Published in final edited form as:

Biochemistry. 2012 July 3; 51(26): 5359–5368. doi:10.1021/bi3001227.

Thermodynamic Characterization of RNA 2×3 Nucleotide Internal Loops

Nina Z. Hausmann and Brent M. Znosko*

Department of Chemistry, Saint Louis University; Saint Louis, MO 63103

Abstract

In order to better elucidate RNA structure-function relationships and to improve design of pharmaceutical agents that target specific RNA motifs, an understanding of RNA primary, secondary, and tertiary structure is necessary. The prediction of RNA secondary structure from sequence is an intermediate step in predicting RNA three-dimensional structure. RNA secondary structure is typically predicted using a nearest neighbor model based on free energy parameters. The current free energy parameters for 2×3 nucleotide loops are based on a dataset of 23 2×3 loops and internal loops of other sizes. A database of representative RNA secondary structures was searched to identify 2×3 nucleotide loops that occur in nature. Seventeen of the most frequent 2×3 nucleotide loops in this database were studied by optical melting experiments. Fifteen of these loops melted in a two-state manner, and the associated experimental $\Delta G^\circ_{37,2\times 3}$ values are, on average, 0.6 and 0.7 kcal/mol different from the values predicted for these internal loops using the predictive models proposed by Lu, Turner, and Mathews [Lu, Z. J., Turner, D. H., and Mathews, D. H. (2006) *Nucleic Acids Res.* 34, 4912–4924] and Chen and Turner [Chen, G. and Turner, D. H. (2006) *Biochemistry* 45, 4025–4043], respectively. These new $\Delta G^\circ_{37,2\times 3}$ values can be used to update the current algorithms that predict secondary structure from sequence. In order to improve free energy calculations for duplexes containing 2×3 nucleotide loops that still do not have experimentally-determined free energy contributions, an updated predictive model was derived. This new model resulted from a linear regression analysis of the data reported here combined with 31 previously studied 2×3 nucleotide internal loops. Most of the values for the parameters in this new predictive model are within experimental error of the previous models, suggesting that approximations and assumptions associated with the derivation of the previous nearest neighbor parameters were valid. The updated predictive model predicts free energies of 2×3 nucleotide internal loops within 0.4 kcal/mol, on average, of the experimental free energy values. Both the experimental values and the updated predictive model can be used to improve secondary structure prediction from sequence.

While about half of RNA nucleotides are found in Watson-Crick regions, the other half are found in bulges, hairpin loops, internal loops, or multi-branch loops (1). One common non-Watson-Crick motif is a 2×3 nucleotide internal loop, which has been found to occur in the human immunodeficiency virus type 1 (HIV-1) Rev response element (2, 3), the large ribosomal subunit from *Haloarcula marismortui* (4), the *Thermus thermophilus* 30S ribosomal subunit (5), the P4-P6 domain of the *Tetrahymena thermophila* (6) and *Pneumocystis carinii* (7) group I introns, the purine-binding domain of the guanine riboswitch from the xpt-pbuX operon of *Bacillus subtilis* (8), and in many other organisms (Fig. 1). Based on the sequences, 2×3 nucleotide loops are predicted to form in the 23S ribosomal subunit of the fungal plant pathogen *Mucor racemosus* (9, 10), the 16S ribosomal subunit of the rodent parasite *Giardia muris* (11), the group I intron of *Chlorella*

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

saccharophila (12, 13), and the signal recognition particle from the flowering plant *Cineraria hybrid* (14), to name a few.

Not only do 2×3 nucleotide loops occur in different types of RNA and in a wide range of organisms, but they also can be found in regions that serve important biological roles. One example is the 2×3 nucleotide loop found in the *Homo sapiens* 18S ribosomal decoding A-site (Fig. 1), which plays an important role in protein synthesis (15). The decoding A-site works to recognize the cognate interactions between the tRNA anticodon and the messenger codon. Specifically, when an activated tRNA with elongation factor EF-Tu and GTP is delivered to the A site, the adenines found within the 2×3 nucleotide loop change their conformation from an 'off' state to an 'on' state. Interactions between the adenosine residues (while in the 'on' state) of the 2×3 nucleotide loop and the codon-anticodon contribute to the fidelity of the tRNA selection step (15). Another example of a biological role played by 2×3 nucleotide loops is the Rev response element in mRNA from HIV-1 (Fig. 1). The Rev response element contains a 2×3 nucleotide loop that interacts with Rev, a small regulatory protein of HIV-1 that controls viral replication (16, 17).

Due to the frequency of occurrence of 2×3 nucleotide loops in nature and their important biological roles, they are of interest to researchers in a wide variety of fields. Being able to accurately predict the thermodynamic contribution of 2×3 nucleotide loops to duplex stability is crucial in order to predict RNA secondary structure from sequence. Accurate secondary structures can provide information about structure-function relationships, three-dimensional structures, and tertiary interactions; thereby aiding in the design of therapeutics. Since the thermodynamics of less than 3% of all possible 2×3 nucleotide loops have been experimentally measured, the thermodynamic contribution of most 2×3 nucleotide loops to duplex stability is predicted by an algorithm (18, 19). One algorithm (18) predicts the free energy contribution of 2×3 nucleotide loops to duplex stability using the following equation:

$$\Delta G^{\circ}_{37,2\times 3} = \Delta G^{\circ}_{37,\text{loop initiation}} + \Delta G^{\circ}_{37,\text{AU/GU}} + \Delta G^{\circ}_{37,\text{asym}} + \Delta G^{\circ}_{37,5'\text{YA}/3'\text{RG}} + \Delta G^{\circ}_{37,5'\text{RG}/3'\text{YA}} + \Delta G^{\circ}_{37,5'\text{YG}/3'\text{RA}} + \Delta G^{\circ}_{37,\text{GG}} + \Delta G^{\circ}_{37,\text{UU}}$$

Here, $\Delta G^{\circ}_{37,\text{loop initiation}}$ is an initiation term for an internal loop containing five nucleotides (2.0 kcal/mol); $\Delta G^{\circ}_{37,\text{AU/GU}}$ is a penalty applied per A-U, U-A, G-U, or U-G pair adjacent to the loop (0.7 kcal/mol); $\Delta G^{\circ}_{37,\text{asym}}$ is a term for internal loops with unequal numbers of nucleotides on each side (0.6 kcal/mol); $\Delta G^{\circ}_{37,5'\text{YA}/3'\text{RG}}$, $\Delta G^{\circ}_{37,5'\text{RG}/3'\text{YA}}$, and $\Delta G^{\circ}_{37,5'\text{YG}/3'\text{RA}}$ are bonuses for the stack of a specific closing base pair with the first "pair" in the loop (−0.5, −1.2, and −1.1 kcal/mol, respectively); and $\Delta G^{\circ}_{37,\text{GG}}$ and $\Delta G^{\circ}_{37,\text{UU}}$ are bonuses for loops that contain a possible G-G or U-U pair as the first or last "pair" in the loop (−0.7 and −0.4 kcal/mol, respectively). While most of the parameters were derived from the published dataset of 2×3 nucleotide loops (7, 20, 21), $\Delta G^{\circ}_{37,\text{AU/GU}}$ and $\Delta G^{\circ}_{37,\text{asym}}$ were derived from data of loops of various sizes (18).

