G·A and U·U Mismatches Can Stabilize RNA Internal Loops of Three Nucleotides†

Susan Schroeder, James Kim, and Douglas H. Turner*

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

Received July 19, 1996; Revised Manuscript Received October 8, 1996⊗

ABSTRACT: Optical melting experiments on oligoribonucleotides containing internal loops of three nucleotides (two nucleotides opposing one nucleotide) flanked by CG pairs show that the free energy of internal loop formation depends on the sequence of nucleotides within the loop. Internal loops with the potential for A·G and U·U mismatch pairing generally have more favorable free energies of formation than internal loops with potential AA, CC, AC, or UC pairings. In most cases, potential A·G and U·U pairings leaving a single unpaired nucleotide 5′ to a flanking CG pair are most stable; those leaving a single nucleotide 3′ to a flanking CG pair have intermediate stability. Internal loops of three nucleotides without potential A·G or U·U pairs are the least stable. The results provide an improved model for predicting the folding stabilities of internal loops of three nucleotides.

RNA plays a role in protein translation, catalysis, regulation of gene expression, and as an anti-sense therapeutic target (Watson et al., 1987). Knowledge of RNA secondary and tertiary structure can improve our understanding of RNA biological functions. Internal loops are an important secondary structural motif because they occur often and contain unpaired nucleotides with functional groups available to serve as binding sites for proteins or for tertiary interactions (Le et al., 1994; Egebjerg et al., 1988; Weeks & Crothers, 1993; Wyatt & Tinoco, 1993). Little is known, however, about the sequence dependence of folding stabilities for internal loops. Such thermodynamic information would improve computer predictions of secondary and tertiary structure based on free energy minimization (Walter et al., 1994a; Jaeger et al., 1989; Turner et al., 1988; Lück et al., 1996). Previous research has shown that G·A and U·U mismatches form stabilizing hydrogen bonds in internal loops of tandem mismatches (SantaLucia et al., 1991a,b; Walter et al., 1994b; Wu et al., 1995) and that asymmetric internal loops containing only A's (Peritz et al., 1991) or C's (Weeks & Crothers, 1993) have less thermodynamic stability than similar symmetric loops. In this paper, thermodynamic parameters for forming internal loops of three nucleotides (two nucleotides opposing one nucleotide) are reported for 17 new sequences containing a variety of mismatches. In most cases, A·G and U·U mismatch pairs also stabilize internal loops of three nucleotides.

MATERIALS AND METHODS

Oligoribonucleotides were synthesized on an ABI 392 DNA/RNA synthesizer using phosphoramidite chemistry (Usman et al., 1987; Scaringe et al., 1990). Incubation in 3:1 (v/v) ammonia:ethanol at 55 °C for 17 h cleaves the RNA from the CPG support and removes the base protecting group; incubation in triethylaminehydrogenfluoride (TEAHF) at 55 °C for 48 h removes the silyl protecting group on the 2′ hydroxyl. After desalting the RNA samples with a Sep-Pak C-18 cartridge, the product was separated from failure sequences on either a Baker 500Si TLC plate (60% 1-propanol, 35% ammonia, 5% H_2O solvent) or a denaturing 8 M urea, 20% polyacrylamide gel. After desalting the RNA, purity was checked by HPLC on a Beckman Ultrasphere C-8 column with a gradient from 10 mM sodium phosphate, pH 7 (buffer A), to 10 mM sodium phosphate, pH 7/50% methanol (buffer B) at a rate of 1 mL/min. Oligomers usually eluted at 20−30% buffer B. Purity of all oligomers was greater than 95%, except for 5′UGAGACGUCA3′ (>90%) and 5′UGAGAUGUCA3′ (>91%).

