

Thermodynamic Characterization of Naturally Occurring RNA Single Mismatches with G-U Nearest Neighbors[†]

Amber R. Davis and Brent M. Znosko*

Department of Chemistry, Saint Louis University, Saint Louis, Missouri 63103

Received March 19, 2008; Revised Manuscript Received May 23, 2008

ABSTRACT: Due to their prevalence and roles in biological systems, single mismatches adjacent to G-U pairs are important RNA structural elements. Since there are only limited experimental values for the stability of single mismatches adjacent to G-U pairs, current algorithms using free energy minimization to predict RNA secondary structure from sequence assign predicted thermodynamic values to these types of single mismatches. Here, thermodynamic data are reported for frequently occurring single mismatches adjacent to at least one G-U pair. This experimental data can be used in place of predicted thermodynamic values in algorithms that predict secondary structure from sequence using free energy minimization. When predicting the thermodynamic contributions of previously unmeasured single mismatches, most algorithms apply the same thermodynamic penalty for an A-U pair adjacent to a single mismatch and a G-U pair adjacent to a single mismatch. A recent study, however, suggests that the penalty for a G-U pair adjacent to a tandem mismatch should be 1.2 ± 0.1 kcal/mol, and the penalty for an A-U pair adjacent to a tandem mismatch should be 0.5 ± 0.2 kcal/mol [Christiansen, M. E. and Znosko, B. M. (2008) *Biochemistry* 47, 4329–4336]. Therefore, the data reported here are combined with the existing thermodynamic dataset of single mismatches, and nearest neighbor parameters are derived for an A-U pair adjacent to a single mismatch (1.1 ± 0.1 kcal/mol) and a G-U pair adjacent to a single mismatch (1.4 ± 0.1 kcal/mol).

The occurrence of the G-U non-Watson–Crick base pair was first proposed by Francis Crick in his wobble hypothesis for codon–anticodon interactions (1). Crick realized that this base pair was able to form two hydrogen bonds between the Watson–Crick faces of guanosine and uridine. Since this initial hypothesis, G-U pairs have been found in virtually every class of functional RNA (2–11). Due to its unique chemical and structural properties (7, 12), the G-U pair has been shown to be an essential component in RNA secondary and tertiary structures and important in various aspects of recognition by a range of different biomolecules (7).¹

Due to the prevalence and biological importance of G-U pairs, there have been several thermodynamic studies on G-U pairs within a Watson–Crick duplex (13–17) and at the helix termini (13, 18, 19). From this data, nearest neighbor parameters for G-U pairs have been derived (16, 20, 21). However, G-U pairs located adjacent to internal loops are also important. In small and large subunit rRNA, 42% of G-U pairs occur at the loop–helix junction (9), and many are important for structure and function (22–26). Therefore, many studies have also thermodynamically characterized G-U pairs adjacent to internal loops (27–36).

Single mismatches adjacent to G-U pairs have been found in a wide variety of organisms (37–42). Only two studies, however, have thermodynamically characterized single mismatches adjacent to G-U pairs (34, 36). One of these studies was from the Turner laboratory (34), and the other study was recently published by Davis and Znosko (36). In the former, G-U pairs are considered noncanonical base pairs; therefore, the loop $\begin{bmatrix} 5'CUUG3' \\ 3'GGUC5' \end{bmatrix}$ is considered to be a tandem mismatch. Here and in the latter, G-U pairs are considered canonical pairs, so $\begin{bmatrix} 5'CUUG3' \\ 3'GGUC5' \end{bmatrix}$ is considered a single mismatch with a G-U nearest neighbor.

Since there are only limited experimental values for the stability of single mismatches adjacent to G-U pairs, current models using free energy minimization to predict RNA secondary structure from sequence, such as *RNAstructure* (21, 43, 44), *mfold* (45, 46), and the Vienna software package (47), assign predicted thermodynamic values to G-U pairs adjacent to most single mismatches. When calculating these predicted thermodynamic values, a 0.7 kcal/mol penalty is applied for each A-U or G-U closure (44). Recently, a new model to predict the thermodynamic contribution of single mismatches to duplex stability was proposed, and the penalty applied for each A-U or G-U closure was found to be 1.2 kcal/mol (36). Another study, however, suggests that the penalty for a G-U pair adjacent to a tandem mismatch should be 1.2 ± 0.1 kcal/mol, and the penalty for an A-U pair adjacent to a tandem mismatch should be 0.5 ± 0.2 kcal/mol (35). To determine if there

[†] Partial funding for this project was provided by the Saint Louis University College of Arts and Sciences, Saint Louis University Department of Chemistry, a Saint Louis University Summer Research Award (B.M.Z.), and the Saint Louis University Beaumont Faculty Development Fund (B.M.Z.). A.R.D. is supported by a Monsanto Scholars Graduate Fellowship.

* To whom correspondence should be addressed. Phone: (314) 977-8567. Fax: (314) 977-2521. E-mail: znoskob@slu.edu.

¹ Abbreviations: R, purine nucleotides; Y, pyrimidine nucleotides.

should be unique thermodynamic penalties for A-U and G-U pairs adjacent to single mismatches or if one penalty is sufficient for both types of adjacent base pairs, additional single mismatches with at least one G-U nearest neighbor were investigated. In this work, thermodynamic data are reported for single mismatches adjacent to G-U pairs that are common in nature. This experimental data can be used in place of predicted thermodynamic values in programs such as *RNAstructure*, *mfold*, and the Vienna software package. Also, based on new and existing thermodynamic data (34, 36, 48–51), A-U and G-U pairs adjacent to single mismatches are treated separately, and thermodynamic penalties are derived for A-U pairs adjacent to single mismatches and G-U pairs adjacent to single mismatches. Thus, the results presented here provide thermodynamic data that may be beneficial when designing therapeutics to bind RNA and should improve both the prediction of RNA secondary structure from sequence using free energy minimization and the prediction of RNA secondary structure by algorithms which combine structural and thermodynamic data (52).

MATERIALS AND METHODS

Compiling and Searching a Database for Single Mismatches with G-U Nearest Neighbors. A database of RNA secondary structures was compiled previously (21, 36, 43). This database was searched for single mismatches with G-U nearest neighbors, and the number of occurrences for each type of mismatch was tabulated.

RNA Synthesis and Purification. Sequences of mismatches and nearest neighbors were designed to represent those found most frequently in the database. Further details for the design of sequences were described previously (36, 53). Oligonucleotides were ordered from the Keck Laboratory at Yale University (New Haven, CT), Azco BioTech, Inc. (San Diego, CA), or Integrated DNA Technologies (Coralville, IA). The synthesis and purification of the oligonucleotides followed standard procedures that were described previously (36, 54).

Optical Melting Experiments and Thermodynamics. The methods used to determine the concentration of the single strands and to form duplexes from the single strands are standard and were described previously (36, 53). Optical melting experiments were performed in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na₂EDTA (pH 7.0). Melting curves (absorbance versus temperature) were obtained and duplex thermodynamics were determined as described previously (36). The thermodynamic contributions of single mismatches to duplex thermodynamics ($\Delta G^{\circ}_{37, \text{single mismatch}}$, $\Delta H^{\circ}_{\text{single mismatch}}$, and $\Delta S^{\circ}_{\text{single mismatch}}$) were calculated by subtracting the nearest neighbor contributions for the canonical pairs (21, 55) from the measured duplex thermodynamics. This type of calculation was described previously (36).

