9058 *Biochemistry* **2010,** *49,* 9058–9062
DOI: 10.1021/bi101164s

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Thermodynamic Characterization of RNA Triloops†

Praneetha Thulasi, Lopa K. Pandya, and Brent M. Znosko*

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

Received July 21, 2010; Revised Manuscript Received September 14, 2010

ABSTRACT: Relatively few thermodynamic parameters are available for RNA triloops. Therefore, 24 stem−loop sequences containing naturally occurring triloops were optically melted, and the thermodynamic parameters $\Delta H°$, $\Delta S°$, $\Delta G°_{37}$, and T_M for each stem−loop were determined. These new experimental values, on average, are 0.5 kcal/mol different from the values predicted for these triloops using the model proposed by Mathews et al. [Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) *Proc. Natl. Acad. Sci. U.S.A. 101*, 7287−7292]. The data for the 24 triloops reported here were then combined with the data for five triloops that were published previously. A new model was derived to predict the free energy contribution of previously unmeasured triloops. The average absolute difference between the measured values and the values predicted using this proposed model is 0.3 kcal/mol. These new experimental data and updated predictive model allow for more accurate calculations of the free energy of RNA stem−loops containing triloops and, furthermore, should allow for improved prediction of secondary structure from sequence.

RNA stem−loops containing three nucleotides in the loop, triloops, are common secondary structure motifs found in naturally occurring RNA. For example, bacterial 16S rRNAs strongly favor tetraloops; however, the UUU triloop is the most common replacement (*1*). In the 16S-like rRNA variable regions, triloops account for 7% of the loops in bacteria and 16% of the loops in eukaryotes (*2*). Triloops are also found in large subunit rRNAs (*3*, *4*), 5S rRNAs (*5*), signal recognition particles (*6*), RNase P RNAs (*7*), and group I introns (*8*, *9*). More specifically, triloops are found in Brome mosaic virus (+) strand RNA (*10*), human rhinovirus isotype 14 (*11*), iron responsive element RNA (*12*), and an RNA aptamer for bacteriophage MS2 coat protein (*13*), to name a few. Although relatively unstable due to the strain in the loop, triloops may be an important structural feature due to the accessibility of the loop nucleotides for recognition by proteins, other nucleic acids, or small molecules. It has been shown that triloops play a role in various biological processes, including virus replication (*11*, *14*), viral synthesis (*15*), and iron response (*12*), to name a few. Nevertheless, only a few studies have thermodynamically characterized RNA triloops (*16*−*20*), and only three of these (*16*−*18*) were done using 1 M NaCl (the salt concentration in which most nearest neighbor parameters are derived).

The current model used by secondary structure prediction algorithms to predict the thermodynamic contribution of RNA triloops to stem−loop stability is sequence independent; all triloops contribute 5.4 kcal/mol to stem−loop stability, with the exception of 5′CCC3′ which contributes 6.9 kcal/mol (*21*). In addition, there are two unstable triloop sequences (5′CAACG3′ and 5′GUUAC3′) for which this predictive model is not used; instead, the $\Delta G°_{37,loop}$ values (6.8 and 6.9 kcal/mol, respectively) for these two triloops are provided in a lookup table (*21*). An interesting study by the Bevilacqua laboratory (*19*) used a combinatorial approach and temperature gradient gel electrophoresis to identify stable and unstable RNA triloops. It was discovered that sequence preferences for exceptionally stable triloops included a U-rich loop and C-G as the closing base pair. Although they used 10 mM NaCl during their melting experiments, they suggested that the rules for predicting triloop stability at 1 M NaCl should be modified; however, this has yet to be done. Here, we report the thermodynamic parameters for 24 previously unmeasured RNA triloops in 1 M NaCl and propose a new algorithm for predicting the contribution of triloops to stem−loop stability, which includes two bonuses for stabilizing sequence features.

