BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

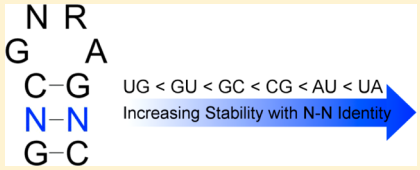
pubs.acs.org/biochemistry

Effects of Non-Nearest Neighbors on the Thermodynamic Stability of RNA GNRA Hairpin Tetraloops

Pamela L. Vanegas, Teresa S. Horwitz, and Brent M. Znosko*

Department of Chemistry, Saint Louis University, 3501 Laclede Avenue, Saint Louis, Missouri 63103, United States

ABSTRACT: Currently, several models for predicting the secondary structure of RNA exist, one of which is free energy minimization using the Nearest Neighbor Model. This model predicts the lowest-free energy secondary structure from a primary sequence by summing the free energy contributions of the Watson−Crick nearest neighbor base pair combinations and any noncanonical secondary structure motif. The Nearest Neighbor Model also assumes that the free energy of the secondary structure motif is dependent solely on the identities of the nucleotides within the motif and the motif's nearest neighbors. To test the current assumption of the Nearest Neighbor Model that the non-nearest neighbors do not affect the stability of the motif, we optically melted different stem−loop oligonucleotides to experimentally determine their thermodynamic parameters. In each of these oligonucleotides, the hairpin loop sequence and the adjacent base pairs were held constant, while the first or second non-nearest neighbors were varied. The experimental results show that the thermodynamic contributions of the hairpin loop were dependent upon the identity of the first non-nearest neighbor, while the second non-nearest neighbor had a less obvious effect. These results were then used to create an updated model for predicting the thermodynamic contributions of a hairpin loop to the overall stability of the stem−loop structure.



Most biologically relevant ribonucleic acids (RNAs) are single strands that fold back upon themselves to create specific secondary and tertiary structures. Even though the Protein Data Bank (PDB)[1] is constantly expanding its collection of three-dimensional structures of small, noncoding RNAs, the vast majority of these small RNAs still have an unknown tertiary structure. One possible step in determining the three-dimensional structure might be through secondary structure prediction, because the secondary structure forms faster and more strongly than the tertiary structure and helps dictate the three-dimensional structure of the macromolecule.[2−5]

A common method of predicting secondary structure of RNA is free energy minimization using the Nearest Neighbor Model. This model can be used to calculate the stability of a completely paired duplex by summing the free energy contributions of pairs of neighboring nucleotides.[6] In addition to completely paired duplexes, the Nearest Neighbor Model is also often used to predict the free energies of RNA duplexes containing secondary structure motifs, such as hairpins, internal bulges, and internal mismatches.[7,8] This is done by summing the free energy contributions of the Watson−Crick stem surrounding the motif and then adding a contribution from the motif itself.[4,7,9] However, there are several limitations to the Nearest Neighbor Model,[4,8,9] including the assumption that the free energy of a secondary structure is solely based on nearest neighbor interactions without any contribution from first or second non-nearest neighbor interactions.

Previous studies[2,3,8,10−17] have observed that non-nearest neighbors do affect the stability of secondary structural motifs. Sheehy et al.[10] observed significant non-nearest neighbor effects in tetraloops; tetraloops with a $^{5'}$GCC$^{3'}$/$^{3'}$CGG$^{5'}$ stem sequence were on average 0.6 kcal/mol more stable than the same tetraloops with a $^{5'}$GGC$^{3'}$/$^{3'}$CCG$^{5'}$ stem sequence. Results of Davis et al.[14] show that the thermodynamic contributions of a single mismatch are dependent on the mismatch's position within the helix. Although this is a significant finding, there were two variables that differed in this study: the position of the mismatch in the helix (centered, 5′-shifted, or 3′-shifted) and the non-nearest neighbors of the mismatch. Therefore, the contributions of only the non-nearest neighbors to the thermodynamics of the mismatch could not be determined.[14] Blose et al.[2] and McCann et al.[3] have studied the effects of non-nearest neighbors on single-nucleotide bulges; both studies noted that there was a direct correlation between the stability of the stem and the energetic cost of inserting a bulged nucleotide. For group I single-nucleotide bulges, as the free energy of the stem decreased, so did the level of destabilization of the duplex from a bulge loop of one nucleotide. Similarly, for group II single-nucleotide bulges, the energetic contribution and identity of the bulged nucleotide were dependent upon the free energy of the second least stable possible stem.[2,3] It is possible that as the stem becomes less stable, the structural perturbations from the addition of a bulge loop can be more easily dissipated through the backbone, reducing the free energy loss.[2] Though these studies demonstrate the importance of non-nearest neighbors for the stability of a secondary structure motif, the extent of such interactions was not investigated. None of the studies mentioned previously have systematically evaluated the effects of all six possible non-nearest neighbor combinations on the thermodynamic stability of a structural motif, which is the focus of this study.

Received: January 3, 2012
Revised: February 10, 2012
Published: February 13, 2012

In this study, the energetic dependence of GNRA tetraloops on non-nearest neighbors was tested. All six possible non-nearest neighbor combinations, including the G·U wobble, were studied in six different loop-closing base pair combinations. It was found that there is a significant contribution to the free energy of the tetraloop by the first non-nearest neighbor, while there was a smaller contribution by the second non-nearest neighbor. The thermodynamic data were then used to update the predictive model[10] for tetraloops by adding both a closing base pair term and a conditional non-nearest neighbor term, improving the accuracy and precision of the model for all tetraloops.