Linear Regression and Single Mismatch Thermodynamic Parameters. Data collected for 16 duplexes in this study were combined with previously published data for 77 single mismatches (34, 36, 48–50). Of the 93 total duplexes, eight melted in a non-two-state manner and were not included in trends, averages, or linear regression. In addition, data from ten sequences were significantly different from what was predicted. Eight of these sequences were discussed previously

(36). The other two sequences are $\begin{bmatrix} 5' \text{GAC GAU CUG} 3' \\ 3' \text{CUG CCG GAC} 5' \end{bmatrix}$ and $\begin{bmatrix} 5' \text{CAG CCG GUC} 3' \\ 3' \text{GUC GUU CAG} 5' \end{bmatrix}$. For these two duplexes, nearest neighbor calculations suggest that the bimolecular association of one of the oligoribonucleotide strands with itself may be competing with the bimolecular association of the two different strands. For example, $\begin{bmatrix} 5' \text{GAC GAU CUG} 3' \\ 3' \text{CUG CCG GAC} 5' \end{bmatrix}$ is predicted to have a ΔG°_{37} of -7.4 kcal/mol, and the bimolecular association of the bottom strand with itself (which forms four G-C pairs in a row) is predicted to have a ΔG°_{37} of -9.3 kcal/mol. Because the formation of $\begin{bmatrix} 5' \text{GAC GAU CUG} 3' \\ 3' \text{CUG CCG GAC} 5' \end{bmatrix}$ was not confirmed, the data for these duplexes were not included in trends, averages, and linear regression. Because three duplexes melted in a non-two-state manner and had possible competing structures, the thermodynamic parameters for 78 (64 reported previously and 14 reported here) single mismatches were included in the linear regression used to derive single mismatch-specific nearest neighbor parameters. Similar to what was done previously (36), three parameters consisting of a total of nine variables were used for linear regression: (1) a mismatch parameter containing variables for an A•G, G•G, or U•U mismatch; (2) a stacking parameter containing variables for $\begin{bmatrix} 5' \text{YRR} 3' \\ 3' \text{RRY} 5' \end{bmatrix}$, $\begin{bmatrix} 5' \text{YYY} 3' \\ 3' \text{YYY} 5' \end{bmatrix}$, $\begin{bmatrix} 5' \text{YRY} 3' \\ 3' \text{RYR} 5' \end{bmatrix}$, and $\begin{bmatrix} 5' \text{RRY} 3' \\ 3' \text{YYR} 5' \end{bmatrix}$ stacking combinations, when cytosine and uracil are classified as pyrimidines (Y) and adenine and guanine are classified as purines (R); and (3) a parameter for an A-U/G-U closure. The calculated experimental contribution of the single mismatch to duplex stability was used as a constant when doing linear regression. To simultaneously solve for each variable, the LINEST function of *Microsoft Excel* used linear regression. To determine if a G-U pair adjacent to a single mismatch should be assigned a different penalty than an A-U pair adjacent to a single mismatch, the linear regression described above was repeated with the third parameter split into two different parameters, one for an A-U pair adjacent to a single mismatch and one for a G-U pair adjacent to a single mismatch. The model derived previously which used the previous dataset and treated G-U pairs adjacent to single mismatches the same as A-U pairs adjacent to single mismatches (36), the same model incorporating the additional data reported here, and a model that treats A-U and G-U pairs adjacent to single mismatches separately are compared.

RESULTS

Database Searching. The database containing 955 RNA secondary structures and 151,503 nucleotides was searched for single mismatches adjacent to at least one G-U pair. In this database, 888 single mismatches with at least one G-U nearest neighbor were found, accounting for 3.5% of the nucleotides in the database. Table 1 shows a summary of the database results obtained. The first set of data lists frequency and percent occurrence when the mismatch nucleotides and nearest neighbors are specified. When categorized in this fashion, 95 types of mismatches were found in the database. The 30 mismatch types listed in dataset 1 (Table 1) account for 75% of the 888 mismatches found in the database. The 65 types of mismatches not shown account for the remaining 25%; however, each type repre-

Table 1: Summary of Database Search Results for Single Mismatches Adjacent to at Least One G-U Pair^a

Dataset 1, Single Mismatch with Nearest Neighbors				Dataset 2, Single Mismatch				Dataset 3, 5' and 3' Adjacent Base Pairs			
mismatch ^b	freq ^c	% ^d	ref	mismatch ^b	freq ^c	% ^d	ref	closing bp	freq ^c	% ^d	ref
CUG GUU	104	11.7	<i>ef</i>	A G	300	33.8	<i>e-g</i>	C G G U	201	22.6	<i>e-g</i>
AAU UGG	89	10.0	<i>f</i>	U U	254	28.6	<i>e-g</i>	G G C U	146	16.4	<i>e-g</i>
GUG CUU	43	4.8	<i>ef</i>	A C	131	14.8	<i>e-g</i>	A U U G	128	14.4	<i>f,g</i>
UAC GGG	40	4.5	<i>f</i>	C U	83	9.4	<i>e,g</i>	C U G G	88	9.9	<i>e,g</i>
UAG GGC	38	4.3	<i>ef</i>	C C	41	4.6	<i>e,g</i>	U G A U	77	8.7	<i>g</i>
UAA GCU	28	3.1	<i>g</i>	A A	41	4.6	<i>e</i>	G U C G	71	8.0	<i>e,g</i>
AUG UUU	27	3.0	<i>g</i>	G G	38	4.3		A G U U	61	6.9	<i>g</i>
CAU GGG	26	2.9	<i>g</i>	previously	850	95.7		G A U U	54	6.1	<i>g</i>
UUG AUU	26	2.9	<i>g</i>	new total	850	95.7		G G U U	27	3.0	<i>g</i>
GAG CGU	20	2.2	<i>e</i>					U G G U	24	2.7	<i>g</i>
GUU CUG	18	2.0	<i>e</i>					G U U G	11	1.2	
CCU GCG	15	1.7	<i>g</i>					previously	634	71.4	
CAG GCU	14	1.6	<i>e</i>					new total	877	98.8	
GAU CCG	14	1.6	<i>g</i>								
GAA UGU	13	1.5	<i>g</i>								
UAU AGG	12	1.3	<i>g</i>								
UAG GGU	12	1.3	<i>g</i>								
GAC UGG	11	1.2	<i>e</i>								
GCG UUC	11	1.2	<i>e</i>								
AUU UUG	11	1.2	<i>g</i>								
CUU GUG	11	1.2	<i>e</i>								
GAG CCU	10	1.1	<i>g</i>								
UAG ACU	10	1.1	<i>g</i>								
GCU CUG	10	1.1	<i>g</i>								
GCU UUA	10	1.1	<i>g</i>								
GAG CAU	9	1.0	<i>e</i>								
CCG GUU	9	1.0	<i>g</i>								
CAG GAU	8	0.9	<i>e</i>								
GAG UCU	8	0.9	<i>g</i>								
CAG GGU	8	0.9	<i>e</i>								
previously	439	49.4									
new total	680	76.6									

^a Not all combinations in dataset 1 are shown due to space limitations. For each set of sequences, the top strand is written 5' to 3' and the bottom strand is written 3' to 5'. ^b Single mismatch is identified by bold letters. Duplexes are written in alphabetical order by the loop nucleotide (A over G, not G over A). If the loop nucleotides are identical, duplexes are written in alphabetical order by the nearest neighbors (CUG over GUU, not GUU over CUG). ^c Frequency of occurrence in the database. ^d Percent out of 888 mismatches, the total number of mismatches adjacent to G-U pairs found in the database. ^e Reference 34. ^f Reference 36. ^g This work.