MATERIALS AND METHODS

Compiling and Searching a Database for RNA Triloops. The initial aim of this project was to identify the most frequently occurring RNA triloops in nature and to thermodynamically characterize these hairpin triloop sequences. Therefore, a database of 1349 RNA secondary structures containing 123 small subunit rRNAs (*22*), 223 large subunit rRNAs (*3*, *4*), 309 5S rRNAs (*5*), 484 tRNAs (*23*), 91 signal recognition particles (*6*), 16 RNase P RNAs (*7*), 100 group I introns (*8*, *9*), and 3 group II introns (*24*) was compiled. This database was searched for triloops, and the number of occurrences for each type of triloop was tabulated. In this work, G-U pairs are considered to be canonical base pairs.

Design of Sequences for Optical Melting Studies. Since most thermodynamic parameters for RNA secondary structure motifs are reported for RNA solutions containing 1 M NaCl, the melting buffer used in this work also contained 1 M NaCl. A major limitation of a thermodynamic analysis of RNA hairpins using this high salt concentration is the possible bimolecular

†This work was supported by Award Number R15GM085699 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute of General Medical Sciences of the National Institutes of Health.

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

 Published on Web 09/15/2010

Table 1: Summary of Database Search Results for RNA Triloops[a]

data set 1: triloop with adjacent base pair			data set 2: triloop			data set 3: adjacent base pairs			data set 4: triloop nucleotides classified as purine (R) or pyrimidine (Y)		
triloop[a]	freq[b]	%[c]	triloop[a]	freq[b]	%[c]	closing bp	freq[b]	%[c]	triloop[a]	freq[b]	%[c]
GGGGC	106	12.9	GGG	108	13.1	C-G	345	42.0	RRR	207	25.2
CGCAG	53	6.4	UAA	96	11.7	G-C	219	26.6	RYR	181	22.0
CUAAG	45	5.5	GCA	66	8.0	U-A	163	19.8	YRR	123	15.0
UUAAA	41	5.0	UUU	53	6.4	A-U	56	6.8	YYY	120	14.6
CAUAG	34	4.1	AUA	48	5.8	U-G	26	3.2	YYR	53	6.4
UGAAA	24	2.9	AAA	32	3.9	G-U	13	1.6	YRY	47	5.7
CUUUG	24	2.9	GAA	32	3.9				RYY	45	5.5
CUCUG	17	2.1	GUA	32	3.9				RRY	41	5.0
CAGAG	15	1.8	UCU	24	2.9						
GGUAC	15	1.8	AGA	22	2.7						

[a]Not all combinations in data sets 1 and 2 are shown due to space limitations. For each set of sequences, the strand is writen 5′ to 3′. [b]Frequency of occurrence in the secondary structure database described in Materials and Methods. [c]Percent out of 822 triloops, the total number of triloops found in the secondary structure database.

association of RNA strands (*19*, *25*). To ensure that unimolecular triloop formation out-competed bimolecular association in a 1 M NaCl solution, the following equations, derived from the equilibrium equations and $\Delta G^\circ = -RT \ln K$, were utilized:

$$[H] = \frac{-1 + \sqrt{1 + ((8K_D[A]_T)/(K_H K_H))}}{(4K_D)/(K_H K_H)} \quad (1)$$

$$[D] = ([A]_T - [H])/2 \quad (2)$$

$$\% H = [H]/([H] + [D]) \times 100 \quad (3)$$

Here, [H] is the concentration of hairpin, K_D is the equilibrium constant for duplex formation, $[A]_T$ is the total strand concentration, K_H is the equilibrium constant for hairpin formation, [D] is the concentration of duplex, and % *H* is the percent hairpin in solution.[1] K_H and K_D values were calculated at 37 °C using ΔG°_{37} values from RNAstructure (*21*, *26*, *27*) for hairpin and duplex formation, respectively. Calculations were done for $[A]_T = 1\ \mu M$ and 0.1 mM, which are the typical concentration range for the melting experiments. Due to potential competition from duplex formation, many of the most frequently occurring triloops were not studied here; only those that were likely to form triloops were used. All of the sequences studied here had % [H] > 84% at $[A]_T = 0.1$ mM and % [H] > 99% at $[A]_T = 1\ \mu M$.