MATERIALS AND METHODS

Loop Selection. In a previous study,[10] tetraloops were ranked on the basis of their frequency of occurrence in a database of secondary structures. The loops selected for this study are the three most common tetraloops, $^{5'}$GAAA$^{3'}$, $^{5'}$GUGA$^{3'}$ (Figure 1), and $^{5'}$GCAA$^{3'}$, with natural abundances

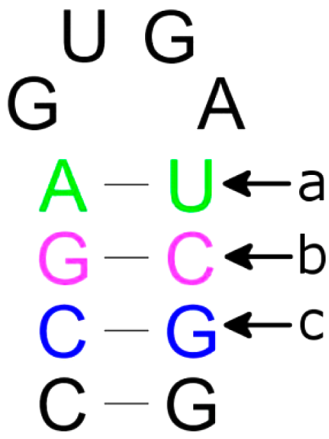


Figure 1. $^{5'}$GUGA$^{3'}$ tetraloop in which the closing base pair or nearest neighbor ($^{5'}$A-U$^{3'}$) is colored green (a), the first non-nearest neighbor ($^{5'}$G-C$^{3'}$) is colored magenta (b), and the second non-nearest neighbor ($^{5'}$C-G$^{3'}$) is colored blue (c).

of 16.1, 8.0, and 7.1%, respectively.[10] Each specific tetraloop sequence was then evaluated with both a $^{5'}$C-G$^{3'}$ closing base pair and a $^{5'}$A-U$^{3'}$ closing base pair. The $^{5'}$C-G$^{3'}$ closing base pair was chosen as it is the most common closing base pair for tetraloops.[10] Additionally, the $^{5'}$C-G$^{3'}$ closing base pair reduces the likelihood of G-quadruplex formation as it eliminates any string of guanines over two residues in length by placing a cytosine beside the guanine of the loop. The $^{5'}$A-U$^{3'}$ closing base pair was chosen to be a comparison between closing base pairs that contain two hydrogen bonds and those that contain three hydrogen bonds. The non-nearest neighbors were systematically altered within the stem to give all six possible combinations, including $^{5'}$G·U$^{3'}$ and $^{5'}$U·G$^{3'}$, and the $^{5'}$G-C$^{3'}$ terminal base pair was chosen to prevent fraying of the hairpin during optical melting studies. For the purposes of this study, the G·U wobble pair is considered to be a canonical base pair. The theoretical concentrations of the hairpin and duplex were calculated using eq 1 as previously described[18]

$$[\mathrm{H}] = \frac{-1 + \sqrt{(8K_\mathrm{D}[\mathrm{A}]_\mathrm{T})/K_\mathrm{H}^2 + 1}}{(4K_\mathrm{D})/K_\mathrm{H}^2} \quad (1)$$

where $[\mathrm{H}]$ is the concentration of hairpins, K_D is the equilibrium constant for duplex formation, K_H is the equilibrium constant for hairpin formation, and $[\mathrm{A}]_\mathrm{T}$ is the total strand concentration. All hairpin concentrations were calculated to be ≥97% of the total concentration.

Synthesis and Purification. Oligoribonucleotides were ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The synthesis and purification of the oligomers followed standard procedures previously described.[13]

Optical Melting Studies and Thermodynamics. All optical melting experiments were performed in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM EDTA (pH 7.0). Each of the stem−loop oligonucleotides was melted at least nine times, with a concentration range of approximately 50-fold. Data points for melting curves were collected at a rate of 1 °C/min from 10 to 90 °C on a Beckman-Coulter DU800 spectrometer with a Beckman-Coulter high-performance temperature controller. There was no observed concentration dependence of the melting temperature (T_M) for any of the oligonucleotides studied here, verifying the formation of the unimolecular stem−loop structure as opposed to the bimolecular duplex structure. The thermodynamic parameters of each stem−loop structure were determined using *MeltWin* version 3.5[19] as previously described,[10,12,18,20] and the thermodynamic contributions of each tetraloop were calculated by subtracting the nearest neighbor values for each canonical base pair in the stem from the measured thermodynamic values.[6,7,11,20]

Linear Regression and the New Predictive Model. After the thermodynamic contributions of each loop were characterized, this data set was added to other tetraloops found in the literature,[10,21−25] resulting in a total of 101 tetraloops whose thermodynamic parameters have been experimentally determined. One published tetraloop, $^{5'}$GGGAUACCCCC-GUAUCCA$^{3'}$,[22] where the underlined nucleotides constitute the loop, was excluded because polycytosine hairpin loops are known to be unusually unstable.[7] This data set was evaluated using the LINEST function in *Microsoft Excel* as described by Wright et al.[13] Multiple parameters and combinations of parameters were evaluated (data not shown), and the parameters that were chosen produced the most accurate predictive model.

RESULTS

Thermodynamic Parameters for First and Second Non-Nearest Neighbors. Table 1 shows the thermodynamic parameters for GNRA tetraloops in which the first non-nearest neighbor was varied, while Table 2 shows the parameters for the tetraloops in which the second non-nearest neighbor was varied. Values for ΔG°_{37}, ΔH°, and ΔS° were determined via fitting the melting curves to the two-state model. All oligonucleotides melted in accordance with the two-state model and showed no correlation between the concentration and the melting temperature.