Concentrations of the single strand oligomers were determined from the absorbance at 280 nm at 80 °C (Borer, 1975). After mixing the oligomers in 1:1 concentration ratio, the RNA was dissolved in 1 M NaCl, 10 mM cacodylate, 0.5 mM EDTA, pH 7 melting buffer. Melting curves were measured at 280 nm on a Gilford 250 spectrophotometer with a heating rate of 1 °C/min controlled by a Gilford 2527 thermoprogrammer. The melting curves were fit to the two-state model with sloping baselines (Petersheim & Turner, 1983). Thermodynamic parameters were also obtained by fitting plots of inverse melting temperature, T_M^{-1}, vs log(total strand concentration/4) to the equation (Borer et al., 1974)

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

where R is the gas constant.

RESULTS

Table 1 lists thermodynamic parameters derived from melting experiments in 1 M NaCl for duplexes containing internal loops of three nucleotides. Duplexes are listed in order of decreasing stability. Figure 1 shows a typical T_M^{-1} vs $\ln(C_T/4)$ plot. Thermodynamic parameters derived from T_M^{-1} vs $\ln(C_T/4)$ plots and from curve fitting agree within 10%, consistent with the two-state transition model (Freier et al., 1986a; Turner et al., 1988). One sequence, $^{5'\mathrm{UGAC}\ \mathrm{A}\mathrm{CUCA}}_{\mathrm{ACUG_{AA}GAGU}}$, has been studied previously (Peritz et al., 1991), and the T_M^{-1} vs $\ln(C_T/4)$ results in Table 1 are similar to those reported. The new results are preferred, however, since the melts are more two-state than those reported

† This work was supported by NIH Grant GM22939.

* Author to whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1996.

S0006-2960(96)01789-8 CCC: $12.00

Table 1: Thermodynamic Parameters of Duplex Formation[a]

duplex	T_M^{-1} vs ln (C_T/4) parameters				curve fit parameters			
	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$(kcal/mol)	T_M[b](°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$\Delta G°_{37}$(kcal/mol)	T_M[b](°C)
5′UGAC U CUCU ACUG$_{CU}$GAGU	65.54±1.23	180.68±3.80	9.50±0.05	51.7	62.82±2.11	172.20±6.64	9.41±0.07	51.9
5′UGAG A GUCA ACUC$_{AG}$CAGU	53.11±3.28	143.05±10.1	8.74±0.15	50.5	55.41±3.83	150.03±11.7	8.88±0.21	50.7
5′UGAG A GUCA ACUC$_{GG}$CAGU	60.77±0.58	168.09±1.89	8.64±0.02	48.2	58.06±2.38	159.49±7.57	8.60±0.05	48.4
5′UGAC G CUCA ACUG$_{GA}$GAGU	66.82±0.68	188.61±2.41	8.32±0.02	45.5	65.88±3.86	185.51±12.1	8.34±0.11	45.8
5′UGAC A CUCA ACUG$_{AG}$GAGU	65.00±1.80	182.76±5.67	8.32±0.05	45.8	65.36±0.90	183.89±2.72	8.32±0.08	45.8
5′UGAC A CUCA ACUG$_{GA}$GAGU	60.19±1.59	168.24±5.01	8.01±0.04	44.8	65.41±4.88	184.75±15.6	8.11±0.12	44.7
5′UGAC G CUCA ACUG$_{AG}$GAGU	65.82±1.24	186.67±3.91	7.92±0.03	43.7	65.11±3.42	184.29±10.9	7.95±0.07	43.9
5′UGAC U CUCA ACUG$_{UU}$GAGU	69.41±1.02	198.33±3.23	7.90±0.02	43.3	63.60±2.03	179.81±6.68	7.83±0.06	43.5
5′UGAC U CUCA ACUG$_{UC}$GAGU	68.45±1.27	196.38±4.06	7.54±0.01	41.7	63.18±2.52	179.54±8.11	7.50±0.06	41.8
5′UGAC A CUCA ACUG$_{CC}$GAGU	65.20±1.14	186.29±3.63	7.42±0.01	41.3	61.47±3.31	174.24±10.7	7.43±0.07	41.6
5′UGAC G CUCA ACUG$_{AA}$GAGU	64.29±0.27	183.60±0.87	7.35±0.01	41.0	63.80±0.05	182.17±0.17	7.34±0.01	41.0
5′UGAC C CUCA ACUG$_{CU}$GAGU	61.43±0.86	174.99±2.76	7.16±0.01	40.2	60.43±3.22	171.67±10.4	7.19±0.04	40.4
5′UGAG A GUCA ACUC$_{GA}$CAGU	60.26±1.03	171.80±3.29	6.97±0.01	39.3	56.68±3.36	160.21±11.0	6.99±0.05	39.5
5′UGAC U CUCA ACUG$_{CC}$GAGU	56.87±0.97	161.27±3.13	6.86±0.01	38.8	54.93±2.45	154.90±8.25	6.88±0.07	39.0
5′UGAC A CUCA ACUG$_{AA}$GAGU	62.26±0.61 (57.98)[c]	178.83±1.97 (166.22)[c]	6.80±0.01 (6.42)[c]	38.4 (36.4)[c]	57.26±4.86 (46.11)[c]	162.51±15.9 (127.34)[c]	6.86±0.08 (6.61)[c]	38.8 (37.6)[c]
5′UGAC C CUCA ACUG$_{AA}$GAGU	61.14±1.46	175.22±4.73	6.78±0.01	38.3	58.82±5.05	167.6±16.5	6.84±0.08	38.7
5′UGAC C CUCA ACUG$_{CC}$GAGU	54.93±0.84	155.75±2.72	6.62±0.01	37.5	54.96±3.72	155.68±12.1	6.67±0.06	37.8
5′UGAC C CUCA ACUG$_{AU}$GAGU	56.26±0.92	160.10±2.99	6.61±0.01	37.4	57.94±5.65	165.26±18.3	6.68±0.06	37.8
5′UGAG A GUCA[c] ACUC$_{AA}$CAGU	55.76±5.76	158.51±15.8	6.59±0.20	37.4	45.07±4.51	123.66±12.4	6.71±0.20	38.3
5′UGACCUCA[c] ACUGGAGU	76.09±1.70	205.55±5.13	12.34±0.11	62.6	69.14±5.90	184.52±17.82	11.90±0.38	63.1

