Thermodynamic characterization of naturally occurring RNA tetraloops

JUSTIN P. SHEEHY, AMBER R. DAVIS, and BRENT M. ZNOSKO
Department of Chemistry, Saint Louis University, Saint Louis, Missouri 63103, USA

ABSTRACT

Although tetraloops are one of the most frequently occurring secondary structure motifs in RNA, less than one-third of the 30 most frequently occurring RNA tetraloops have been thermodynamically characterized. Therefore, 24 stem–loop sequences containing common tetraloops were optically melted, and the thermodynamic parameters ΔH°, ΔS°, ΔG°_{37}, and T_M for each stem–loop were determined. These new experimental values, on average, are 0.7 kcal/mol different from the values predicted for these tetraloops using the model proposed by Vecenie CJ, Morrow CV, Zyra A, Serra MJ. 2006. *Biochemistry* 45: 1400–1407. The data for the 24 tetraloops reported here were then combined with the data for 28 tetraloops that were published previously. A new model, independent of terminal mismatch data, was derived to predict the free energy contribution of previously unmeasured tetraloops. The average absolute difference between the measured values and the values predicted using this proposed model is 0.4 kcal/mol. This new experimental data and updated predictive model allow for more accurate calculations of the free energy of RNA stem–loops containing tetraloops and, furthermore, should allow for improved prediction of secondary structure from sequence. It was also shown that tetraloops within the sequence 5′-GCCNNNNGGC-3′ are, on average, 0.6 kcal/mol more stable than the same tetraloop within the sequence 5′-GGCNNNNGCC-3′. More systemic studies are required to determine the full extent of non-nearest-neighbor effects on tetraloop stability.

Keywords: hairpin; RNA; secondary structure; tetraloops

INTRODUCTION

Most biological RNA is single stranded. In order to fold into active secondary and tertiary structures, these single strands of RNA must fold back onto themselves. In doing so, hairpin loops are created at the end of most base-paired regions. Over 50% of these hairpins are tetraloops (Antao and Tinoco 1992; Wolters 1992). Therefore, RNA tetraloops are widespread and found quite frequently in nature. For example, tetraloops are found in the 16S rRNA of *Thermus thermophilus*, the 23S rRNA of *Deinococcus radiodurans*, the selenocysteine insertion sequence within the mRNA of prokaryotes (Fourmy et al. 2002), the 5′-UTR of coxsackievirus B3 (Du et al. 2003), the P5b stem–loop from a group I intron ribozyme (Kieft and Tinoco 1997), the recognition site for *Saccharomyces cerevisiae* RNase III (Wu et al. 2001), and the encapsidation signals of duck and heron Hepatitis B virus (Girard et al. 2007), to name a few. It is important to note, however, that RNA tetraloops not only occur in a variety of different RNAs and in different organisms, but they also serve functional roles within the RNA beyond allowing secondary structure formation. In general, tetraloops are extensively involved in RNA tertiary interactions with other RNAs and RNA interactions with proteins (Varani 1995; Tinoco and Bustamante 1999). Tetraloops also serve a variety of more specific functional roles, such as preventing reverse transcriptase from reading through mRNAs (Tuerk et al. 1988).

RNA tetraloops have been extensively studied structurally (Cheong et al. 1990; Heus and Pardi 1991; Varani et al. 1991; Allain and Varani 1995; Jucker and Pardi 1995; Cate et al. 1996; Jucker et al. 1996; Ennifar et al. 2000; Wimberly et al. 2000) and thermodynamically (Groebe and Uhlenbeck 1988; Tuerk et al. 1988; Antao et al. 1991; Heus and Pardi 1991; Varani et al. 1991; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000; Williams and Hall 2000; Proctor et al. 2002), especially the stable tetraloops 5′-GNRA-3′ and 5′-UNCG-3′. Nevertheless, less than one-third of the 30 most frequently occurring RNA tetraloops (when considering the identity of the nucleotides in the hairpin loop as well as the closing base pair) have been thermodynamically characterized. This work focuses on those frequently occurring tetraloops that have not yet

Reprint requests to: Brent M. Znosko, Department of Chemistry, Saint Louis University, Saint Louis, MO 63103, USA; e-mail: znoskob@slu.edu; fax: (314) 977-2521.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1773110.

been studied and provides thermodynamic data for frequently occurring tetraloops and a new model to predict the stability of tetraloops that do not yet have experimental data.

Non-nearest-neighbor effects have been observed previously for a wide variety of RNA secondary structure motifs (Longfellow et al. 1990; Kierzek et al. 1999; Badhwar et al. 2007; Davis and Znosko 2007; Siegfried et al. 2007; Wright et al. 2007); however, current algorithms that predict the stability of RNA from sequence (Zuker 1989; Mathews et al. 1999, 2004; Hofacker 2003; Zuker 2003; Lu et al. 2006) ignore non-nearest-neighbor effects. Hairpins are a convenient motif to investigate non-nearest-neighbor effects on motif stability because they are situated at the end of a helix and have nonnearest neighbors in only one direction. Conversely, internal loops and bulges are located within duplexes and contain nonnearest neighbors both 5′ and 3′ of the motif. Therefore, this work investigates nonnearest-neighbor effects on the stability of RNA tetraloops by studying frequently occurring tetraloops that have already been thermodynamically characterized. These tetraloops (along with their closing base pair) are placed within a different stem sequence of the same length and recharacterized. It was concluded that tetraloop stability does depend upon the sequence of the stem.

RESULTS

Database searching

A database of RNA secondary structures was searched for tetraloops. In this database, 4490 tetraloops were found, averaging over three occurrences per secondary structure. Table 1 shows a summary of the database results obtained. The first set of data in Table 1 lists frequency and percent occurrence when the tetraloop nucleotides and the closing base pair are specified. Because the stability of tetraloops depends on both the identity of the nucleotides in the loop and the closing base pair (Groebe and Uhlenbeck 1988; Tuerk et al. 1988; Antao et al. 1991; Heus and Pardi 1991; Varani et al. 1991; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000; Williams and Hall 2000; Proctor et al. 2002), this categorization is most important. Categorizing tetraloops in this fashion results in 524 types of tetraloops in the database. The 30 tetraloop types listed in the first data set (Table 1) account for 62% of the total number of tetraloops found. The 494 types of tetraloops not shown account for the remaining 38%; however, each type represents <0.6% of the total number of tetraloops found. When categorized in this manner, previous thermodynamic studies account for only 36% of the total number of tetraloops found, but after adding the data reported here, this percentage increases to 61%. This increase of 25% is the largest increase of the four sets of data discussed here and the most significant, since the identity of the closing base pair is critical to tetraloop stability. Similarly, previous thermodynamic studies characterized only nine types of tetraloops in the top 30, but after adding the data reported here, 24 of the tetraloops in the top 30 have been studied.

The second set of data (Table 1) lists frequency and percent occurrence when only the tetraloop sequence is specified (the closing base pair is not considered). Categorizing tetraloops in this fashion results in 212 types of tetraloops in the database; however, there are 256 sequence possibilities. Therefore, 44 sequence possibilities are not found in the database. The 30 tetraloops listed in the second data set (Table 1) account for 80% of the total number of tetraloops found. The 182 types of tetraloops not shown account for the remaining 20%, with each tetraloop representing <0.4% of the total number of tetraloops found. If the occurrence of tetraloops were completely random, we would expect each tetraloop sequence to occur 19 times in the database. When categorized in this manner, previous thermodynamic studies account for 66% of the total number of tetraloops found, but after adding the data reported here, this percentage increases slightly to 71%.

The third set of data (Table 1) lists frequency and percent occurrence of the closing base pair. Categorizing tetraloops in this fashion results in six types of closing base pairs in the database, representing all possible types. If the occurrence of tetraloops were completely random, we would expect each closing base pair to occur 832 times in the database. Previous studies have already characterized the stability of tetraloops adjacent to all six types of closing base pairs.