Contributions of Tetraloops to Stem−Loop Free Energy. Each loop's contribution to the overall free energy was determined by subtracting the nearest neighbor values for the stem nucleotides from the overall experimental values. The last columns of both Tables 1 and 2 show the free energy contribution of the loop. The thermodynamic contributions from the tetraloops with a different first non-nearest neighbor can vary by up to 1.51 kcal/mol, as observed in the $^{5'}$CGCAAG$^{3'}$ tetraloop sequence. With a different second non-nearest neighbor, the variation can reach 1.11 kcal/mol,

dx.doi.org/10.1021/bi300008j | *Biochemistry* 2012, 51, 2192−2198

Table 1. Thermodynamic Parameters and Free Energy Contributions of Tetraloops with Varying First Non-Nearest Neighbors[a]

Loop-Closing Base Pair Sequence[b]	Sequence[b]	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°37 (kcal/mol)	Tm (°C)	ΔG°37,tetraloop (kcal/mol)[c]
CGAAAG	GGCGAAAGCC	-33.9 ± 1.6	-98.9 ± 5.2	-3.23 ± 0.16	69.7	3.45
		-35.6 ± 1.5	-104.4 ± 5.2	-3.20 ± 0.16	67.6	3.48
		-33.4 ± 1.7[d]	-97.4 ± 4.0[d]	-3.20 ± 0.20[d]	70.3[d]	3.48[d]
	GCCGAAAGGC	-35.3 ± 1.8	-101.2 ± 5.8	-3.92 ± 0.06	75.7	2.76
		-35.4 ± 2.2[e]	-101.7 ± 6.4[e]	-3.85 ± 0.18[e]	74.8[e]	2.83[e]
	GGCGAAAGUC	-31.5 ± 2.0	-99.1 ± 6.3	-0.81 ± 0.09	45.2	3.23
		-31.9 ± 1.4	-100.2 ± 4.7	-0.80 ± 0.10	44.9	3.24
	GUCGAAAGGC	-30.8 ± 1.0	-97.4 ± 3.4	-0.56 ± 0.08	42.8	3.48
		-30.7 ± 2.1	-97.4 ± 6.7	-0.51 ± 0.05	42.2	3.53
	GACGAAAGUC	-34.5 ± 2.5	-105.4 ± 7.9	-1.78 ± 0.13	53.8	2.81
		-32.9 ± 1.4	-100.3 ± 4.5	-1.83 ± 0.06	55.2	2.76
	GUCGAAAGAC	-35.2 ± 2.2	-107.4 ± 6.9	-1.90 ± 0.10	54.7	2.69
		-32.5 ± 2.1	-98.7 ± 6.9	-1.84 ± 0.14	55.7	2.75
AGAAAU	CGGAGAAAUCCG	-42.8 ± 3.1	-126.2 ± 9.7	-3.70 ± 0.14	66.4	3.82
	CGCAGAAAUGCG	-40.4 ± 1.4	-118.7 ± 4.2	-3.59 ± 0.08	67.3	3.85
	CGGAGAAAUUCG	-36.9 ± 3.7	-116.7 ± 11.9	-0.69 ± 0.02	42.9	4.02
	CGUAGAAAUGCG	-36.9 ± 1.6	-116.2 ± 5.0	-0.89 ± 0.01	44.7	4.53
	CGAAGAAAUUCG	-31.7 ± 4.7	-97.3 ± 15.2	-1.50 ± 0.02	52.4	3.69
	CGUAGAAAUACG	-30.1 ± 4.4	-91.9 ± 14.0	-1.56 ± 0.07	53.9	3.92
CGCAAG	GGCGCAAGCC	-38.5 ± 3.1	-112.5 ± 10.0	-3.66 ± 0.07	69.6	3.02
		-34.3 ± 2.5[d]	-99.8 ± 7.6[d]	-3.40 ± 0.30[d]	71.0[d]	3.28[d]
	GCCGCAAGGC	-37.2 ± 3.0	-106.9 ± 9.6	-4.01 ± 0.06	74.5	2.67
		-33.6 ± 3.4[e]	-96.7 ± 10.1[e]	-3.60 ± 0.27[e]	74.3[e]	3.08[e]
		-37.1 ± 1.4[e]	-107.2 ± 4.6[e]	-3.80 ± 0.11[e]	72.4[e]	2.88[e]
	GGCGCAAGUC	-31.7 ± 2.6	-96.1 ± 8.4	-1.90 ± 0.08	56.8	2.14
	GUCGCAAGGC	-27.4 ± 6.3	-87.1 ± 20.2	-0.39 ± 0.02	41.4	3.65
	GACGCAAGUC	-24.3 ± 5.1	-72.5 ± 16.5	-1.76 ± 0.02	61.2	2.83
	GUCGCAAGAC	-36.2 ± 4.8	-109.5 ± 15.5	-2.23 ± 0.07	57.3	2.36
AGCAAU	CGGAGCAAUCCG	-43.1 ± 4.9	-127.3 ± 15.9	-3.62 ± 0.08	65.4	3.90
	CGCAGCAAUGCG	-34.4 ± 3.8	-99.7 ± 12.5	-3.46 ± 0.06	71.7	3.98
	CGGAGCAAUUCG	-35.0 ± 4.3	-110.6 ± 13.9	-0.70 ± 0.01	43.3	4.01
	CGUAGCAAUGCG	-35.1 ± 4.7	-109.8 ± 15.0	-1.02 ± 0.09	46.3	4.40
	CGAAGCAAUUCG	-23.4 ± 3.8	-69.7 ± 12.1	-1.81 ± 0.06	63.0	3.38
	CGUAGCAAUACG	-37.0 ± 4.2	-112.8 ± 13.5	-2.03 ± 0.04	55.0	3.45
CGUGAG	GGCGUGAGCC	-33.2 ± 3.9	-97.2 ± 12.7	-3.04 ± 0.05	68.3	3.64
		-35.6 ± 2.9[d]	-104.6 ± 8.7[d]	-3.20 ± 0.30[d]	67.8[d]	3.48[d]
	GCCGUGAGGC	-35.1 ± 3.2	-101.5 ± 10.3	-3.60 ± 0.06	72.4	3.08
		-38.3 ± 1.8[e]	-111.5 ± 5.5[e]	-3.73 ± 0.16[e]	70.5[e]	2.95[e]
	GGCGUGAGUC	-33.7 ± 2.9	-106.6 ± 9.5	-0.67 ± 0.07	43.3	3.37
	GUCGUGAGGC	-35.1 ± 7.9	-112.0 ± 25.4	-0.33 ± 0.04	39.9	3.71
	GACGUGAGUC	-33.1 ± 5.4	-101.5 ± 17.4	-1.63 ± 0.09	53.0	2.96
	GUCGUGAGAC	-34.0 ± 2.4	-104.6 ± 7.8	-1.54 ± 0.09	51.8	3.05
AGUGAU	CGGAGUGAUCCG	-38.6 ± 3.4	-114.2 ± 11.1	-3.13 ± 0.07	64.4	4.39
	CGCAGUGAUGCG	-41.6 ± 4.9	-122.8 ± 15.6	-3.49 ± 0.04	65.4	3.95
	CGGAGUGAUUCG	-33.0 ± 7.3	-104.6 ± 23.5	-0.57 ± 0.08	42.5	4.14
	CGUAGUGAUGCG	-36.2 ± 1.9	-114.4 ± 6.1	-0.69 ± 0.05	43.0	4.73
	CGAAGUGAUUCG	-34.1 ± 4.2	-105.6 ± 13.4	-1.30 ± 0.05	49.3	3.89
	CGUAGUGAUACG	-38.5 ± 5.8	-118.7 ± 18.8	-1.72 ± 0.07	51.5	3.76