Table 2: Thermodynamic Parameters for Duplex Formation^a

frequency ^b	sequence ^c	analysis of melt curve fit/errors				analysis of T_m dependence/errors (ln plot)			
		ΔH° (kcal/mol)	ΔS° (cal/K·mol)	ΔG°_{37} (kcal/mol)	T_m^d (°C)	ΔH° (kcal/mol)	ΔS° (cal/K·mol)	ΔG°_{37} (kcal/mol)	T_m^d (°C)
104	CU CUG CUC ^e GA GUU GAG	-68.6 ± 5.4	-201.1 ± 17.5	-6.27 ± 0.1	35.8	-67.0 ± 1.5	-195.9 ± 4.8	-6.21 ± 0.02	35.5
	CAG CUG GUC ^{f,g} GUC GUU CAG	-94.9 ± 7.3	-263.7 ± 22.0	-13.12 ± 0.51	60.1	-94.5 ± 7.2	-262.5 ± 21.7	-13.04 ± 0.48	60.0
89	CAG AAU GUC ^{f,g} GUC UGG CAG	-81.3 ± 12.9	-232.3 ± 40.6	-9.26 ± 0.47	47.8	-81.3 ± 16.4	-232.1 ± 50.8	-9.27 ± 0.87	47.8
43	GA GUG GAG ^e CU CUU CUC	-67.9 ± 4.0	-198.6 ± 13.1	-6.29 ± 0.1	35.8	-70.9 ± 1.7	-208.4 ± 5.6	-6.21 ± 0.02	35.6
	GAC GUG CUG ^f CUG CUU GAC	-83.4 ± 5.3	-241.7 ± 16.7	-8.48 ± 0.12	44.4	-83.7 ± 2.9	-242.7 ± 9.2	-8.48 ± 0.07	44.4
40	CAG UAC GUC ^f GUC GGG CAG	-83.3 ± 6.7	-242.0 ± 21.2	-8.28 ± 0.19	43.7	-87.0 ± 4.7	-253.7 ± 15.0	-8.37 ± 0.10	43.7
38	GAG UAG AG ^e CUC GGC UC	-68.4 ± 7.2	-200.8 ± 23.1	-6.14 ± 0.1	35.3	-63.8 ± 1.8	-186.0 ± 5.8	-6.09 ± 0.02	34.9
	CAG UAG GUC ^f GUC GGC CAG	-56.5 ± 14.6	-154.3 ± 45.4	-8.59 ± 0.58	48.7	-62.6 ± 2.6	-174.2 ± 8.1	-8.57 ± 0.06	47.4
28	CAG UAG GUC GUC GGC CAG	-74.0 ± 7.1	-216.9 ± 22.8	-6.75 ± 0.08	37.9	-72.8 ± 3.6	-213.0 ± 11.5	-6.77 ± 0.04	38.0
27	GAC AUG CUG CUG UUU GAC	-92.4 ± 12.0	-272.8 ± 38.9	-7.79 ± 0.26	41.3	-95.1 ± 10.8	-281.6 ± 34.5	-7.78 ± 0.27	41.1
26	GAC CAU CUG ^h CUG GGG GAC	(-32.8)	(-79.4)	(-8.13)	(52.9)	(-32.2)	(-77.8)	(-8.11)	(53.0)
26	GAC UUG CUG CUG AUU GAC	-82.9 ± 5.8	-246.1 ± 18.6	-6.59 ± 0.20	37.2	-99.9 ± 7.9	-300.9 ± 25.4	-6.60 ± 0.09	37.2
20	GA GAG GAG ^e CU CGU CUC	-52.4 ± 1.1	-149.6 ± 3.2	-5.97 ± 0.1	33.7	-57.9 ± 1.3	-167.9 ± 4.3	-5.87 ± 0.03	33.5
18	GA GUU GAG ^e CU CUG CUC	-58.4 ± 4.4	-172.7 ± 14.7	-4.86 ± 0.2	28.4	-62.5 ± 1.1	-186.4 ± 3.6	-4.68 ± 0.04	28.1
15	GAC CCU CUG CUG GCG GAC	-89.4 ± 8.1	-260.1 ± 25.8	-8.77 ± 0.17	45.0	-91.8 ± 3.8	-267.5 ± 12.1	-8.81 ± 0.09	44.9
14	CU CAG CUC ^e GA GCU GAG	-60.7 ± 2.0	-176.5 ± 6.6	-5.97 ± 0.1	34.2	-60.8 ± 1.4	-176.7 ± 4.6	-5.97 ± 0.03	34.2
14	GAC GAU CUG ^{g,h} CUG CCG GAC	(-76.7)	(-207.1)	(-12.52)	(63.2)	(-78.7)	(-213.1)	(-12.59)	(62.9)
13	GAC GAA CUG CUG UGU GAC	-69.3 ± 3.9	-207.8 ± 12.9	-4.80 ± 0.11	29.4	-69.0 ± 1.9	-206.8 ± 6.3	-4.80 ± 0.06	29.4
12	GAC UAU CUG CUG AGG GAC	-65.6 ± 10.9	-190.1 ± 35.4	-6.65 ± 0.48	37.6	-65.8 ± 13.1	-190.6 ± 42.0	-6.68 ± 0.61	37.7
12	GAC UAG CUG CUG GGU GAC	-55.5 ± 10.5	-157.1 ± 33.9	-6.76 ± 0.15	38.3	-56.8 ± 6.7	-161.2 ± 21.3	-6.79 ± 0.24	38.4
11	CUC GAC UC ^{e,g} GAG UGG AG	-52.1 ± 2.8	-151.3 ± 9.0	-5.23 ± 0.1	29.5	-60.3 ± 1.2	-178.4 ± 4.1	-4.99 ± 0.04	29.3
11	GAG GCG AG ^e CUC UUC UC	-61.3 ± 1.9	-180.2 ± 6.3	-5.42 ± 0.1	31.5	-63.3 ± 1.7	-186.7 ± 5.6	-5.37 ± 0.05	31.4
11	GAC AUU CUG CUG UUG GAC	-90.9 ± 8.0	-274.2 ± 25.6	-5.88 ± 0.20	34.8	-82.3 ± 8.7	-246.3 ± 28.1	-5.93 ± 0.18	34.7
11	CU CUU CUC ^e GA GUG GAG	-58.1 ± 5.1	-169.7 ± 17.3	-5.41 ± 0.3	31.1	-63.7 ± 1.5	-188.6 ± 5.0	-5.19 ± 0.04	30.6
10	GAC GAG CUG CUG CCU GAC	-77.4 ± 3.9	-223.3 ± 12.3	-8.18 ± 0.08	43.8	-77.3 ± 4.2	-222.8 ± 13.2	-8.16 ± 0.07	43.7
10	GAC UAG CUG CUG ACU GAC	-89.3 ± 11.0	-264.2 ± 35.2	-7.42 ± 0.10	40.1	-85.3 ± 2.9	-251.0 ± 9.4	-7.41 ± 0.03	40.2
10	GAC GCU CUG CUG CUG GAC	-43.0 ± 8.2	-118.9 ± 27.4	-6.10 ± 0.44	34.0	-42.9 ± 6.7	-118.9 ± 21.9	-6.01 ± 0.36	33.3
10	GAC GCU CUG CUG UUA GAC	-70.2 ± 3.1	-208.7 ± 10.3	-5.49 ± 0.15	32.5	-70.7 ± 4.5	-210.3 ± 14.6	-5.48 ± 0.10	32.5