[a] Melting experiments were done in 1 M NaCl, 10 mM sodium cacodylate, 0.5 mM EDTA, pH 7.0, buffer solutions. Sequences are listed in order of decreasing loop free energy (see Table 3). Listed errors are standard deviations from reported measurements assuming no correlation of errors in the slope and intercept and are therefore overestimates of this source of error. Estimated errors from all sources are ±10%, ±10%, and ±2% for $\Delta H°$, $\Delta S°$, and $\Delta G°$, respectively. Significant figures beyond error estimates are given to allow accurate calculation of T_M and other parameters. The 10-mers in duplexes $^{5'UGAC\ ^{A}CUCA}_{3'ACUG_{AC}GAGU}$ and $^{5'UGAC\ ^{A}CUCA}_{3'ACUG_{AC}GAGU}$ form self-complementary duplexes with higher T_M's than the duplexes containing an internal loop of three nucleotides. [b] Melting temperatures are given for total strand concentrations of 1×10^{-4} M. [c] Values are from Peritz et al. (1991).

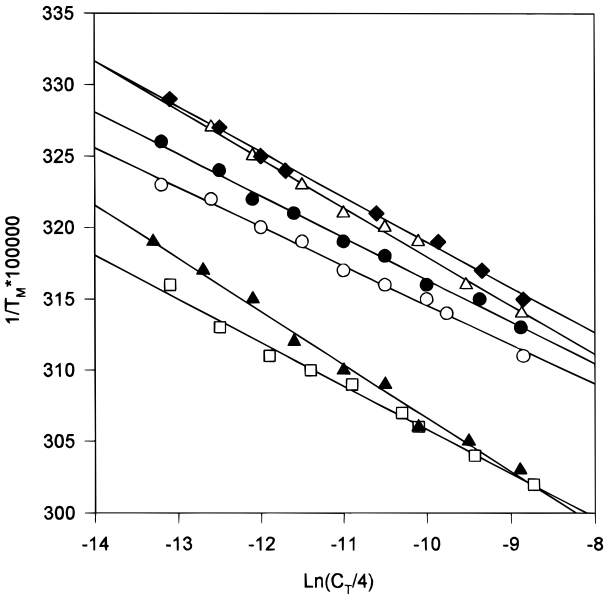

FIGURE 1: Inverse temperature vs ln (C_T/4) plot. (◆) A-AA, (△) A-GA, (●) G-AA, (○) U-UU, (▲) A-AG, (□) U-CU.