The fourth set of data (Table 1) lists frequency and percent occurrence of the tetraloop nucleotides when A and G are categorized as purines (R) and C and U are categorized as pyrimidines (Y). Categorizing tetraloops in this fashion results in 16 types of tetraloops, representing all possible combinations. If the occurrence of tetraloops were completely random, we would expect each type to occur 281 times in the database. When categorized in this manner, previous thermodynamic studies account for 80% of the total number of tetraloops found, but after adding the data reported here, this percentage increases slightly to 86%.

Thermodynamic parameters

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 33 stem–loops containing frequently occurring tetraloops are shown in order of decreasing frequency. However, only 24 unique tetraloops are represented, because seven tetraloops were melted in two different stems and an eighth tetraloop was melted three times (once in one stem and twice in a different stem).

Contribution of tetraloops to stem–loop free energy

The contributions of the 33 tetraloops to stem–loop stability are also listed in Table 2. The examination of the

free energy contributions of tetraloops to stem–loop free energy indicates a large variance, with $\Delta G^{\circ}_{37,\text{tetraloop}}$ ranging from 1.2 to 4.5 kcal/mol.

Updated model for predicting the free energy of previously unmeasured tetraloops

The most recent model for predicting tetraloop stability for tetraloops closed by Watson–Crick pairs is (Vecenie and Serra 2004):

$$\Delta G^{\circ}_{37,\text{tetraloop}} = \Delta G^{\circ}_{37,\text{i}} + \Delta G^{\circ}_{37,\text{MM}} - 0.8\,(\text{if the first mismatch is G•A or U•U}) - 0.8\,(\text{if the first mismatch is G•G and the loop is closed on the 5′ side by a purine}). \quad (1)$$

Here, $\Delta G^{\circ}_{37,\text{i}}$ is the tetraloop initiation term, 4.9 kcal/mol, and $\Delta G^{\circ}_{37,\text{MM}}$ is the free energy of stacking the first and last nucleotides of the tetraloop at the end of the stem. In other words, $\Delta G^{\circ}_{37,\text{MM}}$ accounts for the stacking of the first mismatch of the tetraloop with the closing base pair. This interaction is approximated by terminal mismatch data, the free energy contributions of single mismatches at the end of a helix. The most recent model for predicting tetraloop stability for tetraloops closed by G–U pairs is (Vecenie et al. 2006):

$$\Delta G^{\circ}_{37,\text{tetraloop}} = \Delta G^{\circ}_{37,\text{i}} - 0.8\,(\text{if the first mismatch is G•A}) - 0.8\,(\text{if the first mismatch is G•G and the loop is closed on the 5′side by a purine}). \quad (2)$$

Here, $\Delta G^{\circ}_{37,\text{i}}$ is the tetraloop initiation term, 4.9 kcal/mol. These models were used to predict the free energy contribution of the tetraloops reported here. On average, the predicted value was 0.7 kcal/mol different from the experimental value. It is important to note that these Equations 1 and 2 were derived from experimental data on hairpins with 4–8 nucleotides (nt) in the loop.

The tetraloops studied here were combined with tetraloops from the literature that had been melted in 1 M NaCl (Groebe and Uhlenbeck 1988; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000). Equations 1 and 2 were again used to predict the free energy contribution of the tetraloops. On average, the predicted value was 0.8 kcal/mol different from the experimental value. Because Equation 1 is dependent upon terminal mismatch data, additional data for terminal mismatches were collected. There were three terminal mismatch values that had not been measured: a C•A mismatch adjacent to an A–U pair, a C•A mismatch adjacent to a C–G pair, and a C•C mismatch adjacent to a C–G pair. Also, several terminal mismatches that were measured previously were measured in duplexes with low melting temperatures, and there was difficulty in fitting the lower baselines during data analysis (Vecenie and Serra 2004). Therefore, some of these terminal mismatches were remeasured. The data for these terminal mismatches can be found in Table 3. These data were combined with all of the published terminal mismatch data (Freier et al. 1986; SantaLucia et al. 1991; Serra et al. 1994; Serra and Turner 1995; Giese et al. 1998; Vecenie and Serra 2004; Vecenie et al. 2006), and a new compilation of terminal mismatch data can be found in Table 4.

Using the terminal mismatch data in Table 4, the new tetraloop data reported here, and the tetraloop data published previously (Groebe and Uhlenbeck 1988; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000), the values for the parameters in Equations 1 and 2 were rederived. The updated terminal mismatch-dependent model for predicting tetraloop stability for tetraloops closed by Watson–Crick pairs is:

$$\Delta G^{\circ}_{37,\text{tetraloop}} = \Delta G^{\circ}_{37,\text{i}} + \Delta G^{\circ}_{37,\text{MM}} - 0.5\,(\text{if the first mismatch is G•A or U•U}) - 0.8\,(\text{if the first mismatch is G•G and the loop is closed on the 5′side by a purine}). \quad (3)$$

Here, $\Delta G^{\circ}{}_{37,\text{i}}$ is the tetraloop initiation term, 5.2 kcal/mol, and $\Delta G^{\circ}{}_{37,\text{MM}}$ is the free energy of stacking the first mismatch of the tetraloop at the ends of the stem. The 0.8 kcal/mol bonus was incorporated into the previous model (Vecenie and Serra 2004) based upon data from larger hairpin loops, and this bonus was kept here because no tetraloops with this sequence context have yet to be measured. The updated terminal mismatch-dependent model for predicting tetraloop stability for tetraloops closed by G–U pairs is:

$$\Delta G^{\circ}_{37,\text{tetraloop}} = \Delta G^{\circ}_{37,\text{i}} - 0.5\,(\text{if the first mismatch is G•A}) - 0.8\,(\text{if the first mismatch is G•G and the loop is closed on the 5′side by a purine}). \quad (4)$$

Here, as above, $\Delta G^{\circ}_{37,\text{i}}$ is the tetraloop initiation term, 5.2 kcal/mol, and the 0.8 kcal/mol bonus was kept based on data from larger hairpin loops. These models were used to predict the free energy contribution of the tetraloops measured here and those measured previously (Groebe and Uhlenbeck 1988; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000). On average, the predicted value was 0.6 kcal/mol different from the experimental value. Although the updated values slightly improved prediction, we were interested in testing other models to see if an even better model could be derived.

Many combinations of parameters were tested, but the model described below resulted in predicted values that were closest to the experimental values. This model is

TABLE 1. Summary of database search results for RNA tetraloops[a]