Sequences of triloops and closing base pairs were designed to represent those found in the database described above. Each stem contained three Watson−Crick pairs in addition to the closing base pair. The terminal base pair was always a G-C pair in order to prevent end fraying of the sequence during melting. Care was taken to design the stem−loop sequences so that the triloop of interest would form, with little competition from other secondary structure motifs.

RNA Synthesis and Purification. Oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA). The purification of the oligonucleotides followed standard procedures that were described previously (*28*).

Optical Melting Experiments and Thermodynamics. Optical melting experiments were performed in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA (pH 7.0). All stem−loops were melted about nine times with approximately a 50-fold concentration range. Each stem−loop melting curve resulted in a single transition, and all melts of a given stem−loop were concentration independent, suggesting stem−loop formation. Stem−loop thermodynamics were determined by averaging the thermodynamics derived from each individual curve fit. The thermodynamic contributions of triloops to stem−loop thermodynamics ($\Delta G^\circ_{37,triloop}$, $\Delta H^\circ_{triloop}$, and $\Delta S^\circ_{triloop}$) were determined by subtracting the Watson−Crick contribution (*29*) from the measured thermodynamics. Stem sequences in which the terminal pair of the stem (or the triloop closing base pair) is A-U, U-A, G-U, or U-G utilize the 0.45 kcal/mol terminal A-U penalty (*29*) when calculating the contribution of the Watson−Crick stem; thus, this 0.45 kcal/mol contribution is not included in the parameters for predicting the loop contribution to stem−loop stability.

Linear Regression and Triloop Thermodynamic Parameters. Data collected for the 24 triloops in this study were combined with previously published data for five triloops (*16*−*18*) that were also melted in 1 M NaCl. Data for seven triloops measured in these previous studies (*16*−*18*) were omitted from the analysis and discussion here because, using eqs 1−3, they resulted in % $H < 75\%$. Additional thermodynamic data are available for triloops in 10 mM sodium phosphate (*17*, *19*, *20*) and 0.1 M NaCl (*16*) but, due to different salt concentrations, were not included in the analysis described here.

A predictive model was derived by linear regression as described previously (*28*). The calculated experimental contribution of the triloop to stem−loop stability was used as a constant when doing linear regression. Many combinations of variables were tested. To simultaneously solve for each variable, the LINEST function of Microsoft Excel was used for linear regression.

RESULTS

Database Searching. A database of RNA secondary structures was searched for triloops. In this database, 822 triloops were found, averaging about one triloop for every two secondary structures. Table 1 shows a summary of the database results obtained. Supporting Information Table S1 shows the complete results. The first set of data in Table 1 lists frequency and percent occurrence when the triloop nucleotides and the closing base pair are specified. Because the stability of hairpin loops depends on both the identity of the nucleotides in the loop and the closing

[1]Abbreviations: $[A]_T$, the total strand concentration; [D], concentration of duplex; [H], concentration of hairpin; K_D, equilibrium constant for duplex formation; K_H, equilibrium constant for hairpin formation; R, purines; Y, pyrimidines; % *H*, percent hairpin in solution.

Table 2: Thermodynamic Parameters for Stem−Loop Formation and Contributions of Triloops to Stem−Loop Stability[a]