A second, similar algorithm (19) predicts the free energy contribution of 2×3 nucleotide loops to duplex stability using the same equation (eq 1) with slightly different values. Here, $\Delta G^{\circ}_{37,\text{loop initiation}}$ is 2.15 kcal/mol, $\Delta G^{\circ}_{37,\text{AU/GU}}$ is 0.73 kcal/mol, $\Delta G^{\circ}_{37,\text{asym}}$ is 0.45 kcal/mol, $\Delta G^{\circ}_{37,5'\text{YA}/3'\text{RG}}$ is −0.39 kcal/mol, $\Delta G^{\circ}_{37,5'\text{RG}/3'\text{YA}}$ is −1.06 kcal/mol, $\Delta G^{\circ}_{37,5'\text{YG}/3'\text{RA}}$ is −1.41 kcal/mol, $\Delta G^{\circ}_{37,\text{GG}}$ is −0.74 kcal/mol, and $\Delta G^{\circ}_{37,\text{UU}}$ is −0.34 kcal/mol. Like the first algorithm, most parameter values were derived from 2×3 experimental data, with the exception of $\Delta G^{\circ}_{37,\text{loop initiation}}$ and $\Delta G^{\circ}_{37,\text{asym}}$, which were derived from data of loops of various sizes.

The purpose of this study is two-fold. First, by collecting thermodynamic data for previously unstudied 2×3 nucleotide loops that occur frequently in nature, this thermodynamic data can

be directly incorporated into *RNAstructure*, an RNA secondary structure prediction program based on free energy minimization (1, 22), and other programs that predict RNA secondary structure (23–26). This means that the thermodynamic contribution of these loops to RNA stability can be calculated from the experimental values instead of the predictive models currently used by *RNAstructure*. Second, by collecting thermodynamic data for additional loops, the total database of available 2×3 thermodynamics is increased. The larger database can then be used to derive an updated predictive model to be used to calculate the thermodynamic contribution of 2×3 loops that do not have experimental values.

In order to determine which 2×3 loops to study, a database of 1349 RNA secondary structures was searched for the most frequently occurring 2×3 loops. The results of this search guided our experiments, and the thermodynamics of 17 frequently occurring loops in the database are reported. The Znosko laboratory has previously published the thermodynamics of naturally occurring single mismatches (also referred to as 1×1 nucleotide internal loops) (27, 28), tandem mismatches (also referred to as 2×2 nucleotide internal loops) (29), 1×2 nucleotide internal loops (30), triloops (also referred to as hairpin loops of three nucleotides) (31), and tetraloops (also referred to as hairpin loops of four nucleotides) (32). This study is the next in the series of studies investigating the thermodynamics of small secondary structure motifs in RNA.

MATERIALS AND METHODS

Compiling and Searching a Database for 2×3 Nucleotide Loops

A database of 1349 RNA secondary structures containing various types of RNA was previously compiled (31, 32). This database was searched for 2×3 nucleotide loops, and the number of occurrences for each type of 2×3 nucleotide loop was tabulated (Table 1). For this work, G-U pairs were considered to be canonical base pairs.

Optical Melting Experiments

The general design of sequences to be used for optical melting studies has been previously described (30). The sequences of loops and canonically base-paired nearest neighbors were chosen based on the search results of the database described above. The focus of this study was the most frequently occurring 2×3 loop sequences in the database with their adjacent nearest neighbors. One frequently occurring loop (13 occurrences in the database, Table 1), $\begin{pmatrix} \text{CAAG} \\ \text{GGGGC} \end{pmatrix}$, was not studied here due to the presence of multiple guanines in a row and the possible formation of unwanted structures and/or aggregation. In all duplex sequences, the loop nucleotides and their nearest neighbors were situated in the middle of a duplex, with three Watson-Crick base pairs on either side. Most of the loop-nearest neighbor combinations were inserted in the same stem, $\begin{pmatrix} \text{GACWXXWCUG} \\ \text{CUGWXXXWGAC} \end{pmatrix}$, where W are the nearest neighbor nucleotides, and X are the nucleotides in the 2×3 nucleotide loop. Three loop-nearest neighbor combinations were inserted into a different stem due to the possible formation of competing structures involving this standard stem.

Oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA). The purification of the oligonucleotides followed standard procedures (33, 34) as described previously (35). Concentration calculations and the formation of duplexes from single strands are standard and were also described previously (30).

Newly formed duplexes were lyophilized and re-dissolved in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na₂EDTA, pH = 7.0. The melting scheme and details of the standard melting experiments were described previously (30). *Meltwin* (36) was used to fit melting curves to a two-state model and to derive the duplex thermodynamics (30).

Determination of the Contribution of 2×3 nucleotide loops to Duplex Thermodynamics

The measured free energy change for the duplex can be contributed to an initiation term, the free energy contribution of the 2×3 loop, and the sum of the nearest neighbor interactions of Watson-Crick base pairs in the stem (37). For example:

$$\Delta G^{\circ}_{37}(\text{GACG CC CCUG})_{\text{CUGCAAUGGAC}} = \Delta G^{\circ}_{37,i} + \Delta G^{\circ}_{37}(\text{GA}_{\text{CU}}) + \Delta G^{\circ}_{37}(\text{AC}_{\text{UG}}) + \Delta G^{\circ}_{37}(\text{CG}_{\text{GC}}) + \Delta G^{\circ}_{37,2\times 3} + \Delta G^{\circ}_{37}(\text{CC}_{\text{GG}}) + \Delta G^{\circ}_{37}(\text{CU}_{\text{GA}}) + \Delta G^{\circ}_{37}(\text{UG}_{\text{AC}})$$

where $\Delta G^{\circ}_{37,i}$ is the free energy change for duplex initiation, 4.09 kcal/mol (37), $\Delta G^{\circ}_{37,2\times 3}$ is the free energy contribution from the 2×3 nucleotide loop, and the remainder of the terms are individual nearest neighbor values (37). A rearrangement of the terms in this equation to solve for the stability of the 2×3 loop results in eq 3:

$$\Delta G^{\circ}_{37,2\times 3} = \Delta G^{\circ}_{37}(\text{GACG CC CCUG})_{\text{CUGCAAUGGAC}} - \Delta G^{\circ}_{37,i} - \Delta G^{\circ}_{37}(\text{GA}_{\text{CU}}) - \Delta G^{\circ}_{37}(\text{AC}_{\text{UG}}) - \Delta G^{\circ}_{37}(\text{CG}_{\text{GC}}) - \Delta G^{\circ}_{37}(\text{CC}_{\text{GG}}) - \Delta G^{\circ}_{37}(\text{CU}_{\text{GA}}) - \Delta G^{\circ}_{37}(\text{UG}_{\text{AC}})$$