Table 2: Continued

frequency ^b	sequence ^c	analysis of melt curve fit/errors				analysis of T_m dependence/errors (ln plot)			
		ΔH° (kcal/mol)	ΔS° (cal/K·mol)	ΔG°_{37} (kcal/mol)	T_m^d (°C)	ΔH° (kcal/mol)	ΔS° (cal/K·mol)	ΔG°_{37} (kcal/mol)	T_m^d (°C)
9	GA GAG GAG ^e CU CAU CUC	-55.7 ± 8.1	-162.6 ± 25.9	-5.32 ± 0.3	30.4	-52.2 ± 6.0	-151.3 ± 20.0	-5.30 ± 0.27	29.9
9	CAG CCG CUG ^g GUC GUU GAC	-46.8 ± 11.1	-120.0 ± 33.7	-9.57 ± 0.72	58.6	-49.6 ± 5.0	-129.0 ± 15.4	-9.58 ± 0.35	57.3
8	CU CAG CUC ^e GA GAU GAG	-59.7 ± 8.9	-175.9 ± 30.1	-5.13 ± 0.4	29.9	-60.7 ± 2.5	-179.1 ± 8.3	-5.12 ± 0.08	30.0
8	GAC GAG CUG CUG UCU GAC	-66.4 ± 8.1	-194.9 ± 26.0	-5.94 ± 0.12	34.3	-73.1 ± 4.3	-216.5 ± 14.0	-5.93 ± 0.08	34.5
8	CU CAG CUC ^e GA GGU GAG	-55.8 ± 6.0	-159.9 ± 20.0	-6.23 ± 0.2	35.3	-59.9 ± 1.8	-173.5 ± 5.7	-6.07 ± 0.03	34.7

^a Measurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na₂EDTA, pH 7.0. ^b Frequency of occurrence in the database described in Materials and Methods. ^c Single mismatch is identified by bold letters. The nearest neighbors and the mismatch are set apart for easy identification. The top strand of each duplex is written 5' to 3', and each bottom strand is written 3' to 5'. ^d Calculated at 10⁻⁴ M oligomer concentration. ^e Reference 34. ^f Reference 36. ^g Duplexes that were not included in averages, trends, and the derivation of the predictive model because a bimolecular association of one of the strands with itself may be a competing structure. ^h Data derived from non-two-state melts.

sents ≤0.9% of the total number of mismatches that were found. When categorized in this manner, previous studies account for only 49% of the total number of single mismatches found, but after adding the data reported here, this percentage increases to 77%. Similarly, previous studies thermodynamically characterized only 14 types of mismatches in the top 30, but after adding the data reported here, all of the mismatches in the top 30 have been studied.

Dataset 2 (Table 1) lists frequency and percent occurrence when only the mismatch sequence is specified. When categorized in this fashion, seven types of mismatches were found in the database, representing all possible types of single mismatches. Also, when categorized in this manner, previous studies account for six types of mismatches; however, the current work has also characterized five of these seven types. It is important to note that a G•G single mismatch adjacent to at least one G-U pair has never been thermodynamically measured. Because this single mismatch was not found to be one of the 30 most common single mismatches adjacent to a G-U pair (any G•G mismatch with its nearest neighbors represents less than 1% of all of the single mismatches adjacent to at least one G-U pair that were found in the database), this single mismatch was also not studied here. The thermodynamics of a G•G single mismatch adjacent to at least one G-U pair should eventually be measured, and we will consider this in future studies.

Dataset 3 (Table 1) lists frequency and percent occurrence of 5' and 3' nearest neighbor combinations. When categorized in this fashion, 11 types of nearest neighbor combinations were found in the database, representing all possible types of nearest neighbor combinations containing at least one G-U base pair. Also, when categorized in this manner, previous studies account for 71% of all nearest neighbor combinations, but after adding the data reported here, this percentage increases to 99%.

Thermodynamic Parameters. Table 2 shows the thermodynamic parameters of duplex formation that were obtained from fitting each melting curve to the two-state model and from the van't Hoff plot of T_m^{-1} versus $\log(C_T/4)$. Data for the duplexes containing the 30 most frequently occurring single mismatches with at least one G-U nearest neighbor are shown in order of decreasing frequency. However, data for 33 duplexes are shown because three

mismatches were melted with different stem sequences. Duplexes that melted in a non-two-state manner and duplexes that may have been influenced by competing structures are denoted in Table 2.

Contribution of Single Mismatches to Duplex Thermodynamics. The contributions of the 30 most common single mismatches with at least one G-U pair to duplex stability are listed in Table 3. An additional six single mismatches were added to the 30 from Table 3, and the total list of 36 single mismatches can be found in Supporting Information (Table S1). These six additional single mismatches occur less frequently and were thermodynamically characterized in a previous study (34).

Updated Model for Predicting Thermodynamic Contributions of Single Mismatches. Recently, a model to predict the stability of an RNA duplex containing a single mismatch was proposed by Davis and Znosko (36). Using linear regression with the expanded dataset (data from the 64 mismatches that were used previously plus the data from the 14 mismatches reported here) and the same set of variables (treating G-U pairs adjacent to single mismatches the same as A-U pairs adjacent to single mismatches) used previously (36), nearest neighbor parameters for predicting the contribution of a single mismatch to duplex thermodynamics were derived.

Comparison of the previous result (36) to the results obtained here with the additional 14 mismatches reveals that eight of the nine parameters are within experimental error (data not shown). The one parameter that was not within experimental error, the mismatch-nearest neighbor interaction parameter for $\begin{bmatrix} 5'RY Y3' \\ 3'Y YR5' \end{bmatrix}$, was quite similar in the two models. This parameter was -0.5 ± 0.2 kcal/mol in the model proposed by Davis and Znosko (36) and 0.0 ± 0.2 kcal/mol in the model derived here.

