Published in final edited form as:
Biochemistry. 2012 March 20; 51(11): 2192–2198. doi:10.1021/bi300008j.

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 Ave., Saint Louis, MO 63103

Abstract

Currently, several models exist to predict the secondary structure of RNA, 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 non-canonical 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. In order to test the current assumption of the Nearest Neighbor Model that the non-nearest neighbors do not affect the stability of the motif, different stem-loop oligonucleotides were optically melted 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 to predict the thermodynamic contributions of a hairpin loop to the overall stability of the stem-loop structure.

Keywords

tetraloops; GNRA; non-nearest neighbors; secondary structure prediction; hairpins

INTRODUCTION

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, non-coding 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, since 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

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

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, then adding in 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 from 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 towards the thermodynamics of the mismatch were unable to 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 bulges, as the free energy of the stem decreased, so did the destabilization of the duplex from a bulge loop of one nucleotide. Similarly, for group II single 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 towards the stability of a secondary structural motif, the extent of such interactions was not investigated. None of the studies mentioned previously have systematically evaluated all six possible non-nearest neighbor combinations' effects on the thermodynamic stability of a structural motif, which is the focus of this study.

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 less of a contribution by the second non-nearest neighbor. The thermodynamic data was 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 based upon 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$^{3'}$, and $^{5'}$GCAA$^{3'}$, with natural percent abundances 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 an $^{5'}$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 possibility 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 G·U and U·G, 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 hairpin and duplex were calculated using Equation 1 as previously described (18) where [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 $[A]_T$ is the total strand concentration. All hairpin concentrations were calculated to be ≥97% of the total concentration.

$$[H] = \frac{-1 + \sqrt{\left(8 * K_D * [A]_T\right) / \left(K_H{}^2\right) + 1}}{(4 * K_D) / \left(K_H{}^2\right)} \quad (1)$$

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 were 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 °C 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 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 new predictive model

After the thermodynamic contributions of each loop were characterized, this dataset 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, $^{5'}$GGGAUAC<u>CCCC</u>GUAUCCA$^{3'}$ (22) where the underlined nucleotides are the loop, was excluded because poly cytosine hairpin loops are known to be unusually unstable (7). This dataset 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 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 $\Delta G°_{37}$, $\Delta H°$, and $\Delta S°$ were

determined via fitting the melt 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 of the structure was determined by subtracting the nearest neighbor values for the stem nucleotides from the overall experimental values. The last column in 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 tetraloop sequence $^{5'}$C<u>GCAA</u>G$^{3'}$. With a different second non-nearest neighbor, the variation can be up to 1.11 kcal/mol, as seen with the sequence $^{5'}$A<u>GCAA</u>U$^{3'}$. 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) to predict tetraloop stability contain parameters that are relatively conserved amongst 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 an $^{5'}$A·G$^{3'}$ mismatch. The most recent model (10) for the prediction of tetraloop contributions to overall stability is shown in Equation 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 Sheehy et al. predictive model (Equation 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-nucleotide stem $^{5'}$GCC$^{3'}$/$^{3'}$CGG$^{5'}$, limiting the utility of this model. To improve the Sheehy et al. model, many combinations of parameters were tested, and the most accurate and precise model (Equation 3) resulted in dividing the non-nearest neighbor parameter into two separate parameters, a bonus for a $^{5'}$C<u>GNRA</u>G$^{3'}$ or a $^{5'}$C<u>UNCG</u>G$^{3'}$ tetraloop and a non-nearest neighbor parameter for those two specific tetraloop sequences.