Here, $\Delta G^{\circ}_{37}(\text{GACG CC CCUG})_{\text{CUGCAAUGGAC}}$ is the value determined by optical melting experiments. More specifically:

$$\Delta G^{\circ}_{37,2\times 3} = -7.17 - 4.09 - (-2.35) - (-2.24) - (-2.36) - (-3.26) - (-2.08) - (-2.11) = 3.14 \text{ kcal/mol}$$

This method of calculating $\Delta G^{\circ}_{37,2\times 3}$ values utilizes nearest neighbor parameters (37). A different method of calculating $\Delta G^{\circ}_{37,2\times 3}$ values utilizes reference duplexes. Reference duplexes were not used here because the stability of an additional duplex, the reference duplex, would need to be measured in order to calculate the thermodynamic contribution of each of the 2×3 nucleotide loops studied here. Therefore, all of the $\Delta G^{\circ}_{37,2\times 3}$ values reported here were calculated by the nearest neighbor method. For loops previously studied (7, 20, 21, 38), regardless of how the loop data was calculated in the original publications, $\Delta G^{\circ}_{37,2\times 3}$ values were also calculated by equations similar to eq 4, not from reference duplexes. Similar calculations were done for $\Delta H^{\circ}_{2\times 3}$ and $\Delta S^{\circ}_{2\times 3}$.

Chen and Turner (2006) thermodynamically characterized eight unique 2×3 nucleotide internal loops (19); but all of these loops were inserted into a stem with a dangling purine nucleotide. Reference duplexes were used to calculate the contribution of the 2×3 nucleotide loops to duplex stability. Because the nearest neighbor method was used to calculate the contribution of the 2×3 nucleotide loops studied here to duplex stability (eqs 3 and 4), the authors also wanted to use that same method with the Chen and Turner data. Unfortunately, there are no reliable nearest-neighbor parameters for dangling purine nucleotides. As a result, these loops were omitted from the analysis described here.

One additional previously measured loop (20) was omitted from the analysis. Based on an analysis of the sequence and possible competing secondary structures, the authors were not confident that the duplex $(\text{ACCUGCUUGC})_{\text{UGGAACAAACG}}$ was forming the desired 2×3 internal loop (other competing structures from the association of the two strands are possible); therefore, it was omitted from the analysis.

Linear Regression and Updated Predictive Model

The melting of two duplexes studied here, $(\text{GACGGAGCUG})_{\text{CUGCAUACGAC}}$ and $(\text{GACUAAACUG})_{\text{CUGAGACUGAC}}$, were considered non-two-state. For $(\text{GACGGAGCUG})_{\text{CUGCAUACGAC}}$, the data derived from the individual melting curves and from the T_M versus C_T plot resulted in large error values. For $(\text{GACUAAACUG})_{\text{CUGAGACUGAC}}$, the ΔH° values derived from the analysis of the melt curve fits and the analysis of the T_m

dependence did not agree within 15%. As a result, these two duplexes were not included in the analysis described below. In order to create a larger dataset of 2×3 nucleotide loop data, the data for the 15 sequences studied here were combined with the data for 31 previously published sequences (7, 20, 21, 38), resulting in a dataset of 46 sequences containing 2×3 nucleotide internal loops. Data from these 46 sequences were subjected to linear regression using *Microsoft Excel*'s LINEST function. The loop thermodynamic values were used as constants, and many different combinations of parameters (using a variety of bonuses and penalties) were tried as variables. The combination of variables (which is similar to those in eq 1) that resulted in a model that best agreed with the experimental data included a term for 2×3 nucleotide loop initiation, an A-U or G-U closure, a 5'RG3'/3'YA5' stack, a 5'YG3'/3'RA stack, a potential G-G pair, and a potential U-U pair. Internal loop thermodynamic parameters were derived for $\Delta G^{\circ}_{37,2\times 3}$, $\Delta H^{\circ}_{2\times 3}$, and $\Delta S^{\circ}_{2\times 3}$.

RESULTS

Database Searching

The database containing 1349 structures (described in Materials and Methods) was searched for 2×3 nucleotide loops. As a result, 666 2×3 nucleotide loops were found. On average, about one 2×3 nucleotide loop was found in every two structures. A summary of database results is shown in Table 1.

The first set of data in Table 1 represents the loops when the loop nucleotides and the nearest neighbor canonical base pairs are specified, resulting in a total of 252 different types of loops in the database. Many of the different types of loops do not occur frequently. In fact, 227 types of loops, accounting for 45% of the loops found in the database, individually account for <0.6% of the total number of loops found. Table 1 shows the frequency and percent occurrence of only the 25 most frequently occurring 2×3 nucleotide loops, which represent the other 55% of the 2×3 nucleotide loops found in the database. The most frequent loop is $\begin{pmatrix} \text{GCCC} \\ \text{CAAUG} \end{pmatrix}$, which accounts for 14% of all 2×3 nucleotide loops found in the database. Although it was pioneering work, the initial thermodynamic experiments on 2×3 nucleotide internal loops (7, 20, 21, 38) only began to study the possible sequence combinations. When specifying the loop nucleotides and the nearest neighbor base pairs, previous studies have only characterized one type of loop in the top 25, representing only 3% of the total number of loops found, while the data collected here increases this to 18 loops characterized in the top 25, representing 49% of the total number of loops found.

The second set of data in Table 1 lists the frequency and percent occurrence when only the loop nucleotides are specified. When characterized in this fashion, a total of 135 different types of loops were found. In Table 1, only the 25 most frequent occurring 2×3 nucleotide loops, which account for 71% of all possible types of 2×3 nucleotide loops, are shown. The remaining 110 2×3 nucleotide loops represent the remaining 29% of the 2×3 nucleotide loops found in the database, individually representing <0.9% of the total number of loops found. The most frequent loop is $\begin{pmatrix} \text{CC} \\ \text{AAU} \end{pmatrix}$, accounting for 14% of the 2×3 nucleotide loops found in the database. Previous studies have only characterized 8 different types of loops, representing only 30% of 2×3 nucleotide loops, while the data collected here increases this to 19 different types of loops, representing 64% of 2×3 nucleotide loops.