previously, suggesting a more pure sample. The duplex $^{5'UGAC\ ^{G}CUCA}_{ACUG_{AA}GAGU}$ was synthesized twice, and results of melting experiments on both samples are within experimental error. The duplexes $^{5'UGAC\ ^{A}CUCA}_{ACUG_{AA}GAGU}$ and $^{5'UGAC\ ^{G}CUCA}_{ACUG_{AA}GAGU}$ at roughly 6 $\times$ 10^{-4} M strand concentration were also melted in 10 mM $MgCl_2$, 0.15 M KCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA buffer at pH 7. The thermodynamic parameters were within experimental error of those listed in Table 1.

Table 2 lists melting data for single strand oligomers that gave cooperative melts. The sequences 5′UGAGCAGUCA and 5′UGAGACGUCA form self-complementary duplexes with melting temperatures higher than the duplexes containing internal loops $^{A}_{3'CA5'}$ or $^{A}_{3'AC5'}$. Thus these internal loops could not be studied. Melting temperatures for other single strands were low enough so that the self-complementary association can be neglected (Peritz et al., 1991).

Table 3 lists thermodynamic parameters for internal loop formation. The value for the free energy of internal loop formation is calculated according to the following equation (Gralla & Crothers, 1973):

$$\Delta G^\circ_{\text{loop}} = \Delta G^\circ_{\text{duplex with loop}} - (\Delta G^\circ_{\text{duplex without loop}} - \Delta G^\circ_{\text{interrupted base pair}}) \quad (2)$$

For example,

$$\Delta G^\circ\left(\begin{smallmatrix}C^{A}C\\G_{AG}G\end{smallmatrix}\right) = \Delta G^\circ\left(\begin{smallmatrix}5'UGAC\ ^{A}CUCA\\ACUG_{AG}GAGU\end{smallmatrix}\right) - \Delta G^\circ\left(\begin{smallmatrix}5'UGACCUCA\\ACUGGAGU\end{smallmatrix}\right) + \Delta G^\circ\left(\begin{smallmatrix}CC\\GG\end{smallmatrix}\right) \quad (3)$$

The values for enthalpy and entropy changes are similarly derived. The last term in eq 3 corrects for the nearest neighbor stacking interaction present in the perfectly paired helix but absent in the helix with an internal loop. Thus the thermodynamic parameters calculated are appropriate for the nearest neighbor model. For a simple pairing model, the last term would be omitted. Available data, however, indicate the nearest neighbor model is a better approximation for helix stability (Turner et al., 1988).

The natural occurrence of internal loops of three was studied by searching through known secondary structures of 89 small subunit rRNAs (Gutell, 1994), 78 large subunit rRNAs (Gutell et al., 1993), and 71 group I introns (Damberger & Gutell, 1994). The results are shown in Table 4.

Table 2: Melting Data for Single Strand Oligomers[a]