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

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, $\Delta G^{\circ}_{37,\text{C<u>GNRA</u>G or C<u>UNCG</u>G}}$ and $\Delta G^{\circ}_{37,\text{C<u>GNRA</u>G or C<u>UNCG</u>G non-nearest neighbor}}$, 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, $\Delta G^{\circ}_{37,\text{C<u>GNRA</u>G or C<u>UNCG</u>G}}$, is applicable only if the tetraloop and closing base pair is $^{5'}$C<u>GNRA</u>G$^{3'}$ or $^{5'}$C<u>UNCG</u>G$^{3'}$. 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,

$\Delta G°_{37,\underline{CGNRA}G \text{ or } C\underline{UNCG}G \text{ non-nearest neighbor}}$, is also only applied if the tetraloop and closing base pair sequence is $^{5'}C\underline{GNRA}G^{3'}$ or $^{5'}C\underline{UNCG}G^{3'}$, and the value of the bonus/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 $^{5'}CC\underline{GNRA}GG^{3'}$ or $^{5'}CC\underline{UNCG}GG^{3'}$. A bonus of −0.5 kcal/mol is applied for tetraloops with the sequence $^{5'}UC\underline{GNRA}GA^{3'}$ or $^{5'}UC\underline{UNCG}GA^{3'}$. A penalty of 0.4 kcal/mol is applied for tetraloops with the sequence $^{5'}UC\underline{GNRA}GG^{3'}$ or $^{5'}UC\underline{UNCG}GG^{3'}$. 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 dataset for the second non-nearest neighbor was too small to accurately determine values, thus more data is necessary in order to determine if second non-nearest neighbors contribute to loop stability and, if so, to what extent.

In Table 4, 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 (Equation 4). The $^{5'}CCCC^{3'}$ value was calculated using the initiation term of 4.4 kcal/mol and the value for the destabilization of a poly-cytosine loop presented in 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 3). For example, the $^{5'}C\underline{GAAA}G^{3'}$ tetraloop has a range of free energies between 2.72 kcal/mol with a $^{5'}U\text{-}A^{3'}$ first non-nearest neighbor and 3.51 kcal/mol with a $^{5'}U{\cdot}G^{3'}$ 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-loops that have an $^{5'}A\text{-}U^{3'}$ closing base pair, with fewer hydrogen bonds and therefore less stability, had a range of tetraloop free energies that are approximately the same as hairpins with $^{5'}C\text{-}G^{3'}$ 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; since 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 $^{5'}C\underline{GCAA}G^{3'}$, at 1.51 kcal/mol, while the smallest range was with the loop-closing base pair sequence $^{5'}C\underline{GUGA}G^{3'}$, at 0.75 kcal/mol. The ranges from each loop-closing base pair sequence can be seen in the last column of Table 3.

Also found in Table 3 are rankings of the non-nearest neighbors; all non-nearest neighbors in each loop-closing base pair combination are ranked from one to six, where one is the most stabilized tetraloop, and six is the most destabilized tetraloop. By examining 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 $^{5'}U{\cdot}G^{3'}$ non-nearest neighbor is consistently the most destabilizing of the non-nearest neighbors, with every loop-closing base pair ranked as six, the most unstable of the tetraloops. Additionally, the $^{5'}G\text{-}C^{3'}$ and the $^{5'}G{\cdot}U^{3'}$ non-nearest neighbors were for the most part highly destabilizing to the tetraloop. The $^{5'}A\text{-}U^{3'}$ and $^{5'}U\text{-}A^{3'}$ non-nearest neighbors

were usually the least destabilizing to the tetraloop, while $^{5'}$C-G$^{3'}$ 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 sequence of $^{5'}$GGCGCAAGUC$^{3'}$ 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 $^{5'}$G·U$^{3'}$ 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 least destabilizing and $^{5'}$U·G$^{3'}$ as most destabilizing. Similarly, the sequence $^{5'}$CGUAGAAAUACG$^{3'}$ should be the least destabilized of the $^{5'}$AGAAAU$^{3'}$ sequences, but it is ranked four out of six, 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 better understand these specific anomalies in the data.

