8
GGU A U A A CCA A U A A	29.1 (51.3)	-20.8±1.3 (-26.4)	-68.7±4.2 (-81.4)	+0.5±0.1 (-0.6)	+1.6	-12.8	+5.6 (+4.5)	+6.6
GGC A U A A • • • A CCG A U A A	57.1 (58.9)	-34.8±1.8 (-35.5)	-105.3±5.5 (-106.8)	-2.1±0.2 (-2.3)	-8.4	-40.7	+4.2 (+4.0)	+5.7
GGG A U A A • • • A CCC A U A A	56.8 (53.7)	-34.9±1.8 (-29.5)	-105.8±5.9 (-90.3)	-2.1±0.2 (-1.5)	-10.6	-46.4	+3.8 (+4.4)	+4.9
GGA A U A A • • • A CCU A U A A	34.2 (41.9)	-24.0±2.3 (-29.3)	-78.2±7.7 (-93.1)	+0.2±0.3 (+0.1)	+1.4	-13.0	+5.4 (+5.3)	+6.2
GCG A U A A • • • • A CGGU A U A A	53.8 (61.7)	-42.3±2.9 (-39.4)	-129.2±9.2 (-117.7)	-2.2±0.2 (-2.4)	-6.8	-39.4	+5.5 (+5.3)	+6.3
GCGU A U A A • • • • A CGCA A U A A	51.6 (64.9)	-23.5±4.1 (-36.5)	-72.3±1.3 (-107.9)	-1.0±0.3 (-2.4)	+9.0	+8.2	+6.5 (+5.1)	+7.5

[a] Calculated from equations equivalent to $\Delta G°_{37L} = \Delta G°_{37}$(hairpin formation) − $\Delta G°_{37}$(pred. stem). [b] Calculated from $\Delta G°_{37L} = \Delta G°_{37iL} - \Delta G°_{37mm}$, where $\Delta G°_{37mm}$ is derived from measurements of model duplexes with terminal mismatches (Serra & Turner, 1995) except for loop of three where $\Delta G°_{37iL} = \Delta G°_{37L}$.

from the averages of the fits of the individual melting curves. Melting temperatures and thermodynamic parameters were concentration independent as expected for hairpin formation.

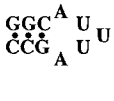
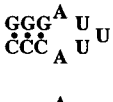
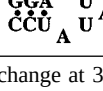
RESULTS

To explore the role of loop size on hairpin stability, the set of RNA hairpins of loop size three to nine (excluding six) with all four Watson−Crick base closures was prepared and the thermodynamics of hairpin formation were measured by optical melting. The first mismatch, AA, was chosen to facilitate comparison with previously studied hairpins (Serra et al., 1993, 1994).

Typical melting curves are shown in Figure 1, and the measured thermodynamic parameters in 1 M NaCl are listed in Table 1. The free energy for folding, $\Delta G°_{37}$, of the RNA hairpins varies with loop size and closing base pair.

The two hairpins closed by AU base pairs with loops of five (AUAUA) in the original series did not melt in a two-state unimolecular manner but rather had concentration dependent melting temperatures. Therefore, an additional set of hairpins, with all four base closures and loops of five (AUUUA), are included in this study. The reliability of the thermodynamic values is seen in the similarity of the $\Delta G°_{37}$ for the two pairs of hairpins with loops of five and the same closing base pair, GC and CG; each value is within 0.2 kcal/mol of the other. The hairpin with a loop of seven with AU closure from the original set also did not melt in a two-state unimolecular manner and was replaced by a hairpin with a

Table 2: Thermodynamic Parameters for Hairpin Formation in 0.1 M NaCl

RNA Hairpin	T_M °C	ΔH° kcal/mol	ΔS° eu	$\Delta G°_{37}$ kcal/mol	$\Delta\Delta G°_{37,0.1\text{-}1.0}$[a] kcal/mol
	48.7	22.6	70.1	-0.84	+0.1
	35.2	17.8	57.8	+0.09	+0.4
	15.0	15.3	53.1	+1.11	+0.5
	20.1	14.1	47.9	+0.76	+0.2
	62.3	31.5	93.9	-2.38	+0.2
	48.2	20.8	64.9	-0.72	+0.6
	32.8	21.9	71.4	+0.28	+0.3
	36.0	19.6	63.4	-0.1	+0.2
	61.8	34.4	102.7	-2.59	+0.1
	53.6	27.4	84.0	-1.37	+0.2
	34.3	31.5	102.3	+0.22	+0.4

[a] Difference in free energy change at 37 °C measured at 0.1 and 1 M NaCl.

stem of four base pairs (Table 1). The extra base pair in the stem does not affect the stability of the hairpin loop. This is seen for the two hairpin loops of nine closed by AU but with different stem sizes in Table 1 (see below).