frequency[b]	sequence[c]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°_{37}$ (kcal/mol)	T_M[d] (°C)	$\Delta H°_{triloop}$[e] (kcal/mol)	$\Delta G°_{37,triloop}$[e] (kcal/mol)	predicted $\Delta G°_{37,triloop}$[f] (kcal/mol)
34	GGCAUAGCC[g]	−21.9 ± 1.9	−67.5 ± 6.0	−0.94 ± 0.15	51.0	6.4	5.74	5.2
15	GGCCAGAGGCC	−39.8 ± 2.1	−113.9 ± 6.2	−4.46 ± 0.22	76.2	1.9	5.48	5.2
14	GCCUUUUAGGC	−36.2 ± 1.8	−105.6 ± 5.4	−3.45 ± 0.15	69.7	2.6	4.86	4.7
12	GGCUAAAAGCC	−29.9 ± 2.5	−88.5 ± 7.8	−2.47 ± 0.18	64.9	8.9	5.84	5.9
10	GGGAUACAAAGUAUCCA[h]	−56.2	−160.9	−6.3	76.4	3.7	5.38	5.2
10	GGCCUUCGGCC	−36.1 ± 2.8	−102.1 ± 7.9	−4.38 ± 0.36	79.9	5.6	5.56	5.2
9	GCCGUUUCGGC	−43.4 ± 3.5	−125.7 ± 10.7	−4.40 ± 0.17	72.0	−4.5	4.64	4.7
8	GGCGACACGCC	−29.1 ± 3.1	−85.4 ± 9.7	−2.66 ± 0.22	68.2	9.8	6.38	5.9
8	GGCUCAAAGCC	−27.9 ± 2.7	−82.2 ± 8.5	−2.40 ± 0.12	66.1	10.9	5.91	5.9
7	GGCAUAUUGCC	−33.6 ± 1.4	−99.1 ± 4.5	−2.85 ± 0.06	65.7	5.1	5.49	5.9
7	GGCUUAUAGCC	−31.7 ± 1.6	−92.9 ± 4.7	−2.85 ± 0.14	67.7	7.1	5.46	5.9
7	GGCCUCCGGCC	−39.4 ± 2.5	−110.6 ± 7.3	−4.77 ± 0.19	80.1	2.3	5.17	5.2
6	GGCGAGACGCC	−37.0 ± 2.5	−109.3 ± 8.0	−3.11 ± 0.25	65.4	1.9	5.93	5.9
6	GGGAUACCC[g]	−27.6 ± 1.7	−88.3 ± 5.1	−0.26 ± 0.22	38.9	−0.8	6.26	5.9
6	GGCGCUUCGCC	−26.7 ± 2.8	−78.7 ± 8.8	−2.28 ± 0.12	66.0	12.2	6.76	5.9
5	GGCUACAAGCC	−30.7 ± 1.9	−90.7 ± 6.0	−2.61 ± 0.15	65.7	8.1	5.70	5.9
5	GCCCGAAGGGC	−43.4 ± 1.9	−122.7 ± 5.8	−5.35 ± 0.07	80.6	−1.7	4.59	5.2
5	GCCCUAUGGGC	−42.6 ± 1.9	−120.4 ± 6.8	−5.24 ± 0.17	80.6	−0.9	4.70	5.2
4	GGCAAAAUGCC	−30.0 ± 2.2	−89.2 ± 6.6	−2.36 ± 0.17	63.5	8.7	5.98	5.9
4	GGCGAAACGCC	−38.0 ± 2.3	−110.9 ± 7.3	−3.63 ± 0.13	69.8	0.9	5.41	5.9
4	GGCCACAGGCC	−38.2 ± 1.7	−109.8 ± 5.0	−4.16 ± 0.17	74.9	3.5	5.78	5.2
4	GGCGCAACGCC	−33.0 ± 1.9	−95.4 ± 5.7	−3.42 ± 0.15	72.9	5.9	5.62	5.9
4	GGCACAUUGCC	−31.6 ± 2.4	−93.8 ± 7.4	−2.47 ± 0.16	63.3	7.1	5.87	5.9
4	GGCCCUUGGCC	−51.8 ± 1.6	−149.3 ± 4.8	−5.47 ± 0.22	73.6	−10.1	4.47	5.2
2	GGCUAACGGCC	−34.5 ± 1.4	−101.8 ± 4.4	−2.91 ± 0.08	65.6	5.9	5.40	5.9
2	GGCGACCCGCC	−35.3 ± 1.2	−103.4 ± 3.6	−3.23 ± 0.19	68.3	3.6	5.81	5.9
1	GGCUACCAGCC	−30.9 ± 2.3	−91.1 ± 7.3	−2.63 ± 0.10	65.9	7.9	5.68	5.9
0	GGGAAAUCC[i]	−18.4 ± 2.9	−60.4 ± 9.2	0.3 ± 0.2	31.8	3.3	6.37	5.9
0	GGAGAAAUUCC[i]	−31.2 ± 3.6	−97.6 ± 11.9	−0.9 ± 0.4	46.1	−2.2	6.34	5.9