The third set of data in Table 1 lists the frequency and percent occurrence of 5' and 3' nearest neighbor combinations. This analysis of the data results in 33 of 36 possible types of nearest neighbor combinations. The three possible nearest neighbor combinations not found in the database all contain a G-U pair. The 25 nearest neighbor combinations listed in the third data set account for 97% of the total number of combinations found, while the

remaining eight nearest neighbor combinations account for the remaining 3%. The most frequently occurring nearest neighbor combination is ($\begin{smallmatrix} \text{GXXC} \\ \text{CXXXG} \end{smallmatrix}$), representing 16% of all nearest neighbor combinations found adjacent to 2×3 nucleotide loops. As shown in Table 1, previous studies account for only 43% of the possible nearest neighbor combinations. Once the data reported here was included, this percentage increased to 82%. Similarly, previous studies only thermodynamically characterized eight nearest neighbor combinations in the top 25. Once the data reported here was included, 14 nearest neighbor combinations in the top 25 have been studied.

The fourth set of data in Table 1 gives the frequency and percent occurrence of the loop nucleotides when cytidine and uridine are classified as pyrimidines, and adenosine and guanosine are classified as purines. This arrangement of data results in 25 different types of loop sequences that are found in the database, all of which are shown in Table 1. An all purine loop is the most frequently occurring combination, representing 19% of all 2×3 nucleotide loops found in the database. Previous studies account for 52%, while the data reported here increases the percentage to 82%.

Thermodynamic Parameters

Thermodynamic parameters for duplex formation derived from an analysis of the individual melt curves and an analysis of the T_M dependence are listed in Table 2. The data shown in Table 2 correspond to 18 of the 19 most frequently occurring loops found in the database (Table 1, first set of data) and are listed in order of decreasing frequency. As previously stated, one sequence with three adjacent guanines was not studied due to possible competing structures.

Contribution of 2×3 Nucleotide Loops to Duplex Thermodynamics

The contribution of 2×3 nucleotide loops to duplex thermodynamics (Table 3) was calculated as described in Materials and Methods and as defined by eq 3 and 4 for $\Delta G^{\circ}_{37,2\times 3}$. In addition to 16 duplexes studied here (and listed in Table 2), 30 additional duplexes that occur less frequently but were studied previously (7, 20, 21, 38) are also included.

Free Energy Parameters for 2×3 Nucleotide Internal Loops

While analyzing the data listed in Table 3 by linear regression, several models (a variety of penalty and bonus combinations) to predict the thermodynamics of 2×3 nucleotide loops were tested. A combination of parameters and their corresponding values (Table 4) similar to those used in eq 1 produced an equation that resulted in predicted $\Delta G^{\circ}_{37,2\times 3}$ values that were very close to the experimental $\Delta G^{\circ}_{37,2\times 3}$ values reported here:

$$\Delta G^{\circ}_{37,2\times 3} = \Delta G^{\circ}_{37,2\times 3 \text{ initiation}} + \Delta G^{\circ}_{37,\text{AU/GU}} + \Delta G^{\circ}_{37,5'\text{RG}/3'\text{YA}} + \Delta G^{\circ}_{37,5'\text{YG}/3'\text{RA}} + \Delta G^{\circ}_{37,\text{GG}} + \Delta G^{\circ}_{37,(5)}$$

Note that the $\Delta G^{\circ}_{37,2\times 3 \text{ initiation}}$ and the $\Delta G^{\circ}_{37,\text{asym}}$ terms in eq 1 were combined into the $\Delta G^{\circ}_{37,2\times 3 \text{ initiation}}$ term in eq 5. Also, the $\Delta G^{\circ}_{37,5'\text{YA}/3'\text{RG}}$ term in eq 1 was not included in this new model because it was not found to contribute significantly to loop stability. The values corresponding to each of these parameters can be found in Table 4. Most of the new $\Delta G^{\circ}_{37,2\times 3}$ parameters are within experimental error of the previously proposed values (18, 19). The same set of parameters was used in a linear regression analysis of the enthalpy and entropy data. All of the values for the new $\Delta H^{\circ}_{2\times 3}$ parameters (Table 4) are within experimental error of previous literature values (18). All of the values for the $\Delta S^{\circ}_{2\times 3}$ parameters are also listed in Table 4.

DISCUSSION

This study is one in a series of studies investigating the thermodynamics of small RNA secondary structure motifs that occur frequently in nature. This study focuses on 2×3 internal loops. A database of 1349 RNA secondary structures was used to determine the frequency of 2×3 nucleotide loops found in naturally occurring RNA (Table 1). The high frequency of occurrence demonstrates the interest in and importance of this particular secondary structure motif. Furthermore, this confirms the importance of generating an updated model for predicting the stability of 2×3 internal loops, which may lead to more accurate prediction of secondary structure from sequence. Although 2×3 nucleotide loops are common in nature, relatively few studies have investigated their thermodynamics (7, 19–21, 38), and most of the loops investigated in these studies were not found in the database compiled (Table 1). As a result, stabilities of many commonly occurring loops are based on the predictive model derived from this dataset of 23 2×3 loops that rarely occur in the database. In this study, 17 frequently occurring 2×3 nucleotide loops were thermodynamically characterized to provide experimental values for frequently occurring loops and to increase the dataset of experimental data from which updated predictive parameters can be derived for those 2×3 loops still without experimental values. It is our hope that the additional experimental values and the updated predictive model can be used to improve secondary structure prediction from sequence.

Database Searching

The database that was searched during this study contained 1349 secondary structures, representing eight different types of RNA. Although this database is not all-inclusive, it provides loop sequences that occur in nature.

Although it was pioneering work, the initial thermodynamic experiments on 2×3 nucleotide internal loops (7, 20, 21, 38) only began to study the possible sequence combinations. After the addition of the data reported here, the percentage of loop sequences with nearest neighbors defined (Table 1, data set 1), loop sequences without the nearest neighbors defined (Table 1, data set 2), nearest neighbors without loop sequences (Table 1, data set 3), and loop sequences classified as purines or pyrimidines (Table 1, data set 4) now studied has increased significantly from the percentages previously studied. It is apparent from the first data set in Table 1 that few of the 2×3 loops found in the database had been thermodynamically characterized previously. In fact, only one of the top 25 most frequently occurring 2×3 loops had been studied. After adding the results reported here, 18 combinations in the top 25 have been thermodynamically characterized. It is interesting to note that the three most frequently occurring loop sequences, $\begin{pmatrix} \text{GCCC} \\ \text{CAAUG} \end{pmatrix}$, $\begin{pmatrix} \text{AAAC} \\ \text{UCACG} \end{pmatrix}$, and $\begin{pmatrix} \text{ACAG} \\ \text{UAUAC} \end{pmatrix}$ do not contain mismatches that have previously been considered stabilizing (G·G, A·G, or U·U) (22, 39). Similar to what was observed for 1×2 nucleotide internal loops (30), the presence or absence of stabilizing mismatches does not determine the frequency of occurrence for 2×3 nucleotide loops.