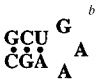
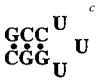
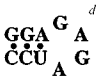
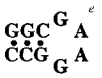
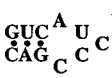
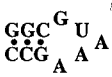
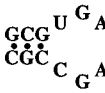
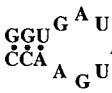
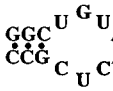
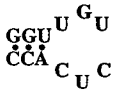
The stability of an RNA hairpin can be dissected into its two structural motifs, the double helical stem and the loop. The thermodynamic parameters for loop formation ($\Delta H°_L$, $\Delta S°_L$, and $\Delta G°_{37L}$) can be calculated by subtracting the double helical stem contribution from the hairpin value. For example, the free energy for loop formation at 37 °C, $\Delta G°_{37L}$, can be calculated from $\Delta G°_{37L} = \Delta G°_{37}$(measured for hairpin formation) $- \Delta G°_{37}$(stem). The results of this analysis are presented in Table 1.

Hairpin loops of three are modeled as independent of loop sequence with no interaction between the closing base pair and first mismatch (Serra & Turner, 1995). Therefore, $\Delta G°_{37L}$ is equal to $\Delta G°_{37iL}$ for hairpin loops of three, where $\Delta G°_{37iL}$ is the free energy for initiating a hairpin loop. As observed previously for hairpin loops of six (Serra et al., 1993), the hairpin loops of three closed by AU or UA (average $\Delta G°_{37L} = 5.8$ kcal/mol) are less stable than loops closed by GC or CG base pairs (average $\Delta G°_{37L} = 5.5$ kcal/mol).

Previous studies (Serra et al., 1994), on hairpins with loop size of six, have shown that the stability of hairpin loops depends on the stacking interactions of the first mismatch with the closing base pair. $\Delta G°_{37L}(n) = \Delta G°_{37iL}(n) + \Delta G°_{37mm}$, where $\Delta G°_{37il}(n)$ is the free energy for initiating a hairpin loop of n nucleotides and $\Delta G°_{37mm}$ is the free energy increment for the interaction of the closing base pair with the first mismatch. The $\Delta G°_{37mm}$ is derived from measurements of model duplexes with terminal mismatches. The measured $\Delta G°_{37iL}$ for hairpin loops larger than three in Table 1 can be determined by subtracting $\Delta G°_{37mm}$ from $\Delta G°_{37L}$. The results of this analysis are presented in Table 1.

The average $\Delta G°_{37iL}$ for hairpin loops of size 4, 5, and 7 to 9 closed by CG or GC base pairs are 5.5, 5.3, 5.0, 5.2, and 5.3 kcal/mol, respectively. It is interesting to note that small hairpin loops closed by CG are more stable than loops closed by GC, while for larger loops the GC closures are more stable (Table 1). The average $\Delta G°_{37iL}$ for hairpin loops

Table 3: Thermodynamic Parameters for Hairpin Formation of Natural Sequences in 1.0 M NaCl

RNA Hairpin	T_M °C Measured (Predicted)	ΔH° kcal/mol Measured (Predicted)	ΔS° eu Measured (Predicted)	$\Delta G°_{37}$ kcal/mol Measured (Predicted)	$\Delta G°_{37L}$[a] kcal/mol Measured (Predicted)
	28.0 (32.1)	-18.7±1.4 (-21.21)	-61.9±4.6 (-69.6)	+0.5±0.2 (+0.3)	5.6 (5.4)
	55.0 (56)	-22.5±2.1 (-26.4)	-68.6±6.7 (-80.1)	-1.2±0.2 (-1.5)	5.1 (4.8)
	52.3 (51.2)	-30.5±1.9 (-28.6)	-93.6±6.1 (-88.3)	-1.5±0.2 (-1.2)	3.7 (4.0)
	69.9 (65.9)	-34.7±2.5 (-35.6)	-101.2±7.2 (-105.0)	-3.2±0.2 (-3.0)	3.1 (3.3)
	53.1 (45.6)	-24.1±2.0 (-29.1)	73.9±4.7 (-91.3)	-1.2±0.2 (-0.9)	3.2 (3.5)
	65.6 (71.2)	-39.8±2.5 (35.3)	-117.4±7.3 (-102.5)	-3.3±0.3 (-3.4)	3.0 (2.9)
	56.8 (56.3)	-29.6±2.0 (22.5)	-89.7±5.7 (-68.3)	-1.8±0.2 (-1.4)	3.6 (4.0)
	44.6 (53.3)	-32.3±3.3 (-26.3)	-101.4±10.6 (-80.6)	-0.8±0.1 (-1.3)	4.3 (3.8)
	57.7 (67.7)	-32.3±2.0 (31.7)	-97.7±6.3 (-9.3)	-2.0±0.2 (-2.8)	4.3 (3.5)
	35.6 (39.5)	-21.8±1.9 (-23.2)	-70.6±6.0 (-74.21)	+0.2±0.1 (-0.2)	5.3 (4.9)
	47.2 (41.6)	-31.8±2.5 (32.4)	-99.4±8.0 (102.8)	-1.0±0.2 (-0.5)	5.1 (5.6)