[a]Measurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA at pH 7.0. [b]Frequency of occurrence in the database described in Materials and Methods. [c]The sequences are written 5′ to 3′. The nucleotides in the triloop are underlined for easy identification. [d]The average T_M difference between the most and least concentrated sample for each stem−loop was 2.7 °C. If the sequences were forming a duplex, nearest neighbor thermodynamic values (*29*) predict that the T_M should change by about 16 °C for the typical concentration range used here, suggesting the formation of stem−loop structures. [e]The enthalpy and free energy contribution of the triloop were calculated by subtracting the Watson−Crick contribution of the stem (*29*) from the experimental enthalpy and free energy of the stem−loop, respectively. [f]Predicted triloop free energy contributions using the new predictive model shown in eq 4. [g]Reference *16*. [h]Reference *17*. [i]Reference *18*.

base pair (*16*−*18*, *30*−*38*), this categorization is most important. Categorizing triloops in this fashion results in 177 types of triloops in the database. The 10 triloop types listed in the first data set (Table 1) account for 45% of the total number of triloops found. The 167 types of triloops not shown account for the remaining 55%; however, each type represents <1.8% of the total number of triloops found.

The second set of data (Table 1) lists frequency and percent occurrence when only the triloop sequence is specified (the closing base pair is not considered). Categorizing triloops in this fashion results in 62 types of triloops in the database; however, there are 64 sequence possibilities. Therefore, two sequence possibilities, 5′AGG3′ and 5′CUG3′, are not found in the database. The 10 triloops listed in the second data set (Table 1) account for 62% of the total number of triloops found. The 52 types of triloops not shown account for the remaining 38%, with each triloop representing <2.5% of the total number of triloops found. If the occurrence of triloops were completely random, we would expect each triloop sequence to occur ~13 times in the database.

The third set of data (Table 1) lists frequency and percent occurrence of the closing base pair. Categorizing triloops in this fashion results in six types of closing base pairs in the database, representing all possible types. If the occurrence of triloops were completely random, we would expect each closing base pair to occur 137 times in the database.

The fourth set of data (Table 1) lists frequency and percent occurrence of the triloop nucleotides when A and G are categorized as purines (R) and C and U are categorized as pyrimidines (Y). Categorizing triloops in this fashion results in eight types of triloops, representing all possible combinations. If the occurrence of triloops were completely random, we would expect each type to occur ~100 times in the database.

Thermodynamic Parameters. As discussed earlier, due to possible bimolecular association of strands, many of the most frequently occurring triloops were not studied. All of the triloops studied here were found in the secondary structure database described above and were most likely in the hairpin conformation during the optical melting studies (as determined by using eqs 1−3). Table 2 shows the thermodynamic parameters of hairpin formation that were obtained from the average of fitting each melting curve to the two-state model. Data for 29 stem−loops containing unique triloops are shown in order of decreasing frequency as determined by the database search described above. Interestingly, all five of the triloops studied previously are of the sequence 5′ANA3′.

Contribution of Triloops to Stem−Loop Free Energy. The contributions of the 29 triloops to stem−loop stability are also listed in Table 2. The examination of the free energy contributions of triloops to stem−loop free energy indicates a large variance, with $\Delta G°_{37,triloop}$ ranging from 4.5 to 6.8 kcal/mol.

Interestingly, the 5′CUU3′ triloop is both the least and most stable triloop reported here, depending on its adjacent pair. It is the least stable triloop ($\Delta G^\circ_{37,\text{triloop}}$ = 6.8 kcal/mol) when adjacent to a G-C pair and the most stable triloop ($\Delta G^\circ_{37,\text{triloop}}$ = 4.5 kcal/mol) when adjacent to a C-G pair.