Data set 1: Tetraloop with closing base pair				Data set 2: Tetraloop				Data set 3: Closing base pairs				Data set 4: Tetraloop nucleotides classified as Purine (R) or Pyrimidine (Y)			
Tetraloop[a]	Freqeuncy[b]	Percent[c]	Reference	Tetraloop[a]	Freqeuncy[b]	Percent[c]	Reference	Closing bp	Freqeuncy[b]	Percent[c]	Reference	Tetraloop[b]	Freqeuncy[b]	Percent[c]	Reference
C**GAAA**G	358	8.0	This work, Dale et al. (2000)	GAAA	802	16.1	This work, Dale et al. (2000)	C–G	2207	49.2	This work, Dale et al. (2000), Groebe and Uhlenbeck (1988), Antao and Tinoco (1992), Serra et al. (1997)	RRRR	1555	34.6	This work, Dale et al. (2000), Groebe and Uhlenbeck (1988)
C**GUGA**G	239	5.3	This work, Dale et al. (2000)	GUGA	399	8.0	This work, Dale et al. (2000)	G–C	903	20.1	This work, Antao and Tinoco (1992), Serra et al. (1997)	RYRR	1133	25.2	This work, Dale et al. (2000), Giese et al. (1998)
C**UUCG**G	211	4.7	This work, Dale et al. (2000)	GCAA	352	7.1	This work, Dale et al. (2000), Giese et al. (1998)	A–U	493	11.0	This work, Serra et al. (1997)	YYYR	525	11.7	This work, Dale et al. (2000), Antao and Tinoco (1992)
G**GAAA**C	191	4.3	This work	GAGA	307	6.2	This work, Dale et al. (2000)	U–G	431	9.6	This work, Giese et al. (1998)	RRYR	187	4.2	This work
C**GCAA**G	173	3.9	This work, Dale et al. (2000)	UUCG	280	5.6	This work, Dale et al. (2000), Antao and Tinoco (1992)	U–A	404	9.0	This work, Serra et al. (1997)	YRRR	157	3.5	
C**GAGA**G	156	3.5	This work, Dale et al. (2000)	GUAA	197	3.9	This work, Dale et al. (2000)	G–U	52	1.2	Giese et al. (1998)	YRYR	147	3.3	This work, Dale et al. (2000)
C**GUAA**G	101	2.2	This work, Dale et al. (2000)	GGAA	153	3.1	This work, Dale et al. (2000)	Previously[d]	4490	100.0		RYYR	131	2.9	Giese et al. (1998), Serra et al. (1997)
G**GCAA**C	95	2.1	This work	GCGA	126	2.5	This work, Dale et al. (2000)	New total[e]	4490	100.0		YRRY	99	2.2	This work
A**GAAA**U	91	2.0	This work	GGGA	113	2.3	This work, Dale et al. (2000)					YYRR	88	2.0	
G**GUGA**C	90	2.0	This work	GAAG	101	2.0	This work					YYYY	88	2.0	Groebe and Uhlenbeck (1988), Antao and Tinoco (1992)

(*continued*)

TABLE 1. *Continued*

Data set 1: Tetraloop with closing base pair				Data set 2: Tetraloop				Data set 3: Closing base pairs				Data set 4: Tetraloop nucleotides classified as Purine (R) or Pyrimidine (Y)			
Tetraloop[a]	Freqeuncy[b]	Percent[c]	Reference	Tetraloop[a]	Freqeuncy[b]	Percent[c]	Reference	Closing bp	Freqeuncy[b]	Percent[c]	Reference	Tetraloop[b]	Freqeuncy[b]	Percent[c]	Reference
G**GGGA**C	88	2.0	This work	AACA	91	1.8	This work					RRRY	87	1.9	
U**GAAA**A	84	1.9	This work	UACG	81	1.6	This work, Dale et al. (2000)					YYRY	74	1.6	
C**UACG**G	77	1.7	This work, Dale et al. (2000)	UAAC	57	1.1	This work					RRYY	59	1.3	
C**GCGA**G	75	1.7	This work, Dale et al. (2000)	GCCA	50	1.0						RYYY	59	1.3	
U**GAAA**G	75	1.7	This work	UCCG	43	0.9						RYRY	55	1.2	
C**GGAA**G	74	1.6	Dale et al. (2000)	CUUG	41	0.8						YRYY	46	1.0	
A**AACA**U	71	1.6	This work	UUAG	38	0.8						Previously[d]	3579	79.7	
U**GAGA**G	63	1.4	This work	AGCA	37	0.7						New total[e]	3865	86.1	
C**GAAG**G	59	1.3	This work	CUCA	35	0.7									
A**GCAA**U	52	1.2	This work	AAAA	33	0.7	Groebe and Uhlenbeck (1988)								
G**GAGA**C	51	1.1	This work	UGAG	32	0.6									
A**GUAA**U	39	0.9		CUCG	31	0.6									
G**GGAA**C	39	0.9	This work	AAAC	30	0.6									
C**UAAC**G	38	0.8	This work	UUUA	29	0.6									
A**AGCA**U	37	0.8		UGAA	29	0.6									
G**GUAA**C	36	0.8	This work	ACCA	29	0.6	Giese et al. (1998)								
A**GUGA**U	35	0.8		UUUU	26	0.5	Groebe and Uhlenbeck (1988), Antao and Tinoco (1992)								
C**GCCA**G	32	0.7		UUUG	23	0.5	Antao and Tinoco (1992)								
A**GAGA**U	29	0.6		AACC	23	0.5									
C**CUCA**G	29	0.6		CAAG	20	0.4									
Previously[d]	1605	35.7		Previously[d]	2976	66.3									
New total[e]	2728	60.8		New total[e]	3225	71.8									

[a]Not all combinations in Datasets 1 and 2 are shown due to space limitations. For each set of sequences, the strand is written 5′–3′.
[b]Frequency of occurrence in the database.
[c]Percent, out of 4490 tetraloops, the total number of tetraloops found in the database.
[d]The number and percentage represented by previous studies.
[e]The number and percentage represented by combining the previous studies with the data reported here.

TABLE 2. Thermodynamic parameters for stem–loop formation and contributions of tetraloops to stem–loop stability[a]

Frequency[b]	Sequence[c]	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)	T_M (°C)[d]	$\Delta G^{\circ}_{37,tetraloop}$ (kcal/mol)[e]
358	GCCGAAAGGC	−35.4 ± 2.2	−101.7 ± 6.4	−3.85 ± 0.18	74.8	2.83
	GGCGAAAGCC[f]	−33.4 ± 1.7	−97.4 ± 4.0	−3.2 ± 0.2	70.3	3.48
239	GCCGUGAGGC	−38.3 ± 1.8	−111.5 ± 5.5	−3.73 ± 0.16	70.5	2.95
	GGCGUGAGCC[f]	−35.6 ± 2.9	−104.6 ± 8.7	−3.2 ± 0.3	67.8	3.48
211	GCCUUCGGGC	−49.3 ± 3.4	−141.5 ± 11.0	−5.44 ± 0.18	75.4	1.24
	GGCUUCGGCC[f]	−40.1 ± 3.8	−116.1 ± 11.1	−4.1 ± 0.3	72.6	2.58
191	GCGGAAACGC	−33.8 ± 1.7	−102.3 ± 5.3	−2.05 ± 0.09	57.1	3.73
173	GCCGCAAGGC	−33.6 ± 3.4	−96.7 ± 10.1	−3.60 ± 0.27	74.3	3.08
	GCCGCAAGGC	−37.1 ± 1.4	−107.2 ± 4.6	−3.80 ± 0.11	72.4	2.88
	GGCGCAAGCC[f]	−34.3 ± 2.5	−99.8 ± 7.6	−3.4 ± 0.3	71.0	3.28
156	GCCGAGAGGC	−33.9 ± 2.5	−98.9 ± 6.4	−3.23 ± 0.14	69.6	3.45
	GGCGAGAGCC[f]	−35.5 ± 2.5	−103.8 ± 7.6	−3.3 ± 0.2	68.8	3.38
101	GCCGUAAGGC	−36.4 ± 3.3	−104.7 ± 9.9	−3.91 ± 0.29	74.4	2.77
	GGCGUAAGCC[f]	−35.1 ± 3.6	−101.8 ± 10.9	−3.5 ± 0.3	71.7	3.18
95	GCGGCAACGC	−34.4 ± 2.9	−103.5 ± 8.6	−2.25 ± 0.27	58.8	3.53
91	GCAGAAAUGC	−34.3 ± 3.8	−105.8 ± 12.0	−1.50 ± 0.10	51.2	3.58
90	GCGGUGACGC	−31.9 ± 2.5	−97.4 ± 8.1	−1.72 ± 0.03	54.7	4.06
88	GCGGGGACGC	−48.3 ± 6.2	−149.0 ± 19.7	−2.05 ± 0.18	50.7	3.73
84	GCUGAAAAGC	−32.9 ± 2.6	−101.1 ± 8.3	−1.50 ± 0.07	51.9	3.55
77	GCCUACGGGC	−39.7 ± 0.9	−114.2 ± 2.9	−4.27 ± 0.08	74.4	2.41
	GGCUACGGCC[f]	−33.3 ± 5.0	−97.0 ± 12.5	−3.2 ± 0.4	69.8	3.48
75	GCCGCGAGGC	−40.2 ± 1.8	−116.4 ± 5.7	−4.11 ± 0.11	72.3	2.57
	GGCGCGAGCC[f]	−36.9 ± 2.2	−107.7 ± 6.9	−3.4 ± 0.2	68.7	3.28
75	GCUGAAAGGC	−34.1 ± 3.8	−105.7 ± 12.2	−1.30 ± 0.05	49.3	3.78
74	GGCGGAAGCC[f]	−33.5 ± 2.3	−97.6 ± 6.9	−3.3 ± 0.2	70.6	3.38
71	GCAAACAUGC	−27.7 ± 10.7	−87.4 ± 34.5	−0.60 ± 0.09	43.9	4.48
63	GCUGAGAGGC	−35.1 ± 3.4	−109.9 ± 10.9	−1.27 ± 0.07	48.6	3.81
59	GCCGAAGGGC	−31.3 ± 2.2	−91.8 ± 6.6	−2.88 ± 0.14	68.4	3.80
52	GCAGCAAUGC	−32.5 ± 2.7	−100.2 ± 8.8	−1.43 ± 0.05	51.3	3.65
51	GCGGAGACGC	−36.1 ± 2.8	−109.7 ± 8.7	−2.02 ± 0.12	55.4	3.76
39	GCGGGAACGC	−33.8 ± 3.3	−102.7 ± 10.5	−1.90 ± 0.10	55.5	3.88
38	GCCUAACGGC	−32.8 ± 1.4	−96.5 ± 4.3	−2.88 ± 0.12	66.8	3.80
36	GCGGUAACGC	−31.6 ± 4.1	−96.7 ± 13.2	−1.61 ± 0.11	53.7	4.17