[a]Measurements were taken in 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA (pH 7.0). [b]Sequences are written 5′ → 3′. Loop nucleotides are underlined, and first non-nearest neighbors are colored magenta. [c]The free energy contribution of the tetraloop was calculated by subtracting the Watson−Crick contribution of the stem[6] from the experimental free energy of the stem−loop structure. [d]From ref 21. [e]From ref 10.

as seen with the $^{5'}$AGCAAU$^{3'}$ sequence. All calculated values for the thermodynamic contribution of the tetraloops are in good agreement with previously published results,[10,21] showing good reproducibility of experimental thermodynamic parameters.

Updated Model for Prediction of Thermodynamics of Tetraloops. Previous models[10,26] for predicting tetraloop stability contain parameters that are relatively conserved among the models. In all models, base pairings for the closing base pair and for the non-nearest neighbor are direction-dependent; e.g., a $^{5'}$G-C$^{3'}$ base pair is not the same as a $^{5'}$C-G$^{3'}$ base pair, and a $^{5'}$G·A$^{3'}$ mismatch is not the same as a $^{5'}$A·G$^{3'}$ mismatch. The most recent model[10] for the prediction of tetraloop contributions to overall stability is shown in eq 2.

$$\Delta G^\circ_{37,\text{tetraloop}} = \Delta G^\circ_{37,\text{initiation}} + \Delta G^\circ_{37,\text{first mismatch bonus}} + \Delta G^\circ_{37,\text{non-nearest neighbor}} \quad (2)$$

The predictive model of Sheehy et al. (eq 2) has an average difference of 0.4 kcal/mol between the predicted free energy contribution and the experimentally determined free energy contribution of the loop.[10] However, the non-nearest neighbor parameter that was used in this previous predictive model is specific to the particular three-base pair $^{5'}$GCC$^{3'}$/$^{3'}$CGG$^{5'}$ stem, limiting the utility of this model. To improve the model of Sheehy et al., many combinations of parameters were tested, and the most accurate and precise model (eq 3) resulted in dividing the non-nearest neighbor parameter into two separate parameters, a bonus for a $^{5'}$CGNRAG$^{3'}$ or a $^{5'}$CUNCGG$^{3'}$ tetraloop and a non-nearest neighbor parameter for those two specific tetraloop sequences.

$$\Delta G^\circ_{37,\text{tetraloop}} = \Delta G^\circ_{37,\text{initiation}} + \Delta G^\circ_{37,\text{first mismatch bonus}} + \Delta G^\circ_{37,\text{CGNRAG or CUNCGG}} + \Delta G^\circ_{37,\text{CGNRAG or CUNCGG non-nearest neighbor}} \quad (3)$$

In this equation, there are four distinct parameters. $\Delta G^\circ_{37,\text{initiation}}$ and $\Delta G^\circ_{37,\text{first mismatch bonus}}$ are based upon the parameters from the previous model,[10] with adjustments to the numerical values. $\Delta G^\circ_{37,\text{initiation}}$ is now 4.4 kcal/mol (decreased from 4.8 kcal/mol), and $\Delta G^\circ_{37,\text{first mismatch}}$ is now −0.5 kcal/mol for a $^{5'}$G·A$^{3'}$ or $^{5'}$U·U$^{3'}$ mismatch (increased from −1.1 kcal/mol) or −1.0 kcal/mol for a $^{5'}$U·G$^{3'}$ mismatch (increased from −1.7 kcal/mol). The other two parameters, $\Delta G^\circ_{37,\text{CGNRAG or CUNCGG}}$ and $\Delta G^\circ_{37,\text{CGNRAG or CUNCGG non-nearest neighbor}}$, are unique to this proposed model. They are derived from the non-nearest neighbor parameter in the Sheehy model[10] but are updated to be more universal. The first new parameter, $\Delta G^\circ_{37,\text{CGNRAG or CUNCGG}}$, is applicable only if the tetraloop and closing base pair sequence is $^{5'}$CGNRAG$^{3'}$ or $^{5'}$CUNCGG$^{3'}$. If this sequence requirement is met, then a bonus of −0.7 kcal/mol is applied. If the tetraloop and closing base pair are any other sequence combination, this parameter is zero. The fourth parameter, $\Delta G^\circ_{37,\text{CGNRAG or CUNCGG non-nearest neighbor}}$, is also only applied if the tetraloop and closing base pair sequence is $^{5'}$CGNRAG$^{3'}$ or $^{5'}$CUNCGG$^{3'}$, and the value of the bonus or penalty is dependent on the non-nearest neighbors to the tetraloop. A bonus of −0.4 kcal/mol is applied for tetraloops with the sequence $^{5'}$CCGNRAGG$^{3'}$ or

dx.doi.org/10.1021/bi300008j | Biochemistry 2012, 51, 2192−2198

Table 2. Thermodynamic Parameters and Free Energy Contributions of Tetraloops with Varying Second Non-Nearest Neighbors[a]