The second set of data in Table 1 shows that 135 loop sequences were found in this database. It is not immediately obvious why these loop sequences are found in the database while the other 66% of possible loop sequences are not. Perhaps there is a structural explanation for this trend; but an extensive structural study would need to be done. Or, these loops may be found in secondary structures that were not included in this database.

The third set of data in Table 1 shows that 33 of the 36 possible nearest neighbor combinations were found in the database. $\begin{pmatrix} \text{GXXC} \\ \text{CXXXG} \end{pmatrix}$ is the most frequent nearest neighbor combination, representing about 16% of the total number of loops, followed closely by

($\begin{smallmatrix} \text{GXXG} \\ \text{CXXXC} \end{smallmatrix}$), also representing about 16% of the total number of loops. As shown here, when flipping the 3' adjacent pair from C-G to G-C, there is relatively no difference in the number of loops that have either combination. However, if the 5' adjacent pair of ($\begin{smallmatrix} \text{GXXC} \\ \text{CXXXG} \end{smallmatrix}$) is flipped from G-C to C-G, the adjacent base pair combination goes from being the most frequent to the 17th most frequent. There are three possible nearest neighbor combinations that were not found in the database, ($\begin{smallmatrix} \text{GXXG} \\ \text{UXXXU} \end{smallmatrix}$), ($\begin{smallmatrix} \text{GXXU} \\ \text{UXXXG} \end{smallmatrix}$), and ($\begin{smallmatrix} \text{UXXU} \\ \text{AXXXG} \end{smallmatrix}$), all of which contain a G-U or U-G pair.

Thermodynamic Contribution of 2×3 Nucleotide Loops to Duplex Thermodynamics

The thermodynamic contribution of a 2×3 loop to duplex thermodynamics is varied. For example, loop contributions to enthalpy, entropy, and free energy changes of the duplex range from −26.9 to 23.0 kcal/mol, −89.2 to 64.3 cal/(K·mol), and 0.2 to 5.2 kcal/mol, respectively (Table 3). The frequency in which a loop was found in the database does not correlate with the stability of the loop. For example, two of the three most stable loops were not found in the database; the remaining was the 32nd most common in the database.

As expected, all of the internal loops studied here contribute unfavorably to the stability of the duplex (Table 3). The nine most stable loops have sequences that could possibly form G·A or U·U pairs, which have been previously shown to stabilize mismatches and loops (30, 39–44). However, some loops with possible G·A or U·U pairs are actually less stable than those with no possible G·A or U·U pairs, showing that free energy contributions are not completely influenced by the possible pairs that could form.

There does seem to be a correlation between the number of G-C or C-G pairs directly adjacent to the loop and the free energy contribution of the loop to duplex stability. For example, the 23 loops with two adjacent G-C or C-G pairs contribute an average of 2.2 kcal/mol to duplex stability. The 12 loops with only one G-C or C-G adjacent base pair and the 11 loops with no adjacent G-C or C-G base pairs contribute an average of 2.8 and 3.1 kcal/mol to duplex stability, respectively.

Derivation of an Updated Model to Predict Loop Thermodynamics

RNAstructure currently uses a predictive model to approximate the free energy contribution of all 2×3 nucleotide loops (see eq 1). Since the dataset of available 2×3 nucleotide loop thermodynamics increased with the addition of the data reported here, this larger dataset was used to derive an updated model to predict 2×3 thermodynamics. Several models were generated by using different combinations of parameters (data not shown). The model that agreed best with the experimental data, resulted in linearly independent parameters, and had small standard deviations was one using parameters similar to the current *RNAstructure* model (eq 1). In fact, some of the new parameters are within experimental error of the current *RNAstructure* model.

In the current *RNAstructure* model, $\Delta G^{\circ}_{37, \text{loop initiation}}$ is 2.0 kcal/mol and was derived from a dataset of 23 2×3 nucleotide loops (18, 22). $\Delta G^{\circ}_{37, \text{asym}}$, a penalty of 0.6 kcal/mol for being an asymmetric internal loop, was derived from data of loops of various sizes (18). Resulting from a linear regression analysis with the larger dataset of 2×3 thermodynamics, a new parameter was derived, $\Delta G^{\circ}_{37, 2 \times 3 \text{ initiation}}$. This value, 2.5 ± 0.1 kcal/mol, is within experimental error of the sum of $\Delta G^{\circ}_{37, \text{loop initiation}}$ and $\Delta G^{\circ}_{37, \text{asym}}$ (Table 4), which are used by the current *RNAstructure* model.

In the current *RNAstructure* model, there is a penalty ($\Delta G^{\circ}_{37, \text{AU/GU}}$) for replacing a closing G-C or C-G pair with a closing A-U, U-A, G-U, or U-G pair. This value was derived from data of loops of various sizes (18). From the new dataset, linear regression was used to

derive a $\Delta G^{\circ}_{37, \text{AU/GU}}$ term based only on the 2×3 nucleotide loop data. This value, 0.9 ± 0.1 kcal/mol, is within experimental error of *RNAstructure*'s current penalty, 0.7 ± 0.1 kcal/mol (18).

One difference between the *RNAstructure* parameters and the parameters proposed here is the term for a 5'YA3'/3'RG5' bonus for the first/last mismatch. *RNAstructure* assigns a bonus of -0.5 ± 0.2 kcal/mol to this stack. Similarly, the Chen and Turner model (19) assigns a bonus of -0.4 ± 0.2 kcal/mol to this stack. With the dataset used here, this stack did not contribute to 2×3 nucleotide loop stability. The 5'RG3'/3'YA5' bonus for the first/last mismatch derived here (-1.4 ± 0.2 kcal/mol) was within experimental error with the same parameter used by *RNAstructure* (-1.2 ± 0.1 kcal/mol). Lastly, the bonus for a 5'YG3'/3'RA5' first/last mismatch derived here (-0.7 ± 0.1 kcal/mol) is less than the bonus currently used by *RNAstructure* (-1.1 ± 0.1 kcal/mol).

The final parameters used by the current model in *RNAstructure* are bonuses for stabilizing mismatches in the loop (18). *RNAstructure* includes a -0.7 ± 0.2 kcal/mol bonus for potential G-G pairs and a -0.4 ± 0.1 kcal/mol bonus for potential U-U pairs. The model proposed here contains a G-G bonus of -0.8 ± 0.3 kcal/mol, which is within experimental error of the *RNAstructure* value, and a more favorable U-U bonus of -0.9 ± 0.2 kcal/mol.