[a]Measurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2 EDTA (pH 7.0).
[b]Frequency of occurrence in the database is described in Materials and Methods.
[c]The sequences are written 5′–3′.
[d]Calculated at 10^{-4} M oligomer concentration.
[e]The free energy contribution of the tetraloop was calculated by subtracting the Watson–Crick contribution of the stem (Xia et al. 1998) from the experimental free energy of the stem–loop.
[f]Dale et al. (2000).

independent of the closing base pair (there is only one equation used for tetraloops, regardless of whether the closing base pair is a Watson–Crick pair or a G–U pair) and independent of terminal mismatch data. The proposed terminal mismatch-independent model is:

$$\Delta G^{\circ}_{37,tetraloop} = \Delta G^{\circ}_{37,i} + \Delta G^{\circ}_{37,first\ mismatch\ bonus} + \Delta G^{\circ}_{37,nonnearest\ neighbors}. \quad (5)$$

Here, $\Delta G^{\circ}_{37,i}$ is the tetraloop initiation term, 4.8 kcal/mol, $\Delta G^{\circ}_{37,first\ mismatch\ bonus}$ is a −1.1 kcal/mol bonus for a G•A or U•U first mismatch and a −1.7 kcal/mol bonus for a U•G first mismatch. $\Delta G^{\circ}_{37,nonnearest\ neighbors}$ is a −0.7 kcal/mol bonus for tetraloops in the sequence 5′-GCC NNNNGGC-3′. Additional $\Delta G^{\circ}_{37,nonnearest\ neighbors}$ parameters may be discovered with additional experiments. This model was used to predict the free energy contribution of the tetraloops measured here and the tetraloops measured previously (Groebe and Uhlenbeck 1988; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000). On average, the predicted value was 0.4 kcal/mol different from the experimental value. A comparison between the measured tetraloop contributions and the contributions predicted by the previous model, updated previous model, and current model can be found in Table 5.

DISCUSSION

Database searching

Due to the size and diversity of the RNA secondary structure database that was searched, we have assumed

TABLE 3. Thermodynamic parameters for duplex formation and contributions of terminal mismatches to duplex stability[a]

	Analysis of melt curve fit/errors				Analysis of T_m dependence/ errors (Ln plot)				
Sequence[b]	ΔH° (kcal/mol)	ΔS° (cal/Kmol)	ΔG°_{37} (kcal/mol)	T_M (°C)[c]	ΔH° (kcal/mol)	ΔS° (cal/Kmol)	ΔG°_{37} (kcal/mol)	T_M (°C)[c]	$\Delta G^\circ_{37,MM}$ (kcal/mol)[d]
CUGGCCA**A** **A**ACCGGU**C**	−56.2 ± 3.7	−147.9 ± 11.0	−10.28 ± 0.32	64.7	−56.2 ± 1.7	−148.0 ± 5.0	−10.29 ± 0.13	64.7	−0.77
CUCGAA[e] **A**AGCU**C**	−27.9 ± 2.5	−78.6 ± 8.3	−3.48 ± 0.20	14.3	−28.0 ± 2.4	−79.0 ± 8.5	−3.45 ± 0.20	14.1	−0.91
GUGGCCA**A** **A**ACCGGU**G**	−58.0 ± 5.6	−153.7 ± 17.0	−10.31 ± 0.39	64.0	−57.2 ± 2.9	−151.1 ± 8.7	−10.28 ± 0.22	64.2	−0.77
GUCGAA[e] **A**AGCU**G**	−28.4 ± 2.1	−80.4 ± 6.9	−3.46 ± 0.2	14.6	−32.6 ± 2.9	−95.2 ± 10.2	−3.09 ± 0.3	14.2	−0.73
CUGGCCA**C** **C**ACCGGU**C**	−57.4 ± 3.1	−151.5 ± 9.2	−10.36 ± 0.24	64.6	−56.6 ± 1.3	−149.2 ± 3.8	−10.30 ± 0.10	64.6	−0.78
UUGGCCA**C** **C**ACCGGU**U**	−56.2 ± 2.1	−148.3 ± 6.1	−10.24 ± 0.19	64.4	−56.8 ± 1.5	−150.0 ± 4.5	−10.27 ± 0.12	64.3	−0.77
UUCGA**C**[e] **C**AGCU**U**	−34.4 ± 2.6	−103.7 ± 8.6	−2.26 ± 0.2	9.0	−30.2 ± 2.8	−88.6 ± 10.1	−2.68 ± 0.3	9.0	−0.52
GUAGCUA**G** **G**AUCGAU**G**	−48.3 ± 4.3	−135.1 ± 13.3	−6.35 ± 0.19	41.4	−51.9 ± 2.7	−146.5 ± 8.7	−6.40 ± 0.05	41.4	−0.79
GUCGA**G**[e] **G**AGCU**G**	−31.2 ± 2.8	−88.0 ± 9.4	−3.96 ± 0.2	20.9	−33.8 ± 2.9	−96.9 ± 10.2	−3.72 ± 0.3	20.0	−1.04
CUAGCUA**U** **U**AUCGAU**C**	−52.2 ± 8.5	−147.4 ± 27.8	−6.51 ± 0.29	42.0	−46.9 ± 4.1	−130.2 ± 12.9	−6.52 ± 0.13	42.7	−0.85
CUCGA**U**[d] **U**AGCU**C**	−36.4 ± 3.0	−107.6 ± 11.2	−3.02 ± 0.5	15.9	−36.6 ± 4.6	−108.5 ± 16.3	−2.99 ± 0.5	15.9	−0.68
UUAGCUA**U** **U**AUCGAU**U**	−57.0 ± 18.4	−163.2 ± 58.2	−6.38 ± 0.47	40.9	−53.4 ± 9.5	−151.5 ± 30.1	−6.42 ± 0.49	41.4	−0.80
UUCGA**U**[e] **U**AGCU**U**	−32.2 ± 2.2	−92.1 ± 7.9	−3.64 ± 0.3	18.6	−34.9 ± 3.4	−101.3 ± 11.7	−3.44 ± 0.3	18.3	−0.90
AGAGCUC**C** **C**CUCGAG**A**	−61.3 ± 3.5	−166.6 ± 10.7	−9.61 ± 0.18	58.2	−64.3 ± 1.5	−175.7 ± 4.5	−9.75 ± 0.09	58.0	−1.00
CGGCC**C** **C**CCGG**C**	−43.1 ± 3.1	−114.7 ± 9.7	−7.47 ± 0.12	50.5	−44.1 ± 0.8	−118.1 ± 2.4	−7.46 ± 0.03	50.1	−1.02
CCCGG**A** **A**GGCC**C**	−45.9 ± 2.9	−126.5 ± 9.1	−6.70 ± 0.11	44.1	−46.1 ± 2.2	−127.2 ± 6.9	−6.69 ± 0.05	44.0	−1.17
CCCGG**A**[f] **A**GGCC**C**	−45.8 ± 3.4	−125.2 ± 11.0	−6.90 ± 0.2	45.8	−43.8 ± 3.8	−119.0 ± 12.2	−6.9 ± 0.1	45.9	−1.20
CCCGG**C** **C**GGCC**C**	−44.8 ± 3.1	−123.7 ± 9.7	−6.46 ± 0.07	42.5	−46.1 ± 1.1	−127.7 ± 3.6	−6.46 ± 0.02	42.4	−1.05
CCCGG**C**[f] **C**GGCC**C**	−38.8 ± 2.4	−104.1 ± 7.9	−6.50 ± 0.2	43.8	−33.8 ± 3.8	−88.0 ± 12.1	−6.5 ± 0.3	44.6	−1.00
UCCGG**C** **C**GGCC**U**	−45.8 ± 2.3	−127.5 ± 7.4	−6.20 ± 0.10	40.6	−45.2 ± 1.6	−125.7 ± 5.2	−6.19 ± 0.02	40.6	−0.92
UCCGG**C**[f] **C**GGCC**U**	−40.6 ± 3.1	−110.1 ± 10.2	−6.40 ± 0.2	43.0	−36.6 ± 4.0	−97.3 ± 13.0	−6.4 ± 0.3	43.6	−0.90