Certain structural elements of these base pairs may help explain why these trends are observed. Since the number of hydrogen bonds, and therefore the strength of the hydrogen bonding, do not correlate with the trends observed, 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 based upon 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, the $^{5'}$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; due to the base pairs' proximity to the tetraloop, it could be possible that they are not in standard A-form geometries. However, these values can be considered as representative of the stacking energies of the closing base pair and non-nearest neighbor of the tetraloop. It appears as if increased 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 due to the fact that 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 the $^{5'}$GUGA$^{3'}$ loops were used with an $^{5'}$A–U$^{3'}$ closing base pair. The $^{5'}$GAAA$^{3'}$ loop

was excluded due to several predicted structures that could out-compete the formation of the tetraloop, and the $^{5'}$C–G$^{3'}$ closing base pair was excluded due to the high thermal stability that a guanine-cytosine-rich, four base pair 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. Due to the large differences in 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 towards the overall stability of the stem-loop (10, 26, 28). After collecting thermodynamic data for 36 new tetraloop sequences and ten previously characterized tetraloop sequences in different stem sequences, 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 the 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 on 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 $\Delta G°_{37,\underline{\text{CGNRAG}}\text{ or }\underline{\text{CUNCGG}}}$ bonus is only for tetraloops with the sequence $^{5'}$C$\underline{\text{GNRA}}$G$^{3'}$ or $^{5'}$C$\underline{\text{UNCG}}$G$^{3'}$; 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 only applied to 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 $^{5'}$C$\underline{\text{GNRA}}$G$^{3'}$ or $^{5'}$C$\underline{\text{UNCG}}$G$^{3'}$. These tetraloops with $^{5'}$U-A$^{3'}$ and $^{5'}$C–G$^{3'}$ nearest neighbors ($^{5'}$CC$\underline{\text{GNRA}}$GG$^{3'}$, $^{5'}$CC$\underline{\text{UNCG}}$GG$^{3'}$, $^{5'}$CC$\underline{\text{GNRA}}$GG$^{3'}$, and $^{5'}$CC$\underline{\text{UNCG}}$GG$^{3'}$) are assigned bonuses due to their stabilizing effect found both in this study and in previous studies (10). Tetraloops with the sequence $^{5'}$C$\underline{\text{GNRA}}$G$^{3'}$ or $^{5'}$C$\underline{\text{UNCG}}$G$^{3'}$ and having a $^{5'}$U·G$^{3'}$ non-nearest neighbor ($^{5'}$UC$\underline{\text{GNRA}}$GG$^{3'}$ and $^{5'}$UC$\underline{\text{UNCG}}$GG$^{3'}$) are assigned a penalty due to 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 the experimental values.

It should be noted that this model is specific towards 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 fourteen 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 non-stable loops with an average difference 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 non-stable 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 Sheehy et al. model to 0.3 kcal/mol. Although these may look like minor differences, there are instances where the new model is significantly better than the Sheehy et al. model. For example, the tetraloop with a sequence of $^{5'}$CGUAGAAAUACG$^{3'}$ has a $\Delta G°_{37}$ 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 $\Delta G°_{37}$ 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.

Acknowledgments

FUNDING This work was supported by the National Institutes of Health [1R15GM085699-01A1 to BMZ].

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]
2. Blose JM, Manni ML, Klapec KA, Stranger-Jones Y, Zyra AC, Sim V, Griffith CA, Long JD, Serra MJ. non-nearest neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops. Biochemistry. 2007; 46:15123–15135. [PubMed: 18047298]
3. McCann MD, Lim GFS, Manni ML, Estes J, Klapec KA, Frattini GD, Knarr RJ, Gratton JL, Serra MJ. Non-nearest-neighbor dependence of the stability for RNA group II single-nucleotide bulge loops. RNA. 2010; 17:108–119. [PubMed: 21088109]
4. Mathews DH. Revolutions in RNA secondary structure prediction. J. Mol. Biol. 2006; 359:526–532. [PubMed: 16500677]
5. Lu ZJ, Turner DH, Mathews DH. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. Nucleic Acids Res. 2006; 34:4912–4924. [PubMed: 16982646]
6. Xia T, Santalucia J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry. 1998; 37:14719–14735. [PubMed: 9778347]
7. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 1999; 288:911–940. [PubMed: 10329189]
8. Serra, MJ.; Turner, DH. Method Enzymol. 1995. Predicting thermodynamic properties of RNA; p. 242-261.
9. Xia T, Mathews DH, Turner DH. Thermodynamics of RNA secondary structure formation, Comp. Natural Products Chem. 199; 6:21–47.
10. Sheehy JP, Davis AR, Znosko BM. Thermodynamic characterization of naturally occurring RNA tetraloops. RNA. 2010; 16:417–429. [PubMed: 20047989]
11. Davis AR, Znosko BM. Thermodynamic characterization of single mismatches found in naturally occurring RNA. Biochemistry. 2007; 47:13425–13436. [PubMed: 17958380]
12. Badhwar J, Karri S, Cass CK, Wunderlish EL, Znosko BM. Thermodynamic characterization of RNA duplexes containing naturally occuring 1×2 nucleotide internal loops. Biochemistry. 2007; 46:14715–14724. [PubMed: 18020450]
13. Wright DJ, Rice JL, Yanker DM, Znosko BM. Nearest neighbor parameters for inosine-uridine pairs in RNA duplexes. Biochemistry. 2007; 46:4625–4634. [PubMed: 17378583]
14. Davis AR, Znosko BM. Positional and neighboring base pair effects on the thermodynamic stability of RNA single mismatches. Biochemistry. 2010; 49:8669–8679. [PubMed: 20681613]