Tetraloop Sequence[b]	Sequence[b]	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°$_{37}$ (kcal/mol)	T_m (°C)	ΔG°$_{37,tetraloop}$ (kcal/mol)[c]
GAGCAAUC	CGGAGCAAUCCG	-43.1 ± 4.9	-127.3 ± 15.9	-3.62 ± 0.08	65.4	3.90
	CCGAGCAAUCGG	-47.6 ± 6.6	139.9 ± 21.4	-4.20 ± 0.04	67.0	3.32
	CGGAGCAAUCUG	-40.9 ± 6.3	-127.0 ± 20.3	-1.55 ± 0.04	49.2	3.87
	CUGAGCAAUCGG	-33.4 ± 11.9	-104.4 ± 38.5	-0.99 ± 0.06	46.5	4.43
	CAGAGCAAUCUG	-36.1 ± 4.5	-110.1 ± 14.6	-1.97 ± 0.03	54.8	4.12
	CUGAGCAAUCAG	-32.3 ± 3.5	-110.6 ± 11.4	-1.96 ± 0.05	54.7	4.13
GAGUGAUC	CGGAGUGAUCCG	-38.6 ± 3.4	-114.2 ± 11.1	-3.13 ± 0.07	64.4	4.39
	CCGAGUGAUCGG	-38.9 ± 4.0	-115.7 ± 12.8	-3.03 ± 0.04	63.2	4.49
	CGGAGUGAUCUG	-33.3 ± 3.0	-104.0 ± 9.8	-1.03 ± 0.01	46.9	4.39
	CUGAGUGAUCGG	-31.0 ± 3.0	-98.0 ± 9.5	-0.63 ± 0.02	43.4	4.55
	CAGAGUGAUCUG	-37.7 ± 3.1	-115.9 ± 10.2	-1.75 ± 0.02	52.1	4.34
	CUGAGUGAUCAG	-35.7 ± 4.1	-109.8 ± 13.3	-1.64 ± 0.06	51.9	4.45

[a]Measurements were taken in 1.0 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA (pH 7.0). [b]Sequences are written 5′ → 3′. Loop nucleotides are underlined, and second non-nearest neighbors are colored blue. [c]The free energy contribution of the tetraloop was calculated by subtracting the Watson−Crick contribution of the stem[6] from the experimental free energy of the stem−loop.

Table 3. Predicted Free Energies of Tetraloops (kilocalories per mole) Using the Proposed Model

		non-nearest neighbor identity[b]						
loop sequence[a]	closing base pair identity[b]	any	AU	UA	CG	GC	GU	UG
NNNN	any	4.4						
GNYA	any	3.9						
UNNU	any	3.9						
UNDG	any	3.0						
GNRA	CG		3.9	3.4	3.5	3.9	3.9	4.3
	any other base pair	3.9						
UNCG	CG		3.9	3.4	3.5	3.9	3.9	4.3
	any other base pair	3.9						
CCCC	Any	7.2						

[a]Loop sequences are written 5′ → 3′. N represents any nucleotide, Y any pyrimidine, R any purine, and D any nucleotide except for cytosine. [b]The identities for the closing base pair and the non-nearest neighbor must be in the 5′ → 3′ direction; i.e., a C-G closing base pair represents a C on the 5′ side of the loop and a G on the 3′ side.

Table 4. Free Energies of GNRA Tetraloops and Ranking of the First Non-Nearest Neighbors[a]

	ΔG°$_{tetraloop}$ (kcal/mol) and ranking[a]												
loop sequence with closing base pairs[b]	G-C		C-G		G·U		U·G		A-U		U-A		range (kcal/mol)
CGAAAG	3.47[c]	5	2.76	2	3.24[c]	4	3.51[c]	6	2.79[c]	3	2.72[c]	1	0.79
CGCAAG	3.02	5	2.67	3	2.14	1	3.65	6	2.83	4	2.36	2	1.51
CGUGAG	3.64	5	3.08	3	3.37	4	3.71	6	2.96	1	3.05	2	0.75
AGAAAU	3.82	2	3.85	3	4.02	5	4.53	6	3.69	1	3.92	4	0.84
AGCAAU	3.90	3	3.98	4	4.01	5	4.40	6	3.38	1	3.45	2	1.02
AGUGAU	4.39	5	3.95	3	4.14	4	4.73	6	3.89	2	3.76	1	0.97

[a]The listed numbers are ΔG° tetraloop values and rankings for the loop sequence with closing base pair when adjacent to the given first non-nearest neighbors. Rankings are based on the calculated free energy of the RNA tetraloops, with 1 being the least destabilizing and 6 the most destabilizing to the tetraloop. ΔG° values were calculated from the predicted free energy of the stem[6] and the experimental free energy of the stem−loop. Rankings 1−6 are based upon the closing base pair and loop sequence; i.e., in the first row XCGAAAGY is ranked from 1 to 6. [b]Underlined nucleotides constitute the tetraloop sequence. [c]ΔG° values are averages from two separate runs of these sequences. Both were within experimental error.