[a] Calculated from $\Delta G°_{37L} = \Delta G°_{37}$(hairpin formation) − $\Delta G°_{37}$(pred. stem). Sequences modeled on [b]*E. coli* large subunit rRNA position 1752; [c]*H. morrhuae* large subunit rRNA position 270 (*E. coli* equivalent position); [d]*E. coli* large ribosomal subunit rRNA position 2660; [e]*E. coli* small ribosomal subunit rRNA position 726; [f]*E. coli* large ribosomal rRNA position 886; [g]*Saccharomyces cerevisae* large ribosomal subunit rRNA position 612 (*E. coli* equivalent position); [h]*E. coli* large ribosomal subunit rRNA 328; [i]*E. coli* large ribosomal subunit rRNA 87; [j]*Aspergillus nidulans* small ribosomal rRNA subunit 1350 (*E. coli* equivalent position); [k]*Methanospirillum hungatei* small ribosomal rRNA subunit 1350 (*E. coli* equivalent position); and [l]*S. cerevisae* mitochondrial small ribosomal subunit rRNA 320 (*E. coli* equivalent position) (Gutell et al., 1993).

of 4, 5, and 7−9 closed by AU or UA base pairs are 5.8, 5.8, 5.8, 5.1, and 6.6 kcal/mol, respectively.

Thermodynamic parameters for some hairpins were also measured in 0.1 M NaCl. These results are presented in Table 2. The trends are the same as in 1 M NaCl, but the hairpins are less stable by an average of 0.3 kcal/mol at 37 °C.

To test the generality of the conclusions from this work, thermodynamic parameters were measured for hairpin loops that naturally occur in small and large subunit rRNA. These results are listed in Table 3. The two hairpin loops of three were chosen with "unusually" stable GA and UU first mismatches. In both cases, no additional stability was observed. These results support the modeling of hairpin loops of three in a "loop" sequence independent manner (see Discussion).

DISCUSSION

Based on a study of hairpins with loops of six, a simple model to predict the stability of RNA hairpin loops was developed (Serra et al., 1994; Serra & Turner, 1995). This model suggests that the free energy increment of hairpin loops of $n > 3$ is given by the following equation:

Table 4: ΔG°_{37iL} for Hairpin Loop Formation in 1 M NaCl

loop size	ΔG°_{37i}(GC)[a]	ΔG°_{37i}(AU)	ΔG°_{37iL}(AU) − ΔG°_{37iL}(GC)
3[b]	4.8 (5)	5.7 (3)	+0.9
4	4.9 (5)	5.4 (4)	+0.5
5	5.0 (9)	5.8 (2)	+0.8
6	5.0 (17)	5.3 (7)	+0.3
7	5.0 (7)	5.8 (4)	+0.8
8	4.9 (3)	5.4 (3)	+0.5
9	5.5 (4)	6.4 (4)	+0.9
			average +0.6

[a] Average calculated ΔG°_{37iL} for measured hairpin loops (excluding C_n hairpins) (Gralla & Crothers, 1973; Riesner et al., 1973; Coutts et al., 1974; Groebe & Uhlenbeck 1988; Serra et al., 1993, 1994). Numbers in parentheses represent number of hairpins used in average. [b] For hairpin loops of three, ΔG°_{37iL} is loop sequence independent.

$$\Delta G^\circ_{37L}(n) = \Delta G^\circ_{37iL}(n) + \Delta G^\circ_{37mm} + 0.4(\text{if closed by AU or UA}) - 0.7(\text{if first mismatch is GA or UU}) \quad (1)$$

Here, $\Delta G^\circ_{37iL}(n)$ is the free energy for initiating a loop of n nucleotides, ΔG°_{37mm} is the free energy for the stacking interactions of the first mismatch with the closing base pair, +0.4 represents the penalty in free energy for closing a hairpin with an AU (or UA) instead of a GC (or CG) pair, and −0.7 is an additional stabilization for hairpins with GA or UU first mismatches. The $\Delta G^\circ_{37iL}(n)$ values used in these calculations for loop sizes other than six often represented a single experimental determination or estimated value.