Updated Model for Predicting the Free Energy of Previously Unmeasured Triloops. Currently, in order to predict the free energy contribution of triloops, all triloops are assigned a free energy penalty of 5.4 kcal/mol. The only modification to this penalty is for a triloop containing all C residues, which includes an additional 1.5 kcal/mol penalty (*21*) and a lookup table for two unstable triloops, 5′CAACG3′ (6.8 kcal/mol) and 5′GUUAC3′ (6.9 kcal/mol) (*21*). When predicting the free energy contributions of the 29 triloops reported here, experimental values can now be used. For triloops that still do not have experimental values, a predictive model can be utilized.

The current predictive model was applied to all of the triloops reported here. On average, the predicted value was 0.5 ± 0.4 kcal/mol different than the experimental value. The average thermodynamic contribution of the triloops reported here was 5.6 kcal/mol. Using this average value in place of the 5.4 kcal/mol penalty (*21*) results in an average difference between the predicted value and the experimental value of 0.4 ± 0.4 kcal/mol.

In an attempt to further improve prediction, various other models were tried. The best model (lowest difference between experimental and predicted values with small deviations for the derived parameters) was based on the findings of Shu and Bevilacqua (*19*). They discovered that sequence preferences for stable triloops included a U-rich loop and C-G as the closing base pair. Based on these observations and the data reported here, the following model was derived:

$$\Delta G^\circ_{37,\text{triloop}} = \Delta G^\circ_{37,\text{i}} + \Delta G^\circ_{37,\text{UUU}} + \Delta G^\circ_{37,\text{C-G closure}} \quad (4)$$

Here, $\Delta G^\circ_{37,\text{i}}$ is the triloop initiation term, 5.9 ± 0.1 kcal/mol, $\Delta G^\circ_{37,\text{UUU}}$ is a −1.2 ± 0.3 kcal/mol bonus for a UUU triloop, and $\Delta G^\circ_{37,\text{C-G closure}}$ is a −0.7 ± 0.2 kcal/mol bonus for triloops adjacent to a C-G pair. Triloops adjacent to a G-C pair are not given the C-G bonus. Additional parameters and/or additional bonuses may be discovered with additional experiments. This model was used to predict the free energy contribution of the triloops measured here and the triloops measured previously (*16*−*18*) (Table 2). On average, the predicted value was 0.3 ± 0.2 kcal/mol different than the experimental value.

DISCUSSION

Database Searching. Due to the size and diversity of the RNA secondary structure database that was searched, we have assumed that the number and type of triloops found in this database are representative of triloops found in naturally occurring RNA.

From the onset of this study, we anticipated that it would be difficult to find triloop sequences that would outcompete bimolecular association of the RNA strand to form a duplex with a 3 × 3 nucleotide internal loop (or smaller depending on the sequence of the nucleotides) in 1 M NaCl. In fact, triloop formation outcompeted bimolecular association with only two of the ten most frequently occurring triloop sequences, and one of those had already been studied thermodynamically (*16*). Therefore, the triloops studied here are not the most frequently occurring triloops (Table 1, data set 1); however, all of the triloops studied here do occur in the database and do appear to outcompete bimolecular association of the RNA strand.

It is interesting to note that the most frequent triloop in this database was 5′GGG3′ (representing 13% of all triloops in the database). In a similar database containing tetraloops, 5′GGGG3′ was not found (*39*). In both the triloop and tetraloop database, a C-G base pair sat atop the list of closing base pairs, representing 49% and 42% of all closing base pairs in tetraloops and triloops, respectively. When categorizing the loop nucleotides as purines and pyrimidines (Table 1, data set 4), it is interesting to note that loop sequences containing all purines were the most common type for triloops (25%) as well as tetraloops (35%).