5′CCUNCGGG3′. A bonus of −0.5 kcal/mol is applied for tetraloops with the sequence 5′UCGNRAGA3′ or 5′UCUNCGGA3′. A penalty of 0.4 kcal/mol is applied for tetraloops with the sequence 5′UCGNRAGG3′ or 5′UCUNCGGG3′. For all other sequences, this parameter is zero. Even though it appears as if the second non-nearest neighbor contributes to the stability of the tetraloop, the updated model does not include a parameter for second non-nearest neighbors. The data set for the second non-nearest neighbor was too small to accurately determine values; thus, more data are necessary to determine if second non-nearest neighbors contribute to loop stability and, if so, to what extent.

dx.doi.org/10.1021/bi300008j | Biochemistry 2012, 51, 2192−2198

In Table 3, all possible stem−loop, closing base pair, and non-nearest neighbor combinations are shown, along with the overall free energy contribution of the tetraloop. These values were calculated using the new model presented here (eq 3). The $^{5'}$CCCC$^{3'}$ value was calculated using the initiation term of 4.4 kcal/mol and the value for the destabilization of a polycytosine loop presented by Mathews et al. (2.8 kcal/mol).[7]

DISCUSSION

Thermodynamic Contributions of Non-Nearest Neighbors to Duplex Formation. The data suggest that there is a significant contribution by the first non-nearest neighbors to the thermodynamics of the loop (Tables 1 and 4). For example, the $^{5'}$CGAAAG$^{3'}$ tetraloop has a range of free energies between 2.72 kcal/mol with a $^{5'}$U-A$^{3'}$ first non-nearest neighbor and 3.51 kcal/mol with a $^{5'}$U·G$^{3'}$ first non-nearest neighbor. In previous studies of single-nucleotide bulges,[2,3] it was found that there was a direct correlation between the stability of the stem and the energetic cost of inserting the bulge. Interestingly, in the study presented here, there was no correlation between the stem stability and hairpin stability. The stem−loop structures that have an $^{5'}$A-U$^{3'}$ closing base pair, with fewer hydrogen bonds and therefore less stability, had a range of tetraloop free energies that are approximately the same as those of hairpins with $^{5'}$C-G$^{3'}$ closing base pairs, even when the rest of the loop sequence remained the same.

For the thermodynamic analysis, hairpins with conserved loop and closing base pair sequences were compared; because the majority of the sequence remains the same within all tetraloops, variations in the overall free energies may be due to the identities and directionality of the non-nearest neighbors. The range is considered to be the difference between the most stable non-nearest neighbor sequence and the least stable non-nearest neighbor sequence. The average range of all six loop-closing base pair combinations is 0.98 kcal/mol. The largest range was with the loop-closing base pair sequence $^{5'}$CGCAAG$^{3'}$, at 1.51 kcal/mol, while the smallest range was with the loop-closing base pair sequence $^{5'}$CGUGAG$^{3'}$, at 0.75 kcal/mol. The ranges from each loop-closing base pair sequence can be seen in the last column of Table 4.

Also found in Table 4 are rankings of the non-nearest neighbors; all non-nearest neighbors in each loop-closing base pair combination are ranked from 1 to 6, where 1 represents the least destabilizing tetraloop and 6 represents the most destabilizing tetraloop. Via examination of the rankings, some generalized patterns emerge throughout each loop-closing base pair combination, while certain anomalies also present themselves. As an example of a common trend, the $^{5'}$U·G$^{3'}$ non-nearest neighbor is consistently the most destabilizing of the non-nearest neighbors, with every loop-closing base pair ranked as 6, the most unstable of the tetraloops. Additionally, the $^{5'}$G-C$^{3'}$ and $^{5'}$G·U$^{3'}$ non-nearest neighbors were for the most part highly destabilizing to the tetraloop. The $^{5'}$A-U$^{3'}$ and $^{5'}$U-A$^{3'}$ non-nearest neighbors were usually the least destabilizing to the tetraloop, while the $^{5'}$C-G$^{3'}$ pair was only moderately destabilizing.

Although these broad patterns can be applied to each of the loop-closing base pair combinations, there are certain aberrations within the thermodynamic parameters that are not easily explained. For instance, the $^{5'}$GGCGCAAGUC$^{3'}$ sequence is a $^{5'}$G·U$^{3'}$ first non-nearest neighbor sequence, which, according to the trends previously discussed, should be one of the more destabilized oligonucleotides. Nevertheless, the overall stability of the tetraloop is 2.14 kcal/mol, making the $^{5'}$G·U$^{3'}$ non-nearest neighbor the least destabilizing of the non-nearest neighbors for that particular loop-closing base pair sequence. The remaining non-nearest neighbor combinations follow the basic trend of $^{5'}$U-A$^{3'}$ being the least destabilizing and $^{5'}$U·G$^{3'}$ the most destabilizing. Similarly, the $^{5'}$CGUAGAAAUACG$^{3'}$ sequence should be the least destabilized of the $^{5'}$AGAAAU$^{3'}$ sequences, but it is ranked 4 of 6, meaning that the $^{5'}$U-A$^{3'}$ non-nearest neighbor base pair is among the most destabilizing first non-nearest neighbors to the loop. Further investigations would be necessary to improve our understanding of these specific anomalies in the data.

Certain structural elements of these base pairs may help explain why these trends are observed. Because the number of hydrogen bonds, and therefore the strength of the hydrogen bonding, do not correlate with the observed trends, it is possible that the disparities stem from different stacking interactions between the nucleotides. It is known that the stability of a specific tetraloop depends on the base stacking of the first mismatch with the closing base pair.[21] If the stability or orientation of the closing base pair can be affected by the non-nearest neighbor, then the loop's stability might be increased or decreased on the basis of the non-nearest neighbor as well. Stacking interactions of standard A-form RNA double helices were calculated in a recently published study.[27] These values may be used to help rationalize the three non-nearest neighbor pairs that were included in the updated model.