[a]Measurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2 EDTA (pH 7.0).
[b]The terminal mismatch is identified by bold letters. The top strand of each duplex is written 5′–3′ and each bottom strand is written 3′–5′.
[c]Calculated at 10^{-4} M oligomer concentration.
[d]The free energy contribution of the terminal mismatch was calculated by subtracting the Watson–Crick free energy contribution (Xia et al. 1998) from the experimental free energy and dividing by 2.
[e]Vecenie and Serra (2004).
[f]Vecenie et al. (2006).

that the number and type of tetraloops found in this database are representative of tetraloops found in naturally occurring RNA.

It is clear from the first set of data in Table 1 that only one previous thermodynamic study (Dale et al. 2000) has focused on the tetraloop-closing base-pair combinations that occur most frequently in nature. When looking at the second set of data in Table 1, however, it appears as if the most frequent tetraloops (considering the loop nucleotides only) have already been studied. It is well known that the 5′-GNRA-3′ and 5′-UNCG-3′ tetraloops are found frequently in nature, and this is confirmed by the fact that

TABLE 4. Updated compilation for free energy contributions of terminal mismatches

Base pair	Y→ X↓	ΔG°_{37} (kcal/mol) A	C	G	U
AX					
UY					
	A	−0.8[a]	−0.8[b]	−0.8[b]	—
	C	−0.6[c]	−0.8	—	−0.6[b]
	G	−0.8[a]	—	−0.9[b]	−0.5[c]
	U	—	−0.8[b]	−1.0[c]	−0.9[b]
CX					
GY					
	A	−1.5[d]	−1.5[d]	−1.4[d]	—
	C	−1.0	−1.0	—	−0.8[d]
	G	−1.4[e]	—	−1.6[d]	−1.2[c]
	U	—	−1.4[d]	−1.9[c]	−1.2[a]
GX					
CY					
	A	−1.1[a]	−1.2[f]	−1.3[a]	—
	C	−1.1[a]	−1.0[f]	—	−0.9[f]
	G	−1.6[a]	—	−1.4[a]	−1.4[c]
	U	—	−1.2[g]	−2.1[c]	−1.0[g]
UX					
AY					
	A	−1.0[a]	−0.8[a]	−1.1[a]	—
	C	−0.7[a]	−0.6[a]	—	−0.5[a]
	G	−1.1[a]	—	−1.2[a]	−0.8[c]
	U	—	−0.6[a]	−1.1[c]	−0.5[a]
GX					
UY					
	A	−0.3[h]	−0.6[g]	−0.7[g]	—
	C	−1.1[g]	−1.1[g]	—	−0.9[g]
	G	−0.6[h]	—	−0.7[g]	
	U	—	−0.7[g]		−0.8[g]
UX					
GY					
	A	−0.8[h]	−0.8[g]	−0.9[g]	—
	C	−0.5[g]	−0.7[g]	—	−0.3[g]
	G	−0.5[h]	—	−0.8[h]	
	U	—	−0.7[g]		−0.6[g]

[a]Freier et al. (1986).
[b]Average of updated values from Vecenie and Serra (2004) and this work.
[c]Serra and Turner (1995).
[d]Serra et al. (1994).
[e]SantaLucia et al. (1991).
[f]Average of values from Vecenie et al. (2006) and this work.
[g]Vecenie et al. (2006).
[h]Giese et al. (1998).

nine out of the top 10 tetraloops fall into one of these two categories. However, it has been shown that the stability of hairpin loops depends not only on the identity of the nucleotides in the loop, but also on the stacking of the first mismatch on the closing base pair (Mathews et al. 1999, 2004; Vecenie and Serra 2004; Vecenie et al. 2006). Therefore, this work focuses on frequently occurring tetraloops when considering both the nucleotides in the loop and the closing base pair.

It is interesting to note that a closing base pair of C–G occurs as frequently as all five of the remaining base pairs combined (see Table 1, data set 3). Recently, Blose et al. (2009) have investigated the molecular basis for the enhanced stability of tetraloops with C–G closing base pairs. Although they may be more stable, it is unclear why C–G closing base pairs are so common in nature, as nature does not select tetraloop sequences based solely on stability (see Table 2).

When categorizing the loop nucleotides as purines and pyrimidines (Table 1, data set 4), it is interesting to note that 5′-RRRR-3′ and 5′-RYRR-3′, representing all of the 5′-GNRA-3′ tetraloops, occurs three and two times as often as the third most frequent type (5′-YYYR-3′, a type of 5′-UNCG-3′), respectively, and eight and six times as often as the fourth most frequent type (5′-RRYR-3′), respectively.

Thermodynamic contributions of tandem mismatches to duplex thermodynamics

From the data in Table 2, it is evident that the stability of a tetraloop alone does not determine its frequency of occurrence. For example, the most stable tetraloop (5′-CUUCGG-5′, $\Delta G^{\circ}_{37,\text{tetraloop}}$ = 1.2 kcal/mol) is only the third most common in the database. Also, one of the most stable tetraloops measured, 5′-GACCAU-3′ (Giese et al. 1998) ($\Delta G^{\circ}_{37,\text{tetraloop}}$ = 2.4 kcal/mol), is the 95$^{\text{th}}$ most common tetraloop in the database. Similarly, a tetraloop that contributes an unfavorable 4.1 kcal/mol toward stem–loop stability (5′-GGUGAC-5′) appears in the top 10 in frequency of occurrence. However, stability may play a partial role in determining frequency of occurrence; the measured tetraloops in the top 30, on average, contribute 3.4 kcal/mol toward stem–loop stability, while the measured tetraloops outside the top 30 contribute 4.2 kcal/mol toward stem–loop stability.