In order to determine if a G-U pair adjacent to a single mismatch should be assigned a unique thermodynamic penalty (as opposed to the same penalty as an A-U pair adjacent to a single mismatch), the third parameter (parameter for an A-U/G-U closure) used for linear regression in the Davis and Znosko model (36) and used above with the expanded single mismatch dataset was separated into two variables, one for an A-U pair adjacent to a single mismatch

Table 3: Contributions of 30 Single Mismatches to Duplex Thermodynamics^a

frequency ^b	sequence ^c	$\Delta H^\circ_{\text{single mismatch}}$ (kcal/mol)		$\Delta S^\circ_{\text{single mismatch}}$ (cal/K·mol)		$\Delta G^\circ_{37, \text{single mismatch}}$ (kcal/mol)	
		measd ^d	predicted ^e	measd ^d	predicted ^e	measd ^d	predicted ^e
104	CUG ^f	−12.2	−16.3 (4.1)	−42.7	−58.5 (15.8)	1.07	1.4 (0.3)
	GUU						
	CUG ^{g,h}	−26.4	−16.3 (10.1)	−75.9	−58.5 (17.4)	−2.82	1.4 (4.2)
	GUU						
89	AAU ^{g,h}	−22.1	−8.8 (13.3)	−68.6	−44.7 (23.9)	−0.82	2.2 (3.0)
43	UGG						
	GUG ^f	−16.6	−18.9 (2.3)	−55.5	−65.0 (9.5)	0.67	0.8 (0.1)
	CUU						
	GUG ^g	−19.3	−18.9 (0.4)	−66.0	−65.0 (1.0)	1.08	0.8 (0.3)
40	CUU						
	UAC ^g	−22.6	−4.8 (17.8)	−77.0	−29.7 (47.3)	1.19	1.1 (0.1)
38	GGG						
	UAG ^f	−9.0	−4.0 (5.0)	−32.8	−23.5 (9.3)	1.19	1.7 (0.5)
	GGC						
	UAG ^g	4.5	−4.0 (8.5)	8.5	−23.5 (32.0)	1.89	1.7 (0.2)
28	GGC						
	UAA	−8.1	−8.0 (0.1)	−33.8	−34.3 (0.5)	2.27	2.5 (0.2)
27	GCU						
	AUG	−30.9	−22.9 (8.0)	104.7	−80.0 (24.7)	1.53	1.9 (0.4)
26	UUU						
	CAU ⁱ	30.7	−4.8 (35.5)	94.3	−29.7 (124.0)	1.37	1.1 (0.3)
26	GGG						
	UUA	−35.7	−20.3 (15.4)	−123.8	−73.5 (50.3)	2.68	2.5 (0.2)
20	GUU						
	GAG ^f	−3.6	−4.8 (1.2)	−15.0	−29.7 (14.7)	1.01	1.1 (0.1)
18	CGU						
	GUC ^f	−14.7	−19.2 (4.5)	−52.2	−66.2 (14.0)	1.50	0.8 (0.7)
15	UUG						
	CCU	−28.9	−4.0 (24.9)	−95.4	−19.3 (76.1)	0.67	1.4 (0.7)
14	GCG						
	CAG ^f	−6.0	−4.0 (2.0)	−23.5	−19.3 (4.2)	1.31	1.4 (0.1)
14	GCU						
	GAU ^{h,i}	−18.6	−21.3 (2.7)	−47.0	−73.8 (26.8)	−4.01	0.5 (4.5)
13	CCG						
	GAA	−10.8	−8.8 (2.0)	−46.3	−44.7 (1.6)	3.54	2.2 (1.3)
12	UGU						
	UAU	−5.8	−8.8 (3.0)	−24.1	−44.7 (20.6)	1.62	2.2 (0.6)
12	AGG						
	UAG	9.1	−8.0 (17.1)	21.0	−42.8 (63.8)	2.52	3.1 (0.6)
11	GGU						
	GAC ^{f,h}	−12.5	−4.8 (7.7)	−44.2	−29.7 (14.5)	1.19	1.1 (0.1)
11	UGG						
	GCG ^f	−12.7	−4.0 (8.7)	−44.1	−19.3 (24.8)	0.93	1.4 (0.5)
11	UUC						
	AUU	−22.4	−23.2 (0.8)	−80.0	−81.2 (1.2)	2.40	1.9 (0.5)
11	UUG						
	CUU ^f	−13.1	−18.9 (5.8)	−46.0	−65.0 (19.0)	1.11	0.8 (0.3)
10	GUG						
	GAG	−12.9	−4.0 (8.9)	−46.1	−19.3 (26.8)	1.40	1.4 (0.0)
10	CCU						
	UAG	−21.1	−8.0 (13.1)	−73.9	−34.3 (39.6)	1.87	2.5 (0.6)
10	ACU						
	GCU	17.2	−4.3 (21.5)	47.2	−20.5 (67.7)	2.57	1.4 (1.2)
10	CUG						
	GCU	−11.5	−8.3 (3.2)	−46.8	−35.5 (11.3)	2.97	2.5 (0.5)
10	UUA						
	UUA						

Table 3: Continued

frequency ^b	sequence ^c	$\Delta H^\circ_{\text{single mismatch}}$ (kcal/mol)		$\Delta S^\circ_{\text{single mismatch}}$ (cal/K·mol)		$\Delta G^\circ_{37,\text{single mismatch}}$ (kcal/mol)	
		measd ^d	predicted ^e	measd ^d	predicted ^e	measd ^d	predicted ^e
9	GAG ^f CAU	2.1	−4.0 (6.1)	1.6	−19.3 (20.9)	1.58	1.4 (0.2)
9	CCG ^h GUU	18.5	−1.4 (19.9)	57.6	−12.8 (70.4)	0.64	2.0 (1.4)
8	CAG ^f GAU	−5.9	−3.2 (2.7)	−25.9	−13.1 (12.8)	2.16	2.0 (0.2)
8	GAG UCU	−13.8	−8.0 (5.8)	−53.0	−38.6 (14.4)	2.68	2.8 (0.1)
8	CAG ^f GGU	−5.1	−4.0 (1.1)	−20.3	−23.5 (3.2)	1.21	1.7 (0.5)

^a Calculations were based on the data obtained from T_m^{-1} vs $\ln(C_T/4)$ plots. ^b Frequency of occurrence in the database described in Materials and Methods. ^c Single mismatch is identified by bold letters. The top strand of each duplex is written 5' to 3', and each bottom strand is written 3' to 5'. ^d Measured values were calculated by subtracting the nearest neighbor contributions for the canonical pairs (21, 55) from the raw optical melting data for the duplex. Watson–Crick nearest neighbor parameters have average uncertainties of 0.07 kcal/mol, 1.4 kcal/mol, and 4.3 eu for free energy, enthalpy, and entropy, respectively (55). G–U nearest neighbor parameters have average uncertainties of 0.27 kcal/mol, 2.4 kcal/mol, and 7.2 eu for free energy, enthalpy, and entropy, respectively (21). ^e Values derived from the proposed model found in Table 4. Differences between the measured values and the predicted values are shown in parentheses. It is important to note that predictions may not be accurate for duplexes in which a bimolecular association of one of the strands with itself may be a competing structure or for data derived from non-two-state melts. ^f Reference 34. ^g Reference 36. ^h Duplexes that were not included in averages, trends, and the derivation of the predictive model because a bimolecular association of one of the strands with itself may be a competing structure. ⁱ Data derived from non-two-state melts.