For the $^{5'}$UC$^{3'}$/$^{5'}$GA$^{3'}$ step, which is the least destabilizing overall, the calculated stacking energy is −9.20 kcal/mol, while the $^{5'}$CC$^{3'}$/$^{5'}$GG$^{3'}$ stem, which is also only slightly destabilizing in the model, contributes −8.16 kcal/mol of stability. The most destabilizing of the non-nearest neighbor pairs, $^{5'}$U·G$^{3'}$, has an overall stacking energy with the closing base pair of only −6.80 kcal/mol. Another factor that could influence the $^{5'}$U·G$^{3'}$ destabilization is that it is not a standard Watson−Crick base pair; the amino group from the guanine is turned away from the pairing nucleotide and toward the minor groove, expanding the size of the minor groove. As the angles between the bases change with this specific base pair, the stacking energies cannot be calculated according to the numbers that are found in Johnson et al.[27] This analysis should be viewed with caution, as these calculated values are for standard A-form RNA; because of the proximity of the base pairs to the tetraloop, it could be possible that they are not in standard A-form geometries. However, these values can be considered as being representative of the stacking energies of the closing base pair and non-nearest neighbor of the tetraloop. It appears as if an increased number of stacking interactions between the non-nearest neighbor and the closing base pair results in a less destabilized tetraloop. In addition to the stacking energies, there are other structural features that could clarify the ranking. For example, the $^{5'}$U-A$^{3'}$ non-nearest neighbor could be favored because it is actually less stable than the $^{5'}$C-G$^{3'}$ base pair. This would make the backbone more flexible and could make the tetraloop less energetically costly, similar to the phenomena described by Blose et al.[2] for single-nucleotide bulges.

All previously discussed results are for the first non-nearest neighbor of the tetraloop, but the second non-nearest neighbor was also analyzed for its thermodynamic contribution to the tetraloop (Table 2). For the analysis of the second non-nearest neighbor, the $^{5'}$GCAA$^{3'}$ and $^{5'}$GUGA$^{3'}$ loops were used with an $^{5'}$A-U$^{3'}$ closing base pair. The $^{5'}$GAAA$^{3'}$ loop was excluded

because of several predicted structures that could outcompete the formation of the tetraloop, and the $^{5'}$C-G$^{3'}$ closing base pair was excluded because of the high thermal stability that a guanine-cytosine-rich, 4 bp stem would confer upon the loop. Within the second non-nearest neighbor sequences, it is interesting to note that the effect on the $^{5'}$GCAA$^{3'}$ loop was much more extreme than the effect on the $^{5'}$GUGA$^{3'}$ loop. The range of differences between the free energies on the $^{5'}$GCAA$^{3'}$ loop was 1.11 kcal/mol, while the range for the $^{5'}$GUGA$^{3'}$ loop was only 0.21 kcal/mol. Because of the large differences in the effects of second non-nearest neighbors, as well as the limited sample size studied, more investigations are necessary to determine exactly how much influence second non-nearest neighbors have on the overall stability of the tetraloop.

Updated Model for Prediction of Thermodynamics of Tetraloops. As discussed previously, there have been several attempts to better predict the thermodynamic contribution of hairpin loops to the overall stability of the stem−loop.[10,26,28] After thermodynamic data for 36 new tetraloop sequences and 10 previously characterized tetraloop sequences in different stem sequences had been collected, these values were added to previously published tetraloop results,[10,21−24] and a new model was created. With the addition of the new tetraloop data, the predictive model described by Sheehy et al.[10] still has an average difference between the predicted and experimental values of 0.4 kcal/mol, showing that this model is accurate. However, the term for the non-nearest neighbor influence is limited in its scope and may not correctly demonstrate the importance of non-nearest neighbors in the stability of the tetraloop. Therefore, the model was updated by separating the non-nearest neighbor parameter into a closing base pair term and a separate conditional non-nearest neighbor term. The new model's values for the initiation and the first mismatch bonuses were modified to account for this change in parameters. The first two parameters in the updated model were modified from previous models,[10,26] with a penalty for initiation and a bonus for first mismatches of $^{5'}$G·A$^{3'}$, $^{5'}$U·U$^{3'}$, and $^{5'}$U·G$^{3'}$ due to stabilization from hydrogen bonding between the two bases.[29,30] As mentioned previously, the $\Delta G^{\circ}_{37,\text{CGNRAG or CUNCGG}}$ bonus is only for tetraloops with the sequence $^{5'}$CGNRAG$^{3'}$ or $^{5'}$CUNCGG$^{3'}$; the $^{5'}$C-G$^{3'}$ closing base pair stabilizes the GNRA and UNCG families of tetraloops but not other families.[29] Therefore, this bonus is applied to only these specific loop-closing base pair sequences. The final parameter in the predictive model is a non-nearest neighbor parameter for tetraloops with the sequence $^{5'}$CGNRAG$^{3'}$ or $^{5'}$CUNCGG$^{3'}$. These tetraloops with $^{5'}$U-A$^{3'}$ and $^{5'}$C-G$^{3'}$ non-nearest neighbors ($^{5'}$UCGNRAGA$^{3'}$, $^{5'}$UCUNCGGA$^{3'}$, $^{5'}$CCGNRAGG$^{3'}$, and $^{5'}$CCUNCGGG$^{3'}$) are assigned bonuses because of their stabilizing effect found both in this study and in previous studies.[10] Tetraloops with the sequence $^{5'}$CGNRAG$^{3'}$ or $^{5'}$CUNCGG$^{3'}$ and having a $^{5'}$U·G$^{3'}$ non-nearest neighbor ($^{5'}$UCGNRAGG$^{3'}$ and $^{5'}$UCUNCGGG$^{3'}$) are assigned a penalty because of the non-nearest neighbor's strongly destabilizing effects on the loop for reasons that are discussed above. The updated model predicts the contribution of the tetraloop with an average difference of 0.3 kcal/mol between the predicted and experimental values.