oligomer	concentration ($\times 10^{-4}$ M)	T_M (°C)	ΔH° (kcal/mol)	ΔS° (eu)
5′UGAGACGUCA	2.73	64.8	68.91	187.6
	0.132	55.1	65.22	176.4
5′UGACGGCUCA	3.58	42.8	45.79	129.1
	0.145	31.1	49.87	141.8
5′UGAGCCGUCA	3.78	40.1	41.54	117.0
	0.150	24.7	28.41	73.32
5′UGACGACUCA	3.59	36.4	63.71	190.1
	0.146	26.7	48.30	138.9
5′UGACAGCUCA	2.59	36.4	38.89	109.2
	0.101	22.7	26.50	66.71
5′UGACGCUCA	3.87	36.0	45.35	131.1
	0.145	22.8	33.96	92.61
5′UGAGUCGUCA	3.84	35.0	48.78	142.6
	0.170	22.8	39.76	112.5
5′UGAGCAGUCA	2.02	34.8	65.37	195.4
	0.100	25.7	47.11	134.8
5′UGAGUAGUCA	3.20	31.1	61.93	187.6
	0.130	18.8	37.64	106.6
5′UGAGAGGUCA	3.77	28.2	26.72	88.65
	0.695	24.5	26.84	90.20
5′UGAGGAGUCA	2.80	29.4	31.27	103.4
	1.52	27.5	29.76	98.99
5′UGACUCUCA	4.32	21.1	42.60	129.37
5′UGAGAGUCA[b]	1.6	22		

[a] Buffer is 1 M NaCl, 10 mM sodium cacodylate, 0.5 mM Na_2 EDTA, pH 7.0. Error in T_M is estimated as ±1 °C. Errors in ΔH° and ΔS° are estimated as ±20%. [b] Peritz et al. (1991).

DISCUSSION

Internal loops of 3 nucleotides are found in known secondary structures of RNA (Table 4). Thus information on the folding stabilities of these motifs will be helpful for predicting the structure of an RNA from its sequence. In this work, we report folding stabilities for internal loops of 3 nucleotides closed by CG pairs. There are 26 possibilities for mismatch sequences in such loops when G·U pairs are excluded. G·U pairs are excluded since most folding algorithms treat them as canonical base pairs. This paper reports results on 16 mismatch combinations, and the results allow assignment of approximate stability increments for 9 of the remaining 10 combinations. The exception is the $^{G}_{3'GG5'}$ motif. The $^{G}_{3'GG5'}$ motif was not studied because the necessary sequences would contain at least three consecutive Gs. Such sequences tend to aggregate (SantaLucia et al., 1991b). Table 5 summarizes all the stability increments derived from this work. They are discussed below.

G·A and U·U Mismatches Can Stabilize Internal Loops of 3 Nucleotides. As shown in Table 5, there is a considerable sequence dependence to the stabilities of internal loops of 3. The free energy increments range from −0.1 to 2.8 kcal/mol, corresponding to a 100-fold range in association

Table 3: Thermodynamic Parameters for Internal Loop Formation[a]

duplex	ΔG°_{37} (kcal/mol)	ΔH° (kcal/mol)	ΔS° (eu)
5′UGAC ^{U} CUCA ACUG$_{CU}$GAGU	-0.09±0.24	−1.65±2.25	−4.83±6.75
5′UGAG ^{A} GUCA ACUC$_{AG}$CAGU	0.67±0.27	10.78±4.00	32.8±11.90
5′UGAG ^{A} GUCA ACUC$_{GG}$CAGU	0.77±0.23	−3.12±2.30	7.76±6.56
5′UGAC ^{G} CUCA ACUG$_{GA}$GAGU	1.09±0.23	−2.93±2.00	−12.76±6.07
5′UGAC ^{A} CUCA ACUG$_{AG}$GAGU	1.09±0.23	−1.11±2.60	−6.91±7.95
5′UGAC ^{A} CUCA ACUG$_{GA}$GAGU	1.40±0.23	3.70±2.46	7.61±7.50
5′UGAC ^{G} CUCA ACUG$_{AG}$GAGU	1.49±0.23	−1.93±2.25	−10.80±6.81
5′UGAC ^{U} CUCA ACUG$_{UU}$GAGU	1.51.±0.23	−5.52±2.14	−22.48±6.44
5′UGAC ^{U} CUCA ACUG$_{UC}$GAGU	1.87±0.23	4.56±2.27	−20.53±6.90
5′UGAC ^{A} CUCA ACUG$_{CC}$GAGU	1.99±0.23	−1.31±2.20	−10.44±6.65
5′UGAC ^{G} CUCA ACUG$_{AA}$GAGU	2.06±0.23	−0.41±2.20	−7.75±5.64
5′UGAC ^{C} CUCA ACUG$_{CU}$GAGU	2.25±0.23	2.46±1.90	0.86±6.22
5′UGAG ^{A} GUCA ACUC$_{GA}$CAGU	2.44±0.23	3.63±2.44	4.05±7.09
5′UGAC ^{U} CUCA ACUG$_{CC}$GAGU	2.55±0.23	7.02±2.11	14.58±6.40
5′UGAC ^{A} CUCA ACUG$_{AA}$GAGU	2.61±0.23	1.63±1.98	−2.98±5.91
5′UGAC ^{C} CUCA ACUG$_{AA}$GAGU	2.63±0.23	2.75±2.23	0.564±7.31
5′UGAC ^{C} CUCA ACUG$_{CC}$GAGU	2.79±0.23	8.96±2.06	20.1±6.20
5′UGAC ^{C} CUCA ACUG$_{AU}$GAGU	2.80±0.23	7.63±2.09	15.75±6.32
5′UGAG ^{A} GUCA[b] ACUC$_{AA}$CAGU	2.82±0.42	8.13±6.17	17.34±17.00