Non-nearest-neighbor effects on the stability of tetraloops

Previous studies have shown that the stability of the iron responsive element hairpin loop is dependent upon the stem sequence. For example, the hexaloop contributes 2.4 kcal/mol to stem–loop stability when in the sequence 5′-GAAGACAGUGCUCUUC-3′ (Laing and Hall 1996). When placed in the sequence 5-′GGACAGUGCUCC-3′, the hexaloop contributes 3.8 kcal/mol to stem–loop stability (Dale et al. 2000). The effect of the stem sequence on tetraloop stability was studied here. One tetraloop reported here was synthesized twice within the same stem (5′-GCC GCAAGGC-3′), and each sample was purified, melted, and

TABLE 5. Predicted values for all measured tetraloops

Frequency[a]	Sequence[b]	$\Delta G^{\circ}_{37,\text{tetraloop}}$ measured[c] (kcal/mol)	$\Delta G^{\circ}_{37,\text{tetraloop}}$ previous model[d] (kcal/mol)	$\Delta G^{\circ}_{37,\text{tetraloop}}$ updated previous model[e] (kcal/mol)	$\Delta G^{\circ}_{37,\text{tetraloop}}$ new model[f] (kcal/mol)
358	GGCGAAAGGC	2.83	2.7	3.3	3.0
	GGCGAAAGCC[g]	3.48	2.7	3.3	3.7
239	GCCGUGAGGC	2.95	2.7	3.3	3.0
	GGCGUGAGCC[g]	3.48	2.7	3.3	3.7
211	GCCUUCGGGC	1.24	3.0	3.3	2.4
	GGCUUCGGCC[g]	2.58	3.0	3.3	3.1
191	GCGGAAACGC	3.73	2.5	3.1	3.7
173	GCCGCAAGGC	3.08	2.7	3.3	3.0
	GCCGCAAGGC	2.88	2.7	3.3	3.0
	GGCGCAAGCC[g]	3.28	2.7	3.3	3.7
156	GCCGAGAGGC	3.45	2.7	3.3	3.0
	GGCGAGAGCC[g]	3.38	2.7	3.3	3.7
101	GCCGUAAGGC	2.77	2.7	3.3	3.0
	GGCGUAAGCC[g]	3.18	2.7	3.3	3.7
95	GCGGCAACGC	3.53	2.5	3.1	3.7
91	GCAGAAAUGC	3.58	3.3	3.9	3.7
90	GCGGUGACGC	4.06	2.5	3.1	3.7
88	GCGGGGACGC	3.73	2.5	3.1	3.7
84	GCUGAAAAGC	3.55	3.0	3.6	3.7
77	GCCUACGGGC	2.41	3.0	3.3	2.4
	GGCUACGGCC[g]	3.48	3.0	3.3	3.1
75	GCCGCGAGGC	2.57	2.7	3.3	3.0
	GGCGCGAGCC[g]	3.28	2.7	3.3	3.7
75	GCUGAAAGGC	3.78	4.1	4.7	3.0
74	GGCGGAAGCC[g]	3.38	2.7	3.3	3.7
71	GCAAACAUGC	4.48	4.1	4.4	4.8
63	GCUGAGAGGC	3.81	4.1	4.7	3.0
59	GCCGAAGGGC	3.80	3.3	3.6	4.1
52	GCAGCAAUGC	3.65	3.3	3.9	3.7
51	GCGGAGACGC	3.76	2.5	3.1	3.7
39	GCGGGAACGC	3.88	2.5	3.1	3.7
38	GCCUAACGGC	3.80	3.5	3.8	4.1
36	GCGGUAACGC	4.17	2.5	3.1	3.7
23	GGGAUACAAAAGUAUCCA[h]	4.08	3.4	3.7	4.8
23	GGAGUUCGCUCC[i]	3.69	2.8	3.1	3.1
20	GGCGGGAGCC[g]	3.48	2.7	3.3	3.7
17	GGACUUUGGUCC[i]	3.65	3.0	3.3	3.1
15	GGACUUUUGUCC[i]	3.65	2.9	3.5	3.7
	GGGAUACUUUUGUAUCCA[h]	3.48	2.9	3.5	4.8
14	GGUGCAAGCC[j]	3.92	4.1	4.7	3.7
12	GGCUCCGGCC[g]	2.68	3.0	3.3	3.1
6	GGGACCAUCC[j]	2.44	4.9	5.2	4.8
5	GGCAUUAGCC[k]	4.17	3.4	3.7	4.8
3	GGUAUUAACC[k]	4.73	3.9	4.2	4.8
1	GGGAUACCCCCGUAUCCA[h]	7.38	3.9	4.1	4.8
1	GGACGCUUGUCC[i]	3.45	3.7	4.0	3.1
1	GGGAUUACCC[k]	5.22	3.8	4.1	4.8
0	GGAAUUAUCC[k]	5.11	4.1	4.4	4.8
0	GCGAUUAUGC[j]	4.08	4.9	5.2	4.8
	GCGGAUUAUCGC[j]	5.46	4.9	5.2	4.8
0	GGUAUUAGCC[j]	5.52	4.9	5.2	4.8
0	GGCGAAAGCC[g]	3.48	2.7	3.3	3.7

[a]Frequency of occurrence in the database is described in Materials and Methods.
[b]The sequences are written 5′–3′.
[c]The free energy contribution of the tetraloop was calculated by subtracting the Watson–Crick contribution of the stem (Xia et al. 1998) from the experimental free energy of the stem–loop.
[d]The free energy of the tetraloop was predicted using the model proposed previously (Vecenie and Serra 2004; Vecenie et al. 2006) (see Eqs. 1, 2).
[e]The free energy of the tetraloop was predicted using the model proposed previously (Vecenie and Serra 2004; Vecenie et al. 2006) after the values for the parameters were rederived (Eqs. 3, 4) based on the data reported here.
[f]The free energy of the tetraloop was predicted using the model proposed in Equation 5.
[g]Dale et al. (2000).
[h]Groebe and Uhlenbeck (1988).
[i]Antao and Tinoco (1992).
[j]Giese et al. (1998).
[k]Serra et al. (1997).

analyzed separately. The $\Delta G^{\circ}_{37,tetraloop}$ values for these two samples varied by only 0.2 kcal/mol, showing that the data are reproducible. Eight tetraloops in the top 30 that were previously characterized (Dale et al. 2000) were remeasured here in a different stem. The sequence used earlier was 5′-GGCNNNNGCC-3′, and the sequence used here is 5′-GCCNNNNGGC-3′. When comparing the two stem sequences, they have the following in common: three G–C pairs, the number and type of nearest-neighbor combinations, a closing C–G pair, and melting conditions. The only difference is that a G–C pair 2 base pairs (bp) away from the hairpin loop in the tetraloops studied previously is switched to a C–G pair in the tetraloops studied here. Interestingly, when tetraloops were placed within the sequence studied here, 5′-GCCNNNNGGC-3′, tetraloops were on average 0.6 kcal/mol more stable than the sequence used previously. It appears as if the orientation of the nonnearest neighbors plays a role in the stability of tetraloops. The $\Delta G^{\circ}_{37,nonnearest\ neighbors}$ term in Equation 5 takes into account this added stability. Because nonnearest-neighbor effects are not well understood and a significant amount of additional data are required, more studies are needed in order to more fully examine the effects of nonnearest neighbors on the stability of tetraloops.