It should be noted that this model is specific to tetraloops and most accurate for the unusually stable loops GNRA and UNCG, as the majority (86.3%) of the published tetraloops used in this model were stable loops. Only 14 of the 101 loops were not members of either family. Nonetheless, the model proposed by Sheehy et al.[10] predicted the contribution of GNRA loops, UNCG loops, and nonstable loops with average differences of 0.3, 0.5, and 0.7 kcal/mol, respectively, between the predicted and experimental free energies. In the updated model presented here, the average differences between the predicted and experimental free energies for GNRA, UNCG, and nonstable loops are 0.2, 0.4, and 0.6 kcal/mol, respectively. The prediction of each type of tetraloop is improved slightly with the more specific conditions set by this model. The standard deviation for these differences also is reduced from 0.4 in the model of Sheehy et al. to 0.3 kcal/mol. Although these may look like minor differences, there are instances in which the new model is significantly better than the model of Sheehy et al. For example, the tetraloop with a sequence of $^{5'}$CGUAGAAAUACG$^{3'}$ has a ΔG°_{37} of −1.56 kcal/mol and a loop contribution of 3.92 kcal/mol. While the previous model predicts the contribution from this tetraloop to be 3.00 kcal/mol,[10] the model proposed here predicts it to be 3.90 kcal/mol. Similarly, the loop sequence $^{5'}$GCUGAGAGGC$^{3'}$ has a ΔG°_{37} of −1.30 kcal/mol[10] and a loop contribution of 3.81 kcal/mol. The previous model predicts the contribution to be 3.00 kcal/mol, while the updated model predicts it to be 3.90 kcal/mol. Although the accuracy and precision are only slightly improved, the new parameters allow for a more universal model.

AUTHOR INFORMATION

Corresponding Author

*Phone: (314) 977-8567. Fax: (314) 977-2521. E-mail: znoskob@slu.edu.

Funding

This work was supported by National Institutes of Health Grant 1R15GM085699-01A1 to B.M.Z.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28*, 235−242.

(2) Blose, J. M., Manni, M. L., Klapec, K. A., Stranger-Jones, Y., Zyra, A. C., Sim, V., Griffith, C. A., Long, J. D., and Serra, M. J. (2007) Non-nearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops. *Biochemistry 46*, 15123−15135.

(3) McCann, M. D., Lim, G. F. S., Manni, M. L., Estes, J., Klapec, K. A., Frattini, G. D., Knarr, R. J., Gratton, J. L., and Serra, M. J. (2011) Non-nearest-neighbor dependence of the stability for RNA group II single-nucleotide bulge loops. *RNA 17*, 108−119.

(4) Mathews, D. H. (2006) Revolutions in RNA secondary structure prediction. *J. Mol. Biol. 359*, 526−532.

(5) 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.

(6) Xia, T., Santalucia, J., 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.

(7) 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.

(8) Serra, M. J., and Turner, D. H. (1995) Predicting thermodynamic properties of RNA. *Methods Enzymol. 259*, 242−261.

(9) Xia, T., Mathews, D. H., and Turner, D. H. (1999) Thermodynamics of RNA secondary structure formation. In *Prebiotic Chemistry, Molecular Fossils, Nucleosides, and RNA* (Söll, D. G., Nishimura, S., and Moore, P. B., Eds.) pp 21−47, Elsevier, New York.
(10) Sheehy, J. P., Davis, A. R., and Znosko, B. M. (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA 16*, 417−429.
(11) Davis, A. R., and Znosko, B. M. (2007) Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry 46*, 13425−13436.
(12) Badhwar, J., Karri, S., Cass, C. K., Wunderlish, E. L., and Znosko, B. M. (2007) Thermodynamic characterization of RNA duplexes containing naturally occuring 1 × 2 nucleotide internal loops. *Biochemistry 46*, 14715−14724.
(13) 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.
(14) Davis, A. R., and Znosko, B. M. (2010) Positional and neighboring base pair effects on the thermodynamic stability of RNA single mismatches. *Biochemistry 49*, 8669−8679.
(15) Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry 38*, 14214−14223.
(16) Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucletodies. *Biochemistry 29*, 278−285.
(17) Siegfried, N. A., Metzger, S. L., and Bevilacqua, P. C. (2007) Folding cooperativity in RNA and DNA is dependent on position in the helix. *Biochemistry 46*, 172−181.
(18) Thulasi, P., Pandya, L. K., and Znosko, B. M. (2010) Thermodynamic characterization of RNA triloops. *Biochemistry 49*, 9058−9062.
(19) McDowell, J. A. (1996) *MeltWin*, version 3.5.
(20) Davis, A. R., and Znosko, B. M. (2008) Thermodynamic characterization of naturally occurring RNA single mismatches with GU nearest neighbors. *Biochemistry 47*, 10178−10187.
(21) 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.
(22) Groebe, D. R., and Uhlenbeck, O. C. (1988) Characterization of RNA hairpin loop stability. *Nucleic Acids Res. 38*, 11725−11735.
(23) Antao, V. P., and Tinoco, I. (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res. 20*, 819−824.
(24) 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.
(25) 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.
(26) Vecenie, C. J., and Serra, M. J. (2004) Stability of RNA hairpin loops closed by AU base pairs. *Biochemistry 43*, 11813−11817.
(27) Johnson, C. A., Bloomingdale, R. J., Ponnusamy, V. E., Tillinghast, C. A., Znosko, B. M., and Lewis, M. (2011) Computational model for predicting experimental RNA and DNA nearest-neighbor free energy rankings. *J. Phys. Chem. B 115*, 9244−9251.
(28) Vecenie, C. J., Morrow, C. V., Zyra, A., and Serra, M. J. (2006) Sequence dependence of the stability of RNA hairpin molecules with six nucleotide loops. *Biochemistry 45*, 1400−1407.
(29) Blose, J. M., Proctor, D. J., Veeraraghavan, N., Misra, V. K., and Bevilacqua, P. C. (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.
(30) Serra, M. J., Axenson, T. J., and Turner, D. H. (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.