[a] Buffer is 1 M NaCl, 10 mM sodium cacodylate, 0.5 mM Na_2EDTA, pH 7.0. Values listed are from T_M^{-1} vs ln (C_T/4) plots. Errors assume ±10% errors for ΔH° and ΔS° of $^{5'GG3'}_{3'CC5'}$. [b] Values from Peritz et al. (1991).

constants. Except for $^{U}_{3'CU5'}$, all the internal loops of 3 destabilize the duplex. The nine most stable sequences have possible G·A and U·U pairings (Table 3). G·A and U·U pairings are known to stabilize tandem mismatches due to hydrogen bonding between the mismatched bases (SantaLucia et al., 1991a,b; SantaLucia & Turner, 1993; Walter et al., 1994b; Wu et al., 1995). The thermodynamic results in Table 3 suggest internal loops of 3 can accommodate similar

Table 4: Natural Occurrence of Internal Loops of 3 Nucleotides, $^{X}_{3'ZY5'}$ [a]

ZY	X = A	ZY	X = C	ZY	X = G
AA	110 (42)[b]	AA	12(1)[b]	AA	42(14)[b]
AC	13	AC	3	AG	7
				GA	6(2)
AG	94	AU	65	GG	3
CA	12(1)[b]	CA	2		X = U
CC	9(1)[b]	CC	4(2)[b]	CC	4
CG	7	CU	1(1)[b]	CU	3
GA	4	UA	5		
GC	14	UC	2	UC	35
GG	49	UU	16	UU	14(3)[b]

[a] Occurrence in 89 small subunit rRNAs (Gutell, 1994), 78 large subunit rRNAs (Gutell et al., 1993), and 71 group I introns (Damberger & Gutell, 1994). [b] Values in parentheses are number of occurrences of $^{CXC}_{G_{YZ}G}$ motif.

Table 5: Free Energy Increments (kcal/mol) at 37 °C in 1 M NaCl, pH 7, for $^{X}_{3'ZY5'}$ Internal Loops Closed by CG Pairs

3′ZY5′	X = A	3′ZY5′	X = C	3′ZY5′	X = G
AA	2.6 2.8	AA	2.6	AA	2.1
AC	(2.6)[a]	AC	(2.6)[a]	AG	1.5
AG	1.1 0.7	AU	2.8		
				GA	1.1
CA	(2.6)[a]	CA	(2.6)[a]	GG	−
CC	2.0	CC	2.8		X = U
CG	(0.9)[a]	CU	2.3	CC	2.6
GA	1.4 2.4	UA	(2.6)[a]	CU	−0.1
GC	(1.9)[a]	UC	(2.6)[a]	UC	1.9
GG	0.8	UU	(2.6)[a]	UU	1.5

[a] Estimated from measurements on other sequences.

pairing schemes. Two exceptions are observed, however, in the 11 sequences studied with potential G·A and U·U pairs. The motifs $^{CGC}_{GAAG}$ and $^{GAG}_{CGAC}$ are not significantly more stable than internal loops without G·A and U·U mismatches. Thus there are some unexplained subtleties to the sequence dependence of stability.