15. Kierzek R, Burkard ME, Turner DH. Thermodynamics of single mismatches in RNA duplexes. Biochemistry. 1999; 38:14214–14223. [PubMed: 10571995]
16. Longfellow CE, Kierzek R, Turner DH. Thermodynamic and spectroscopic study of bulge loops in oligoribonucletodies. Biochemistry. 1990; 29:278–285. [PubMed: 2322546]
17. Siegfried NA, Metzger SL, Bevilacqua PC. Folding cooperativity in RNA and DNA is dependent on position in the helix. Biochemistry. 2007; 46:172–181. [PubMed: 17198387]
18. Thulasi P, Pandya LK, Znosko BM. Thermodynamic characterization of RNA triloops. Biochemistry. 2010; 49:9058–9062. [PubMed: 20843054]
19. McDowell, JA. MeltWin. version 3.5. 1996.
20. Davis AR, Znosko BM. Thermodynamic characterization of naturally occurring RNA single mismatches with GU nearest neighbors. Biochemistry. 2008; 47:10178–10187. [PubMed: 18754680]
21. Dale T, Smith R, Serra MJ. A test of the model to predict unusually stable RNA hairpin loop stability. RNA. 2000; 6:608–615. [PubMed: 10786851]
22. Groebe DR, Uhlenbeck OC. Characterization of RNA hairpin loop stability. Nucleic Acids Res. 1988; 38:11725–11735. [PubMed: 3211748]
23. Antao VP, Tinoco I. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. Nucleic Acids Res. 1992; 20:819–824. [PubMed: 1371866]
24. Giese MR, Betschart K, Dale T, Riley CK, Rowan C, Sprouse KJ, Serra MJ. Stability of RNA hairpins closed by wobble base pairs. Biochemistry. 1998; 37:1094–1100. [PubMed: 9454601]
25. Serra MJ, Barnes TW, Betschart K, Gutierrez MJ, Sprouse KJ, Riley CK, Stewart L, Temel RE. Improved parameters for the prediction of RNA hairpin stability. Biochemistry. 1997; 36:4844–4851. [PubMed: 9125504]
26. Vecenie CJ, Serra MJ. Stability of RNA hairpin loops closed by AU base pairs. Biochemistry. 2004; 43:11813–11817. [PubMed: 15362866]
27. Johnson CA, Bloomingdale RJ, Ponnusamy VE, Tillinghast CA, Znosko BM, Lewis M. Computational model for predicting experimental RNA and DNA nearest neighbor free energy rankings. J. Phys. Chem. B. 2011; 115:9244–9251. [PubMed: 21619071]
28. Vecenie CJ, Morrow CV, Zyra A, Serra MJ. Sequence dependence of the stability of RNA hairpin molecules with six nucleotide loops. Biochemistry. 2006; 45:1400–1407. [PubMed: 16445282]
29. Blose JM, Proctor DJ, Veeraraghavan N, Misra VK, Bevilacqua PC. Contribution of the closing base pair to exceptional stability in RNA tetraloops: Roles for molecular mimicry and electrostatic factors. J. Am. Chem. Soc. 2009; 131:8474–8484. [PubMed: 19476351]
30. Serra MJ, Axenson TJ, Turner DH. A model for the stabilities of RNA hairpins based on a study of the sequence dependence of stability for hairpins of six nucleotides. Biochemistry. 1994; 33:14289–14296. [PubMed: 7524674]