The values for the enthalpy parameters derived here can also be compared to the enthalpy parameters published previously (18). Except for the omission of the 5'YA3'/3'RG5' parameter in the model proposed here, the values for all of the other parameters are within experimental error of the previous values (Table 4). Although the size of the 2×3 loop dataset was increased, we were unable to derive a predictive model that was much different from the one currently used by *RNAstructure*. Although both the predictive model used currently by *RNAstructure* and the updated predictive model derived here work quite well for the dataset of 2×3 internal loops, adding a look-up table with experimental values for 2×3 nucleotide loops may improve free energy calculations and prediction of secondary structure from sequence.

Acknowledgments

This work was supported by National Institutes of Health grant R15GM08569 from the National Institute of General Medical Sciences.

Abbreviations

R	purine nucleotides
Y	pyrimidine nucleotides

REFERENCES

1. 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]
2. Ippolito JA, Steitz TA. The structure of the HIV-1 RRE high affinity Rev binding site at 1.6 angstrom resolution. *J. Mol. Biol.* 2000; 295:711–717. [PubMed: 10656783]
3. Battiste JL, Mao H, Rao NS, Tan R, Muhandiram DR, Kay LE, Frankel AD, Williamson JR. Alpha helix-RNA major groove recognition in an HIV-1 Rev peptide RRE RNA complex. *Science.* 1996; 273:1547–1551. [PubMed: 8703216]
4. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science.* 2000; 289:905–920. [PubMed: 10937989]

5. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*. 2000; 407:340–348. [PubMed: 11014183]
6. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science*. 1996; 273:1678–1685. [PubMed: 8781224]
7. Schroeder SJ, Fountain MA, Kennedy SD, Lukavsky PJ, Puglisi JD, Krugh TR, Turner DH. Thermodynamic stability and structural features of the J4/5 loop in a *Pneumocystis carinii* group I intron. *Biochemistry*. 2003; 42:14184–14196. [PubMed: 14640686]
8. Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*. 2004; 432:411–415. [PubMed: 15549109]
9. Gutell RR, Gray MW, Schnare MN. A compilation of large subunit (23S and 23S-like) ribosomal-RNA structures: 1993. *Nucleic Acids Res*. 1993; 21:3055–3074. [PubMed: 8332527]
10. Schnare MN, Damberger SH, Gray MW, Gutell RR. Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA. *J. Mol. Biol*. 1996; 256:701–719. [PubMed: 8642592]
11. Gutell RR. Collection of small-subunit (16S- and 16S-like) ribosomal-RNA structures: 1994. *Nucleic Acids Res*. 1994; 22:3502–3507. [PubMed: 7524024]
12. Waring RB, Davies RW. Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing - A review. *Gene*. 1984; 28:277–291. [PubMed: 6086458]
13. Damberger SH, Gutell RR. A comparative database of group I intron structures. *Nucleic Acids Res*. 1994; 22:3508–3510. [PubMed: 7937050]
14. Larsen N, Samuelsson T, Zwieb C. The signal recognition particle database (SRPDB). *Nucleic Acids Res*. 1998; 26:177–178. [PubMed: 9399828]
15. Kondo J, Urzhumtsev A, Westhof E. Two conformational states in the crystal structure of the Homo sapiens cytoplasmic ribosomal decoding A site. *Nucleic Acids Res*. 2006; 34:676–685. [PubMed: 16452297]
16. Gait MJ, Karn J. RNA recognition by the human immuno-deficiency virus Tat and Rev proteins. *Trends Biochem. Sci*. 1993; 18:255–259. [PubMed: 8212135]
17. Gait MJ, Karn J. Progress in anti-HIV structure-based drug design. *Trends Biotechnol*. 1995; 13:430–438. [PubMed: 7546568]
18. 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]
19. Chen G, Turner DH. Consecutive GA pairs stabilize medium-size RNA internal loops. *Biochemistry*. 2006; 45:4025–4043. [PubMed: 16548530]
20. Schroeder SJ, Turner DH. Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry*. 2000; 39:9257–9274. [PubMed: 10924119]
21. Peritz AE, Kierzek R, Sugimoto N, Turner DH. Thermodynamic study of internal loops in oligoribonucleotides: Symmetric loops are more stable than asymmetric loops. *Biochemistry*. 1991; 30:6428–6436. [PubMed: 1711369]
22. Mathews DH, Disney MD, Childs JC, Schroeder SJ, Zuker M, Turner DH. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. U.S.A*. 2004; 101:7287–7292. [PubMed: 15123812]
23. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res*. 2003; 31:3429–3431. [PubMed: 12824340]
24. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003; 31:3406–3415. [PubMed: 12824337]
25. Markham NR, Zuker M. UNAFold: Software for nucleic acid folding and hybridization. *Methods Mol. Biol*. 2008; 453:3–31. [PubMed: 18712296]
26. Andronescu M, Condon A, Hoos HH, Mathews DH, Murphy KP. Efficient parameter estimation for RNA secondary structure prediction. *Bioinformatics*. 2007; 23:i19–i28. [PubMed: 17646296]