and one for a G–U pair adjacent to a single mismatch. Linear regression results in a 1.1 ± 0.1 kcal/mol penalty for A–U pairs adjacent to a single mismatch and a 1.4 ± 0.1 kcal/mol penalty for a G–U pair adjacent to a single mismatch (Table 4). All other $\Delta G^\circ_{37,\text{single mismatch}}$ parameters are within experimental error of those derived with the combined A–U/G–U penalty and the expanded single mismatch dataset. The parameters in Table 4 were used to predict the contributions of the 30 most common single mismatches adjacent to at least one G–U pair to duplex thermodynamics. Contributions are calculated using the following equation:

$$\Delta G^\circ_{37,\text{single mismatch}} = \Delta G^\circ_{37,\text{mismatch nt}} + \Delta G^\circ_{37,\text{mismatch-NN interaction}} + \Delta G^\circ_{37,\text{AU}} + \Delta G^\circ_{37,\text{GU}} \quad (1)$$

An example of this calculation is shown below:

$$\Delta G^\circ_{37,\left[\begin{smallmatrix} \text{CUG} \\ \text{GUU} \end{smallmatrix}\right]} = \Delta G^\circ_{37,\left[\begin{smallmatrix} \text{U} \\ \text{U} \end{smallmatrix}\right]} + \Delta G^\circ_{37,\left[\begin{smallmatrix} \text{YYR} \\ \text{RYY} \end{smallmatrix}\right]} + \Delta G^\circ_{37,\text{AU}} + \Delta G^\circ_{37,\text{GU}} \quad (2)$$

$$\Delta G^\circ_{37,\left[\begin{smallmatrix} \text{CUG} \\ \text{GUU} \end{smallmatrix}\right]} = -0.6 + 0.6 + 0.0 + 1.4 = 1.4 \text{ kcal/mol} \quad (3)$$

A second example is as follows:

$$\Delta G^\circ_{37,\left[\begin{smallmatrix} \text{AAU} \\ \text{UGG} \end{smallmatrix}\right]} = \Delta G^\circ_{37,\left[\begin{smallmatrix} \text{A} \\ \text{G} \end{smallmatrix}\right]} + \Delta G^\circ_{37,\left[\begin{smallmatrix} \text{RRY} \\ \text{YRR} \end{smallmatrix}\right]} + \Delta G^\circ_{37,\text{AU}} + \Delta G^\circ_{37,\text{GU}} \quad (4)$$

$$\Delta G^\circ_{37,\left[\begin{smallmatrix} \text{AAU} \\ \text{UGG} \end{smallmatrix}\right]} = -0.3 + 0.0 + 1.1 + 1.4 = 2.2 \text{ kcal/mol} \quad (5)$$

These predictions and the differences between the predictions and the measured contributions are shown in Tables 3 and S1. Parameters for $\Delta H^\circ_{\text{single mismatch}}$ and $\Delta S^\circ_{\text{single mismatch}}$ are also included in Tables 3 and S1.

DISCUSSION

Database Searching. The database used for this study contains 955 secondary structures from eight different

kinds of RNAs. Although this database does not contain all RNA secondary structures available, we have assumed that the number and variety of structures in this database approximates the number and types of single mismatches with at least one G–U nearest neighbor that are found in nature.

It is clear from the first set of data in Table 1 that previous experiments with single mismatches (34, 36) have provided results for only 14 of the top 30 mismatch–nearest neighbor nucleotide combinations with at least one G–U pair found most commonly in the database. The results reported here expand the available measured dataset to include all combinations in the top 30. Dataset 1 in Table 1 provides some interesting results. For example, it is interesting to note that nine of the top ten single mismatches with G–U nearest neighbors contain mismatches that have been previously considered to be stabilizing (A•G and U•U) in single mismatches (21, 43, 44, 46); however, none of the top 30 single mismatches with adjacent G–U base pairs include G•G mismatches, which have been shown to be the most stable mismatch (36). These results indicate that there is no correlation between the stability of a single mismatch closed by a G–U pair and its frequency of occurrence in nature.

Dataset 2 in Table 1 indicates that all possible single mismatches were found in the database. A•G and U•U mismatches are the most prevalent single mismatches adjacent to at least one G–U pair, which is in agreement with what was found previously for single mismatches with any type of nearest neighbor (36). The percent occurrence of each mismatch when adjacent to at least one G–U pair was compared to the percent occurrence of each mismatch with any type of nearest neighbors. For example, it was shown previously that A•G mismatches account for 28% of all single mismatches (36). When examining single mismatches adjacent to at least one G–U pair, A•G mismatches account for 34% of the single mismatches (Table 1, dataset 2). When

Table 4: Nearest Neighbor Parameters for Single Mismatches at 37 °C

	ΔH° (kcal/mol)	ΔS° (cal/K·mol)	ΔG°_{37} (kcal/mol)
mismatch parameters ^a			
A•G	-0.8 ± 3.0	-10.4 ± 8.5	-0.3 ± 0.2
G•G	-17.9 ± 4.1	-52.2 ± 11.6	-2.1 ± 0.2
U•U	-14.9 ± 3.1	-45.7 ± 8.9	-0.6 ± 0.2
mismatch–NN interaction parameters ^b			
YRR RRY	0.8 ± 3.4	6.2 ± 9.7	0.6 ± 0.2
RYY YYR	-0.3 ± 3.6	-1.2 ± 10.3	0.0 ± 0.2
YYR RYY	2.6 ± 4.5	6.5 ± 13.8	0.6 ± 0.3
YRY RYR	-7.0 ± 5.7	-20.3 ± 16.3	-0.5 ± 0.3
RRY YYR	-17.3 ± 7.2	-54.5 ± 20.4	-0.9 ± 0.4
A-U or G-U closure parameters ^c			
A-U	-4.0 ± 1.7	-15.0 ± 4.8	1.1 ± 0.1
G-U	-4.0 ± 1.9	-19.3 ± 5.4	1.4 ± 0.1

^a Mismatches not included in the table do not contribute to duplex thermodynamics. ^b The pairs on the left and right are the adjacent nearest neighbors and the pair in the center is the single mismatch. Mismatch–NN interactions not included in the table do not contribute to duplex thermodynamics. The top strand is written 5' to 3', and the bottom strand is written 3' to 5'. It is important to note that these orientations of the mismatch and nearest neighbors are important for assigning the value of the mismatch–NN parameter. For example, 5'YRR3'/3'RRY5' is assigned a value of 0.6 kcal/mol while 5'RRY3'/3'YRR5' is not listed in the table, so this orientation of the mismatch and nearest neighbors does not contribute to duplex thermodynamics. Adenine and guanine are classified as purines (R), and cytosine and uracil are classified as pyrimidines (Y). ^c These parameters are applied per A-U or G-U closure.

making a similar comparison between all of the other mismatches, there were no differences greater than 5%.