The limited data on the role of hairpin loop size on stability has produced conflicting results. Early studies with C_n loops found hairpins with loops of six to be most stable (Gralla & Crothers, 1973). More recently, Grobe and Uhlenbeck (1988) found small RNA hairpin loops to be more stable. These differences have made modeling the stability of RNA hairpins on the basis of the loop size difficult.

The free energy measurements listed in Tables 1 and 3 when combined with previous measurements (Gralla & Crothers, 1973; Grobe & Uhlenbeck, 1988; Coutts et al., 1974; Reisner et al., 1973), allow more reliable values for predicting hairpin loop stability. Table 4 lists the average ΔG°_{37iL} for hairpin loops from three to nine nucleotides closed by either AU (or UA) or GC (or CG) base pairs. These results suggest the following model for improvement in the parameters for hairpin loop stability prediction:

$$\Delta G^\circ_{37L}(n) = \Delta G^\circ_{37iL}(n) + \Delta G^\circ_{37mm} + 0.6(\text{if closed by AU base pair}) - 0.7(\text{if first mismatch is GA or UU}) \quad (2)$$

Here, $\Delta G^\circ_{37iL}(n)$ for $n = 4-9$ is 4.9, 5.0, 5.0, 5.0, 4.9, and 5.5 kcal/mol, respectively.

It is surprising that hairpin loops from four to eight nucleotides have similar stabilities. It had been suggested (Haasnoot et al., 1985) on the basis of the physical constraints of the RNA helix that hairpins in this range have two favored loop sizes. Larger loops would be favored when the loop crossed the major groove and stacked on the 5′ side of the stem. Small loops of three or four nucleotides were predicted to cross the minor groove. However, it appears as though the energy differences are minimal, and therefore hairpin loops of four to eight have similar stabilities.

Grobe and Uhlenbeck (1988) have previously found that hairpins with loops of nine were less stable than hairpins with smaller loops. They attributed the differences in stability to heterogeneity in their samples. However, we observed a similar decreased stability for the hairpin loops of nine. It is unclear why loops of nine are less stable than smaller loops.

The cyclic nature of the difference in the stability of hairpins closed by AU (or UA) and GC (or CG) base pairs is interesting (Table 4). It is unclear why hairpin loops with odd numbers of nucleotides should display a greater difference in stability than hairpins with even numbers of nucleotides. At this time, the precision of the data prevents us from making too fine a distinction and therefore, we have chosen to average all of the $\Delta\Delta G^\circ_{37iL(AU-GC)}$ to arrive at the penalty for AU closure of 0.6 kcal/mol. Further measurements will be needed to determine if the cyclic fluctuation observed in $\Delta\Delta G^\circ_{37iL(AU-GC)}$ is real.

The improved parameters were used to predict the thermodynamics for the model hairpins in Table 1 and the naturally occurring hairpins in Table 3. In nearly all cases, there was good agreement between the measured and predicted ΔG°_{37L} values. The only two hairpins for which the model predicted a ΔG°_{37L} value greater than 1 kcal/mol from the experimentally measured ΔG°_{37L} value were the hairpin loops of eight closed by AU or UA (Table 1). It is unclear why these two hairpin loops have such different stabilities. In general, the nearest neighbor model allows prediction of RNA hairpin stability to within 5−10% of the experimentally measured values.

It is encouraging that the simple model is able to provide a reasonable prediction of hairpin stability. This is particularly so, because of the number of interactions that have been determined to occur in loop structures (Heus & Pardi, 1991; Borer et al., 1995). Small hairpin loops in Table 3 are modeled slightly better than larger loops. This may be a result of the greater structural flexibility available to larger loops.

The two naturally occurring hairpin loops of three used to test this model contained "unusually stable" first mismatches, GA or UU (Serra et al., 1994). Neither of these hairpins displayed any additional stabilization and could be best modeled as sequence independent, with no interaction between the closing base pair and first mismatch. In contrast to previous models (Jaeger et al., 1989; Serra & Turner, 1995) for hairpin loops of three, we found that the closing base pair does affect stability. The model to predict the stability for hairpin loops of three is "loop" sequence independent:

$$\Delta G^\circ_{37L}(3) = [4.8 + 0.6(\text{if closed by AU base pair})] \text{ kcal/mol} \quad (3)$$

The results in Tables 1 and 3 along with the limited experimental data previously available (Gralla & Crothers, 1973; Riesner et al., 1973; Coutts et al., 1974; Groebe & Uhlenbeck, 1988; Serra et al., 1993, 1994) provide a firm foundation for modeling the free energies of hairpin loops with eqs 2 and 3. The results should help improve predictions of RNA secondary structure from sequence.