Approximations for Unmeasured Sequences. The available data allow approximations for $^{A}_{CG}$ and $^{A}_{GC}$, the only unmeasured sequences with potential G·A and U·U pairs. Three internal loops were studied in both $^{C\ XC}_{GZYG}$ and $^{G\ XG}_{CZYC}$ motifs (see Tables 3 and 5). The largest difference in stability increment is 1.0 kcal/mol and the average difference is 0.6 kcal/mol. To a first approximation, therefore, these two motifs are assumed to have the same stability. The motif $^{A}_{CG}$ is most similar to the measured motif of $^{A}_{AG}$ in that there is only one potential G·A pair and it contains an unpaired nucleotide 3′ to the G·A pair. Thus the average of the two measurements of the $^{A}_{AG}$ motif is used to provide the approximate increment for $^{A}_{CG}$ of 0.9 kcal/mol listed in Table 5. In a similar manner, the motif $^{A}_{GC}$ is most similar to the measured motif of $^{A}_{GA}$, except the unpaired nucleotide is now 5′ to the G·A pair. Again, two measurements are averaged to provide an approximation of 1.9 kcal/mol for the $^{A}_{GC}$ motif in Table 5.

Eight internal loops without potential G·A or U·U pairs were studied. Except for G·U and GG, all the other potential single mismatch pairs are represented. The average free energy increment for the internal loops without G·A or U·U pairs is 2.6 kcal/mol with a standard deviation of 0.1 kcal/mol. Thus the stabilities of these loops without G·A or U·U

mismatches fall in a relatively narrow range. Similar results have been observed with tandem mismatches (SantaLucia et al., 1991b; Wu et al., 1995). Thus a free energy increment of 2.6 kcal/mol is used in Table 5 for the seven unmeasured loop sequences without potential G·A or U·U mismatches.

Comparisons with Approximations Used in RNA Folding Algorithms. The measured values in Table 3 can be compared with the most recent approximation for the free energy increment for an internal loop of 3 nucleotides (Serra & Turner, 1995): $\Delta G^\circ_{37} = 5.4 + \Delta G^\circ_{37}(\text{mm}) + \Delta G^\circ_{37}(\text{de})$. Here $\Delta G^\circ_{37}(\text{mm})$ is the free energy increment for the mismatch in the loop. This is approximated as −2.7, −2.5, and −1.5 kcal/mol, respectively, for G·A, U·U, and other mismatches. The $\Delta G^\circ_{37}(\text{de})$ term is the free energy increment for the unpaired nucleotide as measured for dangling ends on short helixes (Turner et al., 1988). This approximation predicts that internal loops that allow G·A or U·U pairs that leave an unpaired nucleotide 5′ of the mismatch pair will be more stable than those that leave the unpaired nucleotide 3′ of the mismatch. The results in Table 3 show the opposite trend. On average, the predicted free energy increments for loops with an unpaired nucleotide 3′ of a G·A or U·U mismatch are 1.5 kcal/mol less favorable than those measured. For loops without G·A or U·U pairs, the predicted free energy increments are on average 0.6 kcal/mol less favorable than those measured. Thus the previous model, which was based on only two experimental measurements, was not particularly good.