Dataset 3 in Table 1 shows the nearest neighbor combinations of single mismatches and their relative frequency. The most frequent nearest neighbor combination is $\begin{bmatrix} \text{CXG} \\ \text{GXU} \end{bmatrix}$, representing 23% of the total number of loops. Furthermore, it is interesting to note that nearest neighbor combinations containing two G-U base pairs only represent 7% of the single mismatches with at least one G-U nearest neighbor.

Thermodynamic Contributions of Single Mismatches to Duplex Thermodynamics. The examination of the data listed in Tables 3 and S1 indicates a large variance in the obtained thermodynamic parameters. Single mismatch contributions to enthalpy, entropy, and free energy changes range from -35.7 to 17.2 kcal/mol, -123.8 to 47.2 cal/(K·mol), and -0.20 to 3.54 kcal/mol, respectively. There does not appear to be a correlation between the thermodynamic contribution of a loop and its frequency of occurrence in the database. For example, the most stabilizing loop that was measured, $\begin{bmatrix} 5'\text{GAC}3' \\ 3'\text{UGC}5' \end{bmatrix}$ (34), does not occur in the top 30.

There does seem to be a correlation between the identity of the base pairs directly adjacent to the mismatch and the

Table 5: Order of Stability of Single Mismatches Adjacent to G-U Base Pairs^a

	range $\Delta G^\circ_{37, \text{single mismatch}}$ (kcal/mol)	average $\Delta G^\circ_{37, \text{single mismatch}}$ (kcal/mol)
single mismatch ^b		
C•C	0.7–1.7	1.2 ± 0.7
U•U	0.7–2.7	1.5 ± 0.7
A•G	-0.2 –3.5	1.5 ± 1.0
A•A	0.6–2.2	1.5 ± 0.7
A•C	0.5–2.7	1.7 ± 0.8
C•U	0.9–3.0	2.2 ± 1.1

^a Values are based on the data obtained from T_m^{-1} vs $\ln(C_7/4)$ plots which are shown in Tables 3 and S1. G•G mismatches directly adjacent to G-U base pairs have not been measured. ^b Single mismatches are listed from least destabilizing (top) to most destabilizing (bottom) based on the average contribution.

free energy contribution to duplex stability. For example, the 22 single mismatches with one adjacent G-U pair and one adjacent G-C pair contribute an average of 1.3 kcal/mol to duplex stability. The eight single mismatches with one adjacent G-U pair and one adjacent A-U pair contribute an average of 2.4 kcal/mol to duplex stability. The two single mismatches with two adjacent G-U pairs contribute an average of 2.6 kcal/mol to duplex stability. This trend appears consistent with the ability of G-C nearest neighbors to form three hydrogen bonds while A-U and G-U nearest neighbors can only form two hydrogen bonds.

Examination of the single mismatch free energy values (Tables 3 and S1) gives the following order of single mismatch stability (from least destabilizing to most destabilizing) (Table 5): C•C (range of 0.7 to 1.7 kcal/mol, average of 1.2 ± 0.7 kcal/mol) > U•U (range of 0.7 to 2.7 kcal/mol, average of 1.5 ± 0.7 kcal/mol) \approx A•G (range of -0.2 to 3.5 kcal/mol, average of 1.5 ± 1.0 kcal/mol) \approx A•A (range of 0.6 to 2.2 kcal/mol, average of 1.5 ± 0.7 kcal/mol) > A•C (range of 0.5 to 2.7 kcal/mol, average of 1.7 ± 0.8 kcal/mol) > C•U (range of 0.9 to 3.0 kcal/mol, average of 2.2 ± 1.1 kcal/mol). It is important to note that G•G mismatches directly adjacent to G-U base pairs have not been measured. From the ranges observed for every mismatch, it is evident that the identity of the single mismatch nucleotides alone does not determine the stability of single mismatches, which is similar to the results observed previously (36).

Updated Model for Predicting Thermodynamic Contributions of Single Mismatches. Due to their prevalence and roles in several biological systems, single mismatches adjacent to G-U pairs are important motifs to characterize. Prior to this study, only 14 of the 30 most frequently occurring single mismatches adjacent to G-U pairs had been thermodynamically characterized (Table 1). With the data reported here, experimental values for the thermodynamic contribution of single mismatches with G-U nearest neighbors can be used when predicting the stability of RNAs which contain these motifs; a predictive model is no longer needed for these mismatches.

With these new data, the predictive model for previously unmeasured single mismatches is updated. As described above, adding this new dataset to the dataset of previously measured single mismatches and recalculating the nearest neighbor parameters using the same parameters does not change the nearest neighbor parameters as proposed previously (36). However, when separating the A-U/G-U closure

parameter into two separate parameters, one for A-U closures and one for G-U closures, a slight difference (Table 4) was obtained for each type of closure (1.1 ± 0.1 kcal/mol for A-U closures and 1.4 ± 0.1 kcal/mol for G-U closures). Even though the nearest neighbor parameters have been updated, nearest neighbor parameters alone are not sufficient to include all factors known to affect single mismatch stability. For example, non-nearest neighbors and distance of the single mismatch from the helix end also seem to play a role in single mismatch stability (50).

SUPPORTING INFORMATION AVAILABLE

A table listing the contributions of 36 single mismatches to duplex thermodynamics and a table providing a model grid for calculating single mismatch contributions to duplex thermodynamics (Table 4 in a different format). This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES

- Crick, F. H. C. (1966) Condon-anticodon pairing—wobble hypothesis. *J. Mol. Biol.* 19, 548–555.
- Erdmann, V. A., and Wolters, J. (1986) Collection of published 5S, 5.8S and 4.5S ribosomal-RNA sequences. *Nucleic Acids Res.* 14, R1–R59.
- Noller, H. F. (1984) Structure of ribosomal RNA. *Annu. Rev. Biochem.* 53, 119–162.
- Sprinzl, M., Moll, J., Meissner, F., and Hartmann, T. (1985) Compilation of tRNA sequences. *Nucleic Acids Res.* 13, r1–r49.
- Stern, S., Weiser, B., and Noller, H. F. (1988) Model for the three-dimensional folding of 16 S ribosomal RNA. *J. Mol. Biol.* 204, 447–481.
- Woese, C. R., Gutell, R. R., Gupta, R., and Noller, H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.* 47, 621–629.
- Varani, G., and McClain, W. H. (2000) The G·U wobble base pair - A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1, 18–23.
- Szymanski, M., Barciszewska, M. Z., Erdmann, V. A., and Barciszewski, J. (2000) An analysis of G-U base pair occurrence in eukaryotic 5S rRNAs. *Mol. Biol. Evol.* 17, 1194–1198.
- Gautheret, D., Konings, D., and Gutell, R. R. (1995) GU base-pairing motifs in ribosomal-RNA. *RNA* 1, 807–814.
- Mokdad, A., Krasovska, M., Sponer, J., and Leontis, N. B. (2006) Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* 34, 1326–1341.
- Damberger, S. H., and Gutell, R. R. (1994) A comparative database of group I intron structures. *Nucleic Acids Res.* 22, 3508–3510.
- Masquida, B., and Westhof, E. (2000) On the wobble G-U and related pairs. *RNA* 6, 9–15.
- Uhlenbeck, O. C., Martin, F. H., and Doty, P. (1971) Self-complementary oligoribonucleotides: Effects of helix defects and guanylic acid-cytidylic acid base pairs. *J. Mol. Biol.* 57, 217–229.
- Romaniuk, P. J., Hughes, D. W., Gregoire, R. J., Bell, R. A., and Neilson, T. (1979) Effects of internal nonbonded bases and a G·U base pair on the stability of a short ribonucleic acid helix. *Biochemistry* 18, 5109–5116.
- Alkema, D., Hader, P. A., Bell, R. A., and Neilson, T. (1982) Effects of flanking G·C base pairs on internal Watson-Crick, G·U, and nonbonded base pairs within a short ribonucleic acid duplex. *Biochemistry* 21, 2109–2117.
- Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helices. *Biochemistry* 25, 5755–5759.
- McDowell, J. A., and Turner, D. H. (1996) Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry* 35, 14077–14089.
- Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Free energy contributions of G·U and other terminal mismatches to helix stability. *Biochemistry* 25, 3209–3213.
- Sugimoto, N., Kierzek, R., and Turner, D. H. (1987) Sequence dependence for the energetics of dangling ends and terminal base pairs in ribonucleic acid. *Biochemistry* 26, 4554–4558.
- He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A. E., and Turner, D. H. (1991) Nearest neighbor parameters for G-U mismatches: 5'GU3'/3'UG5' is destabilizing in the contexts CGUG/GUGC, UGUA/AUGU, and AGUU/UUGA but stabilizing in GGUC/CUGG. *Biochemistry* 30, 11124–11132.
- Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288, 911–940.
- Strobel, S. A., and Cech, T. R. (1995) Minor-groove recognition of the conserved G-U pair at the *Tetrahymena* ribozyme reaction site. *Science* 267, 675–679.
- Strobel, S. A., and Ortoleva-Donnelly, L. (1999) A hydrogen-bonding triad stabilizes the chemical transition state of a group I ribozyme. *Chem. Biol.* 6, 153–165.
- Bevilacqua, P. C., Kierzek, R., Johnson, K. A., and Turner, D. H. (1992) Dynamics of ribozyme binding of substrate revealed by fluorescence-detected stopped-flow methods. *Science* 258, 1355–1358.
- Michel, F., and Westhof, E. (1994) Slippery substrates. *Nat. Struct. Biol.* 1, 5–7.
- Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferre-D'Amare, A. R., Strobel, S. A., and Doudna, J. A. (1998) A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat. Struct. Biol.* 5, 986–992.
- Walter, A. E., Wu, M., and Turner, D. H. (1994) The stability and structure of tandem GA mismatches in RNA depend on closing base pairs. *Biochemistry* 33, 11349–11354.
- Wu, M., McDowell, J. A., and Turner, D. H. (1995) A periodic table of symmetric tandem mismatches in RNA. *Biochemistry* 34, 3204–3211.
- Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry* 39, 9257–9274.
- Schroeder, S. J., and Turner, D. H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA. *Biochemistry* 40, 11509–11517.
- Znosko, B. M., Kennedy, S. D., Wille, P. C., Krugh, T. R., and Turner, D. H. (2004) Structural features and thermodynamics of the J4/5 loop from the *Candida albicans* and *Candida dubliniensis* group I introns. *Biochemistry* 43, 15822–15837.
- Schroeder, S. J., Fountain, M. A., Kennedy, S. D., Lukavsky, P. J., Puglisi, J. D., Krugh, T. R., and Turner, D. H. (2003) Thermodynamic stability and structural features of the J4/5 loop in a *Pneumocystis carinii* group I intron. *Biochemistry* 42, 14184–14196.
- Chen, G., Znosko, B. M., Jiao, X., and Turner, D. H. (2004) Factors affecting thermodynamic stabilities of RNA 3 × 3 internal loops. *Biochemistry* 43, 12865–12876.
- Xia, T. B., McDowell, J. A., and Turner, D. H. (1997) Thermodynamics of nonsymmetric tandem mismatches adjacent to G-C base pairs in RNA. *Biochemistry* 36, 12486–12497.
- Christiansen, M. E., and Znosko, B. M. (2008) Thermodynamic characterization of the complete set of sequence symmetric tandem mismatches in RNA and an improved model to predict the free energy contribution of sequence asymmetric tandem mismatches. *Biochemistry* 47, 4329–4336.
- Davis, A. R., and Znosko, B. M. (2007) Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry* 46, 13425–13436.
- Hirao, I., Harada, Y., Nojima, T., Osawa, Y., Masaki, H., and Yokoyama, S. (2004) *In vitro* selection of RNA aptamers that bind to colicin E3 and structurally resemble the decoding site of 16S ribosomal RNA. *Biochemistry* 43, 3214–3221.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* 289, 905–920.
- Calin-Jageman, I., and Nicholson, A. W. (2003) Mutational analysis of an RNA internal loop as a reactivity epitope for *Escherichia coli* ribonuclease III substrates. *Biochemistry* 42, 5025–5034.
- Kikuchi, K., Umehara, T., Fukuda, K., Kuno, A., Hasegawa, T., and Nishikawa, S. (2005) A hepatitis C virus (HCV) internal ribosome entry site (IRES) domain III-IV-targeted aptamer inhibits

- translation by binding to an apical loop of domain III_d. *Nucleic Acids Res.* 33, 683–692.
41. Shi, P. Y., Brinton, M. A., Veal, J. M., Zhong, Y. Y., and Wilson, W. D. (1996) Evidence for the existence of a pseudoknot structure at the 3' terminus of the flavivirus genomic RNA. *Biochemistry* 35, 4222–4230.
 42. Dunn, J. J., Studier, F. W., and Gottesman, M. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* 166, 477–535.
 43. Mathews, D. H., Disney, M. D., Childs, J. C., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A.* 101, 7287–7292.
 44. Lu, Z. J., Turner, D. H., and Mathews, D. H. (2006) A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res.* 34, 4912–4924.
 45. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science* 244, 48–52.
 46. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.
 47. Hofacker, I. L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.* 31, 3429–3431.
 48. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs. *Biochemistry* 30, 8242–8251.
 49. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) Thermodynamic study of internal loops in oligoribonucleotides: Symmetric loops are more stable than asymmetric loops. *Biochemistry* 30, 6428–6436.
 50. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry* 38, 14214–14223.
 51. Bevilacqua, J. M., and Bevilacqua, P. C. (1998) Thermodynamic analysis of RNA combinatorial library contained in a short hairpin. *Biochemistry* 37, 15877–15884.
 52. Andronescu, M., Condon, A., Hoos, H. H., Mathews, D. H., and Murphy, K. P. (2007) Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics* 23, i19–i28.
 53. Badhwar, J., Karri, S., Cass, C. K., Wunderlich, E. L., and Znosko, B. M. (2007) Thermodynamic characterization of RNA duplexes containing naturally occurring 1 × 2 nucleotide internal loops. *Biochemistry* 46, 14715–14724.
 54. Wright, D. J., Rice, J. L., Yanker, D. M., and Znosko, B. M. (2007) Nearest neighbor parameters for inosine-uridine pairs in RNA duplexes. *Biochemistry* 46, 4625–4634.
 55. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37, 14719–14735.

BI800471Z