Updated model for predicting thermodynamics of tetraloops

Because we have collected thermodynamic data for 15 tetraloops that previously did not have experimental values, when predicting the free energy contributions of these tetraloops in an RNA stem–loop, the experimental values can be used. These new experimental values, on average, are 0.8 kcal/mol different from the values predicted for these mismatches using the previous model (Vecenie et al. 2006). For tetraloops that still do not have experimental values, the predictive model can be utilized.

Using the data reported here and the data for previously measured tetraloops (Groebe and Uhlenbeck 1988; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000), the accuracy of the current model (Vecenie et al. 2006) (Eqs. 1, 2) to predict tetraloop stability was tested. On average, this model predicted the experimental free energies within 0.8 kcal/mol. Because this model was derived from data for hairpins of various sizes, and because the database of tetraloop experimental data has nearly doubled with the data reported here, the values for this model were recalculated using tetraloop data only. On average, this model (Eqs. 3, 4) predicted the experimental free energies within 0.6 kcal/mol.

Other models to predict tetraloop stability were also investigated. The model that gave the closest prediction to the experimental free energies is shown in Equation 5. Although more studies are needed to determine if additional $\Delta G^{\circ}_{37,nonnearest\ neighbors}$ parameters should be included, this model, on average, predicted the experimental free energies within 0.4 kcal/mol. Interestingly, this model does not depend upon the terminal mismatch data as does previous models. The conformation of a terminal mismatch and the interaction between the terminal mismatch and adjacent base pair is likely different from the conformation of the first mismatch in a tetraloop and the interaction between the first mismatch and the closing base pair. Therefore, using terminal mismatch data to predict tetraloop stability may not be the best approach. Not only does the model proposed here (Eq. 5) result in more accurate free energy predictions, but because it relies on free energy bonuses derived from tetraloop data and not on data from terminal mismatches, the new model may be a more realistic approach.

Because this model is derived from tetraloop data only, and most of the tetraloops that have been studied thermodynamically are 5′-GNRA-3′ and 5′-UNCG-3′, this model may be biased for these loops. In order to test for this bias, the available thermodynamic database for tetraloops was broken down into 5′-GNRA-3′, 5′-UNCG-3′, and all other tetraloops. For the 28 5′-GNRA-3′ tetraloops, the average difference between the predicted and experimental values are 0.7, 0.4, and 0.3 kcal/mol for the previous model (Eqs. 1, 2), updated previous model (Eqs. 3, 4), and the model proposed here (Eq. 5), respectively. For the six 5′-UNCG-3′ tetraloops, the average difference between the predicted and experimental values are 0.7, 0.7, and 0.5 kcal/mol for the previous model, updated previous model, and the model proposed here, respectively. For the 18 other tetraloops, the average difference between the predicted and experimental values are 0.9, 0.8, and 0.7 kcal/mol for the previous model, updated previous model, and the model proposed here, respectively. This shows that although 54% of the tetraloops used to derive the proposed model are 5′-GNRA-3′ tetraloops, it predicts the stability of 5′-GNRA-3′, 5′-UNCG-3′, and all other tetraloops better than both the previous model and updated previous model.

MATERIALS AND METHODS

Compiling and searching a database for RNA tetraloops

A database of 1349 RNA secondary structures containing 123 small subunit rRNAs (Gutell 1994), 223 large subunit rRNAs (Gutell et al. 1993; Schnare et al. 1996), 309 5S rRNAs (Szymanski et al. 1998), 484 tRNAs (Sprinzl et al. 1998), 91 signal recognition particles (Larsen et al. 1998), 16 RNase P RNAs (Brown 1998), 100 group I introns (Waring and Davies 1984; Damberger and Gutell 1994), and three group II introns (Michel et al. 1989) was compiled. This database was searched for tetraloops, and the number of occurrences for each type of tetraloop was tabulated. In this work, G–U pairs are considered to be canonical base pairs.

Design of sequences for optical melting studies

Sequences of tetraloops and closing base pairs were designed to represent those found most frequently in the database described above. Each stem contained two 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. The duplexes were also designed to have a melting temperature between 40 and 75°C. Care was taken to design the stem–loop sequences so that the tetraloop of interest would form. To ensure that unimolecular tetraloop formation outcompeted bimolecular association between strands, the following equations, derived from the equilibrium equations and $\Delta G = -RT \ln K$, were utilized:

$$[H] = \frac{-1 + \sqrt{(8 * K_D * [A]_T)/(K_H * K_H)}}{(4 * K_D)(K_H * K_H)}, \quad (6)$$

$$[D] = ([A]_T - [H])/2, \quad (7)$$

$$\%H = [H]/([H] + [D]) * 100. \quad (8)$$

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 of hairpin in solution. K_H and K_D values were calculated at 37°C using ΔG°_{37} values from RNAstructure (Mathews et al. 1999, 2004; Lu et al. 2006) for hairpin and duplex formation, respectively. Calculations were done for $[A]_T$ = 1 μM and 0.1 mM, the usual concentration range for the melting experiments. All of the sequences studied here had %[H] > 79% at $[A]_T$ = 0.1 mM and %[H] > 99% at $[A]_T$ = 1 μM.

The current model used to predict the stability of RNA tetraloops closed by Watson–Crick pairs (Vecenie and Serra 2004; Vecenie et al. 2006) depends upon terminal mismatch values. There were three terminal mismatch values not previously measured; therefore, these terminal mismatches were analyzed. In addition, several terminal mismatch values were determined from duplexes with low melting temperatures, which resulted in difficulty during data analysis (Vecenie and Serra 2004). Therefore, these terminal mismatches were remeasured in more stable duplexes.

RNA synthesis and purification

Oligonucleotides were ordered from Integrated DNA Technologies. The synthesis and purification of the oligonucleotides followed standard procedures that were described previously (Davis and Znosko 2007; Wright et al. 2007; Christiansen and Znosko 2008).

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). Melting curves (absorbance versus temperature) were obtained, and duplex thermodynamics (for terminal mismatches) were determined as described previously (Davis and Znosko 2007; Wright et al. 2007; Christiansen and Znosko 2008). 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 each individual curve fit. The thermodynamic contributions of tetraloops to stem–loop thermodynamics ($\Delta G^\circ_{37,tetraloop}$, $\Delta H^\circ_{tetraloop}$, and $\Delta S^\circ_{tetraloop}$) and terminal mismatches to duplex thermodynamics were determined by subtracting the Watson–Crick contribution (Xia et al. 1998) from the measured thermodynamics.

Linear regression and tetraloop thermodynamic parameters

Data collected for the 24 tetraloops in this study were combined with previously published data for 28 tetraloops (Groebe and Uhlenbeck 1988; Antao and Tinoco 1992; Serra et al. 1997; Giese et al. 1998; Dale et al. 2000) that were also melted in 1 M NaCl. Updated values for parameters in previous predictive models and values for new predictive models were derived by linear regression as described previously (Badhwar et al. 2007; Davis and Znosko 2007; Wright et al. 2007; Christiansen and Znosko 2008; Davis and Znosko 2008). The calculated experimental contribution of the tetraloop to stem–loop stability was used as a constant when doing linear regression. To simultaneously solve for each variable, the LINEST function of Microsoft Excel was used for linear regression.

ACKNOWLEDGMENTS

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.

Received June 12, 2009; accepted November 4, 2009.

REFERENCES

Allain FH, Varani G. 1995. Structure of the P1 helix from group I self-splicing introns. *J Mol Biol* **250:** 333–353.

Antao VP, Tinoco I. 1992. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res* **20:** 819–824.

Antao VP, Lai SY, Tinoco I. 1991. A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res* **19:** 5901–5905.

Badhwar J, Karri S, Cass CK, Wunderlich EL, Znosko BM. 2007. Thermodynamic characterization of RNA duplexes containing naturally occurring 1×2 nucleotide internal loops. *Biochemistry* **46:** 14715–14724.