Stability Increments for Mismatches Depend on Context. The ultimate goal of this work is to predict the folding stability of an RNA from its sequence. A large fraction of nucleotides in known RNA secondary structures are not in Watson−Crick or G·U base pairs, but little is known about the sequence dependence of folding stability for motifs other than Watson−Crick and G·U pairs. The number of potential combinations of different mismatches in various types of loops precludes thermodynamic measurements on all possibilities. Thus models must be developed that allow prediction of stability for various motifs. The results in Table 3 indicate that internal loops of 3 containing only potential AA, AC, CC, and CU mismatches have similar stabilities. Symmetric tandem mismatches have shown similar results (Wu et al., 1995). Thus it should be reasonably straightforward to predict stabilities of loops containing only these mismatches. G·A and U·U mismatches, however, have different stabilities depending on context. For example, the free energy increments at 37 °C for $^{\text{U}}_{\text{CU}}$ and $^{\text{U}}_{\text{UC}}$ are −0.1 and 1.9 kcal/mol, respectively (see Table 3). While G·A or U·U mismatches are almost always more stable than others, the differences between the free energy increments for G·A or U·U and other mismatches are also context dependent. For example, in comparing G·A and A·A mismatches adjacent to CG pairs in RNA, the average difference in free energy increment per mismatch is 0.1, 0.5, 1.3, and 1.7 kcal/mol for terminal mismatches (Turner et al., 1988), hairpins (Serra et al., 1994), internal loops of 3 (Table 3), and tandem mismatches (Wu et al., 1995), respectively. A similar comparison between U·U and U·C mismatches gives average differences of −0.2, 0.2, 1.0, and 0.9 kcal/mol, respectively. Thus there is an emerging trend that the advantage for G·A and U·U mismatches in promoting folding is larger when the mismatch is shielded from water and/or is more constrained by its environment. In principle, theoretical methods like free energy perturbation (McCammon & Harvey, 1987) should be able to predict the various sequence dependences by predicting the effect of changing individual nucleotides. The results in Table 3 provide useful benchmarks for testing such calculations.

REFERENCES

Borer, P. N., (1975) in *Handbook of Biochemistry and Molecular Biology: Nucleic Acids*, 34d ed. (Fasman, G. D., Ed.) Vol. 1, p 597, CRC Press, Cleveland, OH.

Borer, P. N. Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843−853.

Damberger, S. H., & Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.

Egebjerg, J., Christiansen, J., Brown, R. S., Larsen, N., & Garrett, R. A. (1989) *J. Mol. Biol. 206*, 651−668.

Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373−9377.

Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., & Turner, D. H. (1986b) *Biochemistry 25*, 3214−3219.

Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301−319.

Damberger, S. H., & Gutell, R. R. (1994) *Nucleic Acids Res. 21*, 3508−3510.

Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3502−3507.

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

Han, K., & Kim. H. (1993) *Nucleic Acids Res. 21*, 1251−1257.

Jaeger, J. A., Turner, D. H., & Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.

Le, S.-Y., Pattabiraman, N., & Maizel, J. V., Jr. (1994) *Nucleic Acids Res. 22*, 3966−3976.

Lück, R., Steger, G., & Riesner, D. (1996) *J. Mol. Biol. 258*, 813−826.

McCammon, J. A., & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, U.K.

Peritz, A. E., Kierzek, R., Sugimoto, N., & Turner, D. H. (1991) *Biochemistry 30*, 6428−6436.

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

SantaLucia, J., Jr., & Turner D. H. (1993) *Biochemistry 32*, 12612−12623.

SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991a) *J. Am. Chem. Soc. 113*, 4313−4322.

SantaLucia, J., Jr., Kierzek, R., & Turner, D. H. (1991b) *Biochemistry 30*, 8242−8251.

Scaringe, S. A., Francklyn, C., & Usman, N. (1990) *Nucleic Acids Res. 18*, 5433−5441.

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

Turner, D. H., Sugimoto, N., & Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.

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

Walter, A. E., Turner, D. H., Kim, J., Lyttle, M. H., Müller, P., Matthews, D. H., & Zuker, M. (1994a) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9218−9222.

Walter, A. E., Wu, M., & Turner, D. H. (1994b) *Biochemistry 33*, 11349−11354.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., & Weiner, A. M. (1987) *Molecular Biology of the Gene*, 4th ed., Benjamin/Cummings Publishing Co., Menlo Park, CA.

Weeks, K. M., & Crothers, D. M. (1993) *Science 261*, 1574−1577.

Wu, M., McDowell, J. A., & Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.

Wyatt, J. R., & Tinoco, I., Jr. (1993) in *The RNA World* (Gesteland, R. F., & Atkins, J. F., Eds.) pp 465−496, Cold Spring Harbor Laboratory Press, Plainview, New York.

BI961789M