27. Davis AR, Znosko BM. Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry*. 2007; 46:13425–13436. [PubMed: 17958380]
28. Davis AR, Znosko BM. Thermodynamic characterization of naturally occurring RNA single mismatches with G-U nearest neighbors. *Biochemistry*. 2008; 47:10178–10187. [PubMed: 18754680]
29. Christiansen ME, Znosko BM. Thermodynamic characterization of tandem mismatches found in naturally occurring RNA. *Nucleic Acids Res*. 2009; 37:4696–4706. [PubMed: 19509311]
30. Badhwar J, Karri S, Cass CK, Wunderlich EL, Znosko BM. Thermodynamic characterization of RNA duplexes containing naturally occurring 1×2 nucleotide internal loops. *Biochemistry*. 2007; 46:14715–14724. [PubMed: 18020450]
31. Thulasi P, Pandya LK, Znosko BM. Thermodynamic characterization of RNA triloops. *Biochemistry*. 2010; 49:9058–9062. [PubMed: 20843054]
32. Sheehy JP, Davis AR, Znosko BM. Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA*. 2010; 16:417–429. [PubMed: 20047989]
33. Stawinski J, Stromberg R, Thelin M, Westman E. Evaluation of the use of the tert-butyltrimethylsilyl group for 2'-protection in RNA-synthesis via the H-phosphonate approach. *Nucleosides Nucleotides*. 1988; 7:779–782.
34. Chou SH, Flynn P, Reid B. Solid-phase synthesis and high-resolution NMR-studies of two synthetic double-helical RNA dodecamers – r(CGCGAAUUCGCG) and r(CGCGUAUACGCG). *Biochemistry*. 1989; 28:2422–2435. [PubMed: 2471547]
35. 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]
36. McDowell JA, Turner DH. Investigation of the structural basis for thermodynamic stabilities of tandem GU mismatches: Solution structure of (rGAGGUCUC)₂ by two-dimensional NMR and simulated annealing. *Biochemistry*. 1996; 35:14077–14089. [PubMed: 8916893]
37. Xia T, SantaLucia J Jr, 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]
38. Schroeder SJ, Turner DH. Thermodynamic stabilities of internal loops with GU closing pairs in RNA. *Biochemistry*. 2001; 40:11509–11517. [PubMed: 11560499]
39. Schroeder S, Kim J, Turner DH. G-A and U-U mismatches can stabilize RNA internal loops of three nucleotides. *Biochemistry*. 1996; 35:16105–16109. [PubMed: 8973181]
40. Walter AE, Wu M, Turner DH. The stability and structure of tandem GA mismatches in RNA depend on closing base pairs. *Biochemistry*. 1994; 33:11349–11354. [PubMed: 7537087]
41. SantaLucia J Jr, Kierzek R, Turner DH. 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*. 1991; 30:8242–8251. [PubMed: 1714301]
42. SantaLucia J, Kierzek R, Turner DH. Functional-group substitutions as probes of hydrogen-bonding between GA mismatches in RNA internal loops. *J. Am. Chem. Soc*. 1991; 113:4313–4322.
43. SantaLucia J Jr, Turner DH. Structure of (rGGCGAGCC)₂ in solution from NMR and restrained molecular dynamics. *Biochemistry*. 1993; 32:12612–12623. [PubMed: 8251479]
44. Wu M, McDowell JA, Turner DH. A periodic table of symmetric tandem mismatches in RNA. *Biochemistry*. 1995; 34:3204–3211. [PubMed: 7533535]

Figure 1.

Secondary structures of naturally occurring 2×3 nucleotide internal loops. These loops are found in a) the human immunodeficiency virus type 1 (HIV-1) Rev response element (2, 3), b) the large ribosomal subunit from *Haloarcula marismortui* (4), c) the *Thermus thermophilus* 30S ribosomal subunit (5), d) the P4-P6 domain of the *Tetrahymena thermophila* group I intron (6), e) the P4-P6 domain of the *Pneumocystis carinii* group I intron (7), f) the purine-binding domain of the guanine riboswitch from the xpt-pbuX operon of *Bacillus subtilis* (8), and g) the *Homo sapiens* 18S ribosomal decoding A-site (15).

Table 1

Summary of Database Search for 2×3 Nucleotide Loops^a

Loop with Nearest Neighbors					5' and 3' Adjacent Base Pairs					Loop Nucleotides Classified as Purine or Pyrimidine				
Loop	Freq ^b	% ^c	Ref ^d		Closing bp	Freq ^b	% ^c	Ref ^d		Loop	Freq ^b	% ^c	Ref ^d	
G CC C	90	13.5	<i>e</i>		G C	105	15.8	<i>e</i>		RR	127	19.1	<i>e-i</i>	
C AAU G					C G					RRR				
A AA C	39	5.9	<i>e</i>		G G	104	15.6	<i>e,g,i</i>		YY	91	13.7	<i>e,g</i>	
U CAC G					C C					RRY				
A CA G	31	4.7	<i>e</i>		A G	60	9.0	<i>e,g</i>		RY	89	13.4	<i>e-f</i>	
U AUA C					U C					RRY				
G AU G	18	2.7	<i>e</i>		A C	48	7.2	<i>e</i>		RR	48	7.2	<i>e</i>	
C CAU C					U G					RYR				
G AU G	18	2.7	<i>e</i>		C G	43	6.5	<i>e,g,h</i>		RR	42	6.3	<i>e</i>	
C CGU C					G C					YRY				
U AU C	17	2.6	<i>e</i>		U C	43	6.5	<i>e,f</i>		RY	42	6.3	<i>e</i>	
G GGU G					G G					YRY				
U AA A	16	2.4	<i>e</i>		A U	28	4.2	<i>e</i>		YR	41	6.2		
A GAC U					U U					YRR				
C AA G	13	2.0			U G	28	4.2	<i>e</i>		YR	35	5.3	<i>e</i>	
G GGG C					G U					RYR				
U GG G	13	2.0	<i>e</i>		U G	27	4.1	<i>e-f</i>		RR	23	3.5	<i>e</i>	
G GGA U					G C					RRY				
U AU G	12	1.8	<i>f</i>		C U	24	3.6	<i>e</i>		YY	15	2.3	<i>e</i>	
G AAU C					G A					YRR				
G GA G	11	1.7	<i>e</i>		C G	18	2.7			YY	15	2.3		
C AUA C					G U					YRY				
G GA G	10	1.5	<i>e</i>		U A	18	2.7	<i>e</i>		YY	15	2.3	<i>g</i>	
C AAA C					A U					YYY				
C GC U	10	1.5	<i>e</i>		G A	15	2.3			RY	11	1.7	<i>e,g</i>	
G AAA A					C U					RRR				
U AU C	9	1.4	<i>e</i>		G U	12	1.8			YR	10	1.5		
G AAU G					C U					RRR				
A AC A	7	1.1	<i>e</i>		U C	11	1.7			RR	9	1.4		
U GAU U					A G					YYR				
A CC A	6	0.9	<i>e</i>		A U	10	1.5			RR	7	1.1		
U UAA U					U A					RYR				
A CC G	6	0.9	<i>e</i>		C C	8	1.2	<i>g</i>		YR	7	1.1		
U UAA C					G G					YRY				

Loop with Nearest Neighbors					Loop			5' and 3' Adjacent Base Pairs				Loop Nucleotides Classified as Purine or Pyrimidine				
Loop	Freq ^b	% ^c	Ref ^d		Loop	Freq ^b	% ^c	Ref ^d	Closing bp	Freq ^b	% ^c	Ref ^d	Loop	Freq ^b	% ^c	Ref ^d
U G	GG AGA	G U	6 0.9	<i>e</i>	GC AAA	11	1.7	<i>e,g</i>	U A	8 A	1.2	<i>g</i>	YR YYR	7	1.1	
U G	GG AGA	G C	5 0.8	<i>e</i>	GA AAG	9	1.4	<i>f,g,h</i>	U G	8 A	1.2		RR YRR	6	0.9	<i>g</i>
G C	AA GAA	G C	5 0.8		AC GAC	8	1.2		A U	6 U	0.9		RY RYY	5	0.8	
U G	GG AGA	U A	5 0.8		GA ACG	8	1.2		C G	6 U	0.9		RY RYY	4	0.6	<i>g</i>
G C	GU AAU	G C	5 0.8		AA GAA	7	1.1	<i>g</i>	U A	6 C	0.9		RY YYY	4	0.6	
C G	UA UAG	G U	5 0.8		CA AAG	7	1.1		U G	5 U	0.8	<i>g</i>	YY YYR	4	0.6	
G C	UA UAG	G U	5 0.8		GU AAU	7	1.1		G C	4 A	0.6		RR YYY	3	0.5	
C G	AA GGG	G U	4 0.6		AU AGU	6	0.9		U A	4 U	0.6		YY RYY	3	0.5	
Previously ^j New Total ^k	17 329	2.6 49.4			Previously ^j New Total ^k	198 424	29.7 63.7		Previously ^j New Total ^k	288 544	43.2 81.7		Previously ^j New Total ^k	343 548	51.5 82.3	

^aAs described in the text, not all combinations are shown due to space limitations.