Blose JM, Proctor DJ, Veeraraghavan N, Misra VK, Bevilacqua PC. 2009. Contribution of the closing base pair to exceptional stability in RNA tetraloops: Roles for molecular mimicry and electrostatic factors. *J Am Chem Soc* **131:** 8474–8484.

Brown JW. 1998. The ribonuclease P database. *Nucleic Acids Res* **26:** 351–352.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273:** 1678–1685.

Cheong C, Varani G, Tinoco I. 1990. Solution structure of an unusually stable RNA hairpin, 5′-GGAC(UUCG)GUCC. *Nature* **346:** 680–682.

Christiansen ME, Znosko BM. 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.

Dale T, Smith R, Serra MJ. 2000. A test of the model to predict unusually stable RNA hairpin loop stability. *RNA* **6:** 608–615.

Damberger SH, Gutell RR. 1994. A comparative database of group I intron structures. *Nucleic Acids Res* **22:** 3508–3510.

Davis AR, Znosko BM. 2007. Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry* **46:** 13425–13436.

Davis AR, Znosko BM. 2008. Thermodynamic characterization of naturally occurring RNA single mismatches with G–U nearest neighbors. *Biochemistry* **47:** 10178–10187.

Du Z, Yu J, Andino R, James TL. 2003. Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from coxsackie virus B3. *Biochemistry* **42:** 4373–4383.

Ennifar E, Nikulin A, Tishchenko S, Serganov A, Nevskaya N, Garber M, Ehresmann B, Ehresmann C, Nikonov S, Dumas P. 2000. The crystal structure of UUCG tetraloop. *J Mol Biol* **304:** 35–42.

Fourmy D, Guittet E, Yoshizawa S. 2002. Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB. *J Mol Biol* **324:** 137–150.

Freier SM, Kierzek R, Caruthers MH, Neilson T, Turner DH. 1986. Free energy contributions of G•U and other terminal mismatches to helix stability. *Biochemistry* **25:** 3209–3213.

Giese MR, Betschart K, Dale T, Riley CK, Rowan C, Sprouse KJ, Serra MJ. 1998. Stability of RNA hairpins closed by wobble base pairs. *Biochemistry* **37:** 1094–1100.

Girard FC, Ottink OM, Ampt KA, Tessari M, Wijmenga SS. 2007. Thermodynamics and NMR studies on duck, heron, and human HBV encapsidation signals. *Nucleic Acids Res* **35:** 2800–2811.

Groebe DR, Uhlenbeck OC. 1988. Characterization of RNA hairpin–loop stability. *Nucleic Acids Res* **16:** 11725–11735.

Gutell RR. 1994. Collection of small-subunit (16s- and 16s-like) ribosomal-RNA structures—1994. *Nucleic Acids Res* **22:** 3502–3507.

Gutell RR, Gray MW, Schnare MN. 1993. A compilation of large subunit (23s-like and 23s-like) ribosomal-RNA structures—1993. *Nucleic Acids Res* **21:** 3055–3074.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253:** 191–194.

Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res* **31:** 3429–3431.

Jucker FM, Pardi A. 1995. Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif. *Biochemistry* **34:** 14416–14427.

Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264:** 968–980.

Kieft JS, Tinoco I. 1997. Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexamine. *Structure* **5:** 713–721.

Kierzek R, Burkard ME, Turner DH. 1999. Thermodynamics of single mismatches in RNA duplexes. *Biochemistry* **38:** 14214–14223.

Laing LG, Hall KB. 1996. A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochemistry* **35:** 13586–13596.

Larsen N, Samuelsson T, Zwieb C. 1998. The signal recognition particle database (SRPDB). *Nucleic Acids Res* **26:** 177–178.

Longfellow CE, Kierzek R, Turner DH. 1990. Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry* **29:** 278–285.

Lu ZJ, Turner DH, Mathews DH. 2006. A set of nearest-neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Res* **34:** 4912–4924.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288:** 911–940.

Mathews DH, Disney MD, Childs JC, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci* **101:** 7287–7292.

Michel F, Umesono K, Ozeki H. 1989. Comparative and functional anatomy of group-Ii catalytic introns—a review. *Gene* **82:** 5–30.

Proctor DJ, Schaak JE, Bevilacqua JM, Falzone CJ, Bevilacqua PC. 2002. Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry* **41:** 12062–12075.

SantaLucia J Jr, Kierzek R, Turner DH. 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.

Schnare MN, Damberger SH, Gray MW, Gutell RR. 1996. Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA. *J Mol Biol* **256:** 701–719.

Serra MJ, Turner DH. 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol* **259:** 242–261.

Serra MJ, Axenson TJ, Turner DH. 1994. A model for the stabilities of RNA hairpins based on a study of the sequence dependence of stability for hairpins of six nucleotides. *Biochemistry* **33:** 14289–14296.

Serra MJ, Barnes TW, Betschart K, Gutierrez MJ, Sprouse KJ, Riley CK, Stewart L, Temel RE. 1997. Improved parameters for the prediction of RNA hairpin stability. *Biochemistry* **36:** 4844–4851.

Siegfried NA, Metzger SL, Bevilacqua PC. 2007. Folding cooperativity in RNA and DNA is dependent on position in the helix. *Biochemistry* **46:** 172–181.

Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26:** 148–153.

Szymanski M, Specht T, Barciszewska MZ, Barciszewski J, Erdmann VA. 1998. 5S rRNA data bank. *Nucleic Acids Res* **26:** 156–159.

Tinoco I, Bustamante C. 1999. How RNA folds. *J Mol Biol* **293:** 271–281.

Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, Stormo G, d'Aubenton-Carafa Y, Uhlenbeck OC, Tinoco I, et al. 1988. CUUCGG hairpins: Extraordinary stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci* **85:** 1364–1368.

Varani G. 1995. Exceptionally stable nucleic acid hairpins. *Annu Rev Biophys Biomol Struct* **24:** 379–404.

Varani G, Cheong C, Tinoco I Jr. 1991. Structure of an unusually stable RNA hairpin. *Biochemistry* **30:** 3280–3289.

Vecenie CJ, Serra MJ. 2004. Stability of RNA hairpin loops closed by AU base pairs. *Biochemistry* **43:** 11813–11817.

Vecenie CJ, Morrow CV, Zyra A, Serra MJ. 2006. Sequence dependence of the stability of RNA hairpin molecules with six nucleotide loops. *Biochemistry* **45:** 1400–1407.

Waring RB, Davies RW. 1984. Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing—a review. *Gene* **28:** 277–291.

Williams DJ, Hall KB. 2000. Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J Mol Biol* **297:** 1045–1061.

Wimberly BT, Brodersen DE, Clemons WM, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407:** 327–339.

Wolters J. 1992. The nature of preferred hairpin structures in 16S-like rRNA variable regions. *Nucleic Acids Res* **20:** 1843–1850.

Wright DJ, Rice JL, Yanker DM, Znosko BM. 2007. Nearest-neighbor parameters for inosine–uridine pairs in RNA duplexes. *Biochemistry* **46:** 4625–4634.

Wu H, Yang PK, Butcher SE, Kang S, Chanfreau G, Feigon J. 2001. A novel family of RNA tetraloops structure forms the recognition site for *Saccharomyces cerevisiae* RNase III. *EMBO J* **20:** 7240–7249.

Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry* **37:** 14719–14735.

Zuker M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science* **244:** 48–52.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406–3415.

Thermodynamic characterization of naturally occurring RNA tetraloops

Justin P. Sheehy, Amber R. Davis and Brent M. Znosko

RNA 2010 16: 417-429 originally published online January 4, 2010
Access the most recent version at doi:10.1261/rna.1773110