Thermodynamic Contributions of Triloops to Duplex Thermodynamics. From the data in Table 2, it is evident that the stability of a triloop alone does not determine its frequency of occurrence. For example, the most stable triloop reported here (5′CCUUG5′, $\Delta G^\circ_{37,\text{triloop}}$ = 4.5 kcal/mol) is only the 45th most common in the database. Interestingly, due to its abnormally low destabilizing free energy contribution, it is also the worst predicted (using eq 4) of those reported here.

The sequence preferences observed here for stable triloops are consistent with the general trends reported by Shu and Bevilacqua (*19*). For example, triloops containing all uracils in the loop were, on average, 0.9 kcal/mol more stable than other triloop sequences. Triloops containing two uracils were, on average, 0.2 kcal/mol more stable than other triloop sequences; therefore, only triloops with three uracils are given a bonus in the proposed model (eq 4). Similarly, triloops adjacent to a C-G pair were, on average, 0.6 kcal/mol more stable than triloops with other adjacent pairs. More specifically, when adjacent to a C-G pair, both an 5′AUA3′ triloop and an 5′AGA3′ triloop are 0.5 kcal/mol more stable than when adjacent to a G-C adjacent pair. Surprisingly, when adjacent to a C-G pair, a 5′CUU3′ triloop is 2.3 kcal/mol more stable than when adjacent to a G-C pair. Although this difference in stability is almost four times the average stabilization by a C-G closing pair, this magnitude of stabilization is not anomalous. When adjacent to a C-G pair, the tetraloop 5′UUCG3′ was found to be 2.3 kcal/mol more stable than when adjacent to a G-C pair in 10 mM NaCl (*32*). Similarly, when adjacent to a C-G pair, the 5′UUA3′ triloop found to be 2.1 kcal/mol more stable than when adjacent to a G-C pair in 10 mM NaCl (*19*).

Updated Model for Predicting Thermodynamics of Triloops. Because we have collected thermodynamic data for 24 triloops that previously did not have experimental values, when predicting the free energy contributions of these triloops in an RNA stem−loop, the experimental values can be used. For triloops that still do not have experimental values, the predictive model (eq 4) can be utilized. It is unclear why 5′UUU3′ is more stable than any other triloop sequence. Similarly, it is unclear why triloops with adjacent C-G pairs are more stable than triloops with adjacent G-C pairs. Structural studies are currently underway to help to understand these thermodynamic patterns. Additional studies with more triloop sequences and adjacent base pairs may result in additional free energy bonuses.