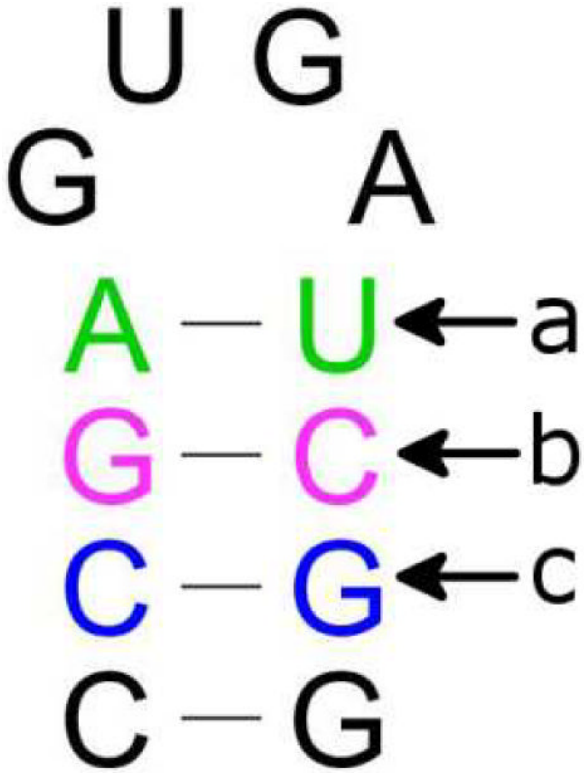

Figure 1.
A $^{5'}$GUGA$^{3'}$ tetraloop in which the closing base pair/ nearest neighbor ($^{5'}$A–U$^{3'}$) is colored green (a), the first non-nearest neighbor ($^{5'}$G–C$^{3'}$) is colored magenta (b), and the second non-nearest neighbor ($^{5'}$C–G$^{3'}$) is colored blue (c).

Biochemistry. Author manuscript; available in PMC 2013 March 20.

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)	T_m(°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

Loop-Closing Base Pair Sequence[b]	Sequence[b]	ΔH° (kcal/mol)	ΔS° (eu)	ΔG° $_{37}$(kcal/mol)	T_m(°C)	ΔG° $_{37,tetraloop}$ (kcal/mol)[c]
	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 made in 1.O M NaCl, 10 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 in 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

[d] (21)

[e] (10)

Table 2

Thermodynamic parameters and free energy contributions of tetraloops with varying the 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
	CAGAGUGACUG	−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 made in 1.0 M NaCl, 10 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 neighbor are in 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

Free energies of GNRA tetraloops and rankings of first non-nearest neighbors[a]

		Non-nearest neighbor identities												
		G–C		C–G		G·U		U·G		A–U		U–A		
		$\Delta G°_{tetraloop}$ values (kcal/mol) and rankings[a]												Range (kcal/mol)
Loop sequence with closing base pairs[b]	**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] Rankings are based on calculated free energy of the RNA tetraloops, with 1 as least destabilizing and 6 as 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 structure. Rankings are bold and italicized. Rankings 1–6 are based upon the closina base pair and loop sequence, i.e. XCGAAAGY is ranked from 1 throuah 6.

[b] Underlined nucleotides are the tetraloop sequence.

[c] ΔG° values are averages from two separate trials run of these sequences. Both were within experimental.

Table 4

Predicted free energies of tetraloops (kcal/mol) using proposed model

Loop Sequence[a]	Closing Base Pair Identity[b]	Non-Nearest Neighbor 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 Any other base pair	3.9	3.9	3.4	3.5	3.9	3.9	4.3
UNCG	CG Any other base pair	3.9	3.9	3.4	3.5	3.9	3.9	4.3
CCCC	Any	7.2						

[a] Loop sequences are written 5'→3'. N represents any nucleotide, Y represents any pyrimidine, R represents any purine, and D represents 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.