^bFrequency of occurrence in database.

^cPercent out of 666 loops, the total number of loops found in the database.

^dReference where data are reported.

^eThis work.

^fRef 38.

^gRef 20.

^hRef 7

ⁱRef 21.

^jThe total number of 2×3 loops accounted for and the corresponding percentage of the total number of loops found in the database by previous studies.

k_y The total number of 2×3 loops accounted for and the corresponding percentage of the total number of loops found in the database when the data from previous studies were combined with the data reported here.

Table 2

Thermodynamic Parameters for Duplex Formation^a

Frequency ^b	Sequence	Analysis of Melt Curve Fit/Errors				Analysis of T _m Dependence/ Errors (In Plot)			
		-ΔH° (kcal/mol)	-ΔS° (cal/K·mol)	-ΔG° ₃₇ (kcal/mol)	T _m ^c (°C)	-ΔH° (kcal/mol)	-ΔS° (cal/K·mol)	-ΔG° ₃₇ (kcal/mol)	T _m ^c (°C)
90	GACGCCCCUG CUGCAAUUGGAC	69.4 ± 6.1	200.7 ± 19.7	7.21 ± 0.08	40.1	73.2 ± 4.1	213.0 ± 13.3	7.17 ± 0.06	39.7
39	GUCAAAACAG CAGUCACGGUC	77.1 ± 5.6	227.0 ± 18.0	6.72 ± 0.12	37.8	77.2 ± 7.2	227.3 ± 23.3	6.70 ± 0.15	37.7
31	GUCACAGCAG CAGUAUACGUC	58.0 ± 6.0	165.7 ± 19.2	6.63 ± 0.22	37.5	53.5 ± 8.2	151.0 ± 26.8	6.66 ± 0.42	37.8
18	GACGAUGCUG CUGCCAUCGAC	84.5 ± 6.9	243.3 ± 22.1	9.02 ± 0.19	46.4	87.5 ± 6.6	253.0 ± 20.9	9.06 ± 0.19	46.2
18	GACGAUGCAG CUGCCGUCGUC	79.2 ± 7.1	226.6 ± 22.3	8.90 ± 0.23	46.6	79.1 ± 3.1	226.6 ± 9.9	8.88 ± 0.08	46.5
17	GACUAUCCUG CUGGGGUGGAC	81.8 ± 8.3	238.6 ± 26.0	7.74 ± 0.37	41.7	86.8 ± 13.1	255.0 ± 41.9	7.75 ± 0.44	41.4
16	GACUAAACUG CUGAGACUGAC	(62.3)	(180.9)	(6.22)	(35.4)	(45.4)	(126.2)	(6.27)	(35.2)
13	GACUGGGCUG CUGGGGAUGAC	46.7 ± 12.5	129.9 ± 39.7	6.43 ± 0.30	36.3	42.9 ± 6.0	117.9 ± 19.5	6.36 ± 0.35	35.8
12	CUGUAUGACG ^d GACGAAUCUGC	69.4 ± 3.3	202.2 ± 10.9	6.6 ± 0.1	37.5	77.7 ± 2.2	229.3 ± 7.0	6.6 ± 0.1	37.2
11	GACGGAGCUG CUGCAUACGAC	(24.5)	(53.5)	(7.92)	(55.7)	(23.5)	(50.7)	(7.82)	(54.9)
10	GACGGAGCUG CUGCAAAACGAC	82.1 ± 2.1	235.6 ± 6.6	9.06 ± 0.12	46.9	83.3 ± 4.5	239.2 ± 14.2	9.09 ± 0.12	46.8
10	GACCGCUCUG CUGGAAAAGAC	66.4 ± 4.3	190.4 ± 13.7	7.37 ± 0.12	41.0	67.7 ± 3.4	194.5 ± 10.8	7.36 ± 0.05	40.9
9	GACUAUCCUG CUGGAAUUGGAC	80.2 ± 5.1	234.9 ± 16.3	7.36 ± 0.10	40.3	79.3 ± 2.1	232.0 ± 6.7	7.34 ± 0.02	40.2
7	GACAAACACAG CUGUGAUUGUC	67.5 ± 4.2	201.0 ± 14.3	5.15 ± 0.27	30.8	67.1 ± 5.2	199.8 ± 17.2	5.14 ± 0.17	30.7
6	GACACCCACUG CUGUUAUUGAC	75.4 ± 5.4	230.9 ± 18.0	3.81 ± 0.26	26.2	76.0 ± 3.4	232.6 ± 11.3	3.81 ± 0.14	26.3
6	GUCACCCGUG CAGUUAACGAC	62.4 ± 3.6	181.4 ± 12.0	6.11 ± 0.10	34.9	62.4 ± 1.2	181.4 ± 3.8	6.11 ± 0.02	34.9

Frequency ^b	Sequence	Analysis of Melt Curve Fit/Errors				Analysis of T _m Dependence/ Errors (ln Plot)			
		-ΔH° (kcal/mol)	-ΔS° (cal/K·mol)	-ΔG° ₃₇ (kcal/mol)	T _m ^c (°C)	-ΔH° (kcal/mol)	-ΔS° (cal/K·mol)	-ΔG° ₃₇ (kcal/mol)	T _m ^c (°C)
6	GACUGGGCUG CUGGAGAUGAC	75.9 ± 8.2	228.3 ± 26.5	5.04 ± 0.20	31.0	74.7 ± 9.3	224.3 ± 30.8	5.10 ± 0.38	31.2
5	GACUGGGCUG CUGGAGACGAC	58.5 ± 11.3	167.8 ± 36.6	6.47 ± 0.59	36.7	59.9 ± 10.4	172.1 ± 33.5	6.46 ± 0.58	36.6

^aMeasurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na₂EDTA, pH 7.0. Values in parentheses are approximate