SUPPORTING INFORMATION AVAILABLE

A table listing all of the triloops found in the secondary structure database and the type of RNA in which each was found. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture of ribosomal RNA: Constraints on the sequence of tetra-loops. *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467–8471.
2. Wolters, J. (1992) The nature of preferred hairpin structures in 16S-like rRNA variable regions. *Nucleic Acids Res. 20*, 1843–1850.
3. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) A compilation of large subunit (23s and 23s-like) ribosomal-RNA structures—1993. *Nucleic Acids Res. 21*, 3055–3074.
4. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA. *J. Mol. Biol. 256*, 701–719.
5. Szymanski, M., Specht, T., Barciszewska, M. Z., Barciszewski, J., and Erdmann, V. A. (1998) 5S rRNA data bank. *Nucleic Acids Res. 26*, 156–159.
6. Larsen, N., Samuelsson, T., and Zwieb, C. (1998) The signal recognition particle database (SRPDB). *Nucleic Acids Res. 26*, 177–178.
7. Brown, J. W. (1998) The ribonuclease P database. *Nucleic Acids Res. 26*, 351–352.
8. Waring, R. B., and Davies, R. W. (1984) Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing—A review. *Gene 28*, 277–291.
9. Damberger, S. H., and Gutell, R. R. (1994) A comparative database of group I intron structures. *Nucleic Acids Res. 22*, 3508–3510.
10. Kim, C. H., Kao, C. C., and Tinoco, I. (2000) RNA motifs that determine specificity between an viral replicase and its promoter. *Nat. Struct. Biol. 7*, 415–423.
11. Huang, H., Alexandrov, A., Chen, X., Barnes, T. W., Zhang, H. Y., Dutta, K., and Pascal, S. M. (2001) Structure of an RNA hairpin from HRV-14. *Biochemistry 40*, 8055–8064.
12. McCallum, S. A., and Pardi, A. (2003) Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. *J. Mol. Biol. 326*, 1037–1050.
13. Convery, M. A., Rowsell, S., Stonehouse, N. J., Ellington, A. D., Hirao, I., Murray, J. B., Peabody, D. S., Phillips, S. E., and Stockley, P. G. (1998) Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat. Struct. Biol. 5*, 133–139.
14. Olsthoorn, R. C. L., and Bol, J. F. (2002) Role of an essential triloop hairpin and flanking structures in the 3′-untranslated region of alfalfa mosaic virus RNA in in vitro transcription. *J. Virol. 76*, 8747–8756.
15. Haasnoot, P. C. J., Bol, J. F., and Olsthoorn, R. C. L. (2003) A plant virus replication system to assay the formation of RNA pseudotriloop motifs in RNA-protein interactions. *Proc. Natl. Acad. Sci. U.S.A. 100*, 12596–12600.
16. Serra, M. J., Barnes, T. W., Betschart, K., Gutierrez, M. J., Sprouse, K. J., Riley, C. K., Stewart, L., and Temel, R. E. (1997) Improved parameters for the prediction of RNA hairpin stability. *Biochemistry 36*, 4844–4851.
17. Groebe, D. R., and Uhlenbeck, O. C. (1988) Characterization of RNA hairpin loop stability. *Nucleic Acids Res. 16*, 11725–11735.
18. Giese, M. R., Betschart, K., Dale, T., Riley, C. K., Rowan, C., Sprouse, K. J., and Serra, M. J. (1998) Stability of RNA hairpins closed by wobble base pairs. *Biochemistry 37*, 1094–1100.
19. Shu, Z., and Bevilacqua, P. C. (1999) Isolation and characterization of thermodynamically stable and unstable RNA hairpins from a triloops combinatorial library. *Biochemistry 38*, 15369–15379.
20. Davis, P. W., Thurmes, W., and Tinoco, I. (1993) Structure of a small RNA hairpin. *Nucleic Acids Res. 21*, 537–545.
21. 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.
22. Gutell, R. R. (1994) Collection of small-subunit (16s- and 16s-like) ribosomal-RNA structures: 1994. *Nucleic Acids Res. 22*, 3502–3507.
23. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res. 26*, 148–153.
24. Michel, F., Umesono, K., and Ozeki, H. (1989) Comparative and functional-anatomy of group-Ii catalytic introns—A review. *Gene 82*, 5–30.
25. Bevilacqua, P. C., and Blose, J. M. (2008) Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem. 59*, 79–103.
26. 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.
27. 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.
28. 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.
29. 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.
30. Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A., and Turner, D. H. (1993) RNA hairpin loop stability depends on closing base pair. *Nucleic Acids Res. 21*, 3845–3849.
31. Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., d'Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Borody, E. N., and Gold, L. (1988) CUUCGG hairpins: Extraordinary stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U.S.A. 85*, 1364–1368.
32. Antao, V. P., Lai, S. Y., and Tinoco, I. (1991) A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res. 19*, 5901–5905.
33. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry 30*, 3280–3289.
34. Heus, H. A., and Pardi, A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science 253*, 191–194.
35. Antao, V. P., and Tinoco, I. (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res. 20*, 819–824.
36. Williams, D. J., and Hall, K. B. (2000) Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J. Mol. Biol. 297*, 10445–11061.
37. Dale, T., Smith, R., and Serra, M. J. (2000) A test of the model to predict unusually stable RNA hairpin loop stability. *RNA 6*, 608–615.
38. Proctor, D. J., Schaak, J. E., Bevilacqua, J. M., Falzone, C. J., and Bevilacqua, P. C. (2002) Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry 41*, 12062–12075.
39. Sheehy, J. P., Davis, A. R., and Znosko, B. M. (2010) Thermodynamic characterization of naturally occurrring RNA tetraloops. *RNA 16*, 417–429.