ACKNOWLEDGMENT

We thank Dr. Douglas H. Turner for many stimulating discussions and careful reading of this manuscript.

REFERENCES

Borer, P. N., Lin, Y., Wang, S., Roggenbuck, M. W., Gott, J. M., Uhlenbeck, O. C., & Pelczer, I. (1995) *Biochemistry 34*, 6488−6503.

Borer, P. N. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, C. D., Ed.) 3rd ed., Vol. 1, p 589, CRC Press, Cleveland, OH.

Coutts, S. M., Gangloff, J., & Dirheimer, G. (1974) *Biochemistry 13*, 3938−3948.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J., & Ehresmann, B. (1987) *Nucleic Acids Res. 15*, 9109−9127.

Fountain, M. A., Serra, M. J., Krugh, T. R., & Turner, D. H. (1996) *Biochemistry 35*, 6539−6548.

Freier, S. M., Burger, B. J., Alkema, D., Neilson, T., & Turner, D. H. (1983) *Biochemistry 22*, 6198−6206.

Gutell, R. R. (1993) *Nucleic Acids Res. 21*, 3051−3054.

Gutell, R. R., Gray, M. W., & Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.

Haasnoot, C. A. G., de Bruin, S. H., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1985) *J. Biosci. 8*, 767−780.

Heus, H. A., & Pardi, A. (1991) *Science 253*, 191−194.

Huang, S., Wang, Y.-X., & Draper, D. E. (1996) *J. Mol. Bio. 258*, 308−321.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1990) *Methods Enzymol. 183*, 281−306.

Jaeger, J. A., SantaLucia, J., Jr., & Tinoco, I., Jr. (1993) *Annu. Rev. Biochem. 62*, 255−287.

Lazinski, D., Grzadienska, E., & Das, A. (1989) *Cell 59*, 207−218.

Marino, J. P., Gregorian, R. S., Jr. Csankovski, G., & Crothers, D. M. (1995) *Science 268*, 1448−1454.

Martinez, H. M. (1990) *Methods Enzymol. 183*, 306−317.

Murphy, F. L., Wang, Y., Griffith, J. D., & Ceck, T. R. (1994) *Science 265*, 1709−1712.

Parker, R. (1989) in *Methods in Enzymology* (Dahlberg, J. E., & Abelson, J. N., Eds.) Vol. 180, p 510, Academic Press, San Diego, CA.

Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 256−263.

Pley, H. W., Flaherty, K. M., & McKay, D. B. (1994) *Nature 372*, 68−74.

Reisner, D., Maass, G., Thiebe, R., Philippsen, P., & Zachau, H. G. (1973) *Eur. J. Biochem. 36*, 76−88.

Richards, E. G. (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids* (Fasman, C. D., Ed.) 3rd ed., Vol. 1, p 197, CRC Press, Cleveland, OH.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. D., Clark, B. F. C., & Klug, A. (1974) *Nature (London) 250*, 546−551.

Scott, W. G., Finch, J. T., & Klug, A. (1995) *Cell 81*, 991−1002.

Serra, M. J., & Turner, D. H. (1995) *Methods Enzymol. 259*, 242−261.

Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A., & Turner, D. H. (1993) *Nucleic Acids Res. 21*, 3845−3849.

Serra, M. J., Axenson, T. J., & Turner, D. H. (1994) *Biochemistry 33*, 14289−14296.

Sussman, J. L., Holbrook, S. R., Warrent, R. W., Church, G. M., & Kim, S.-H. (1978) *Nature 288*, 607−630.

Usman, N., Ogilvie, K. K., Jiang, M. Y., & Cedergren, R. L. (1987) *J. Am. Chem. Soc. 109*, 7845−7854.

Woese, C. R., & Pace, N. R. (1993) in *The RNA World* (Gesteland, R. F., & Atkins, J. F., Eds.) pp 91−117, Cold Spring Harbor Laboratory Press, Plainview, NY.

Woo, N. H., Roe, B. A., & Rich, A. (1980) *Nature 286*, 346−351.

Wu, H. N., & Uhlenbeck, O. C. (1987) *Biochemistry 26*, 8221−8227.

Wu, T., Ogilvie, K. K., & Pon, R. T. (1989) *Nucleic Acids Res. 17*, 3501−3517.

Zuker, M. (1989) *Science 244*, 48−52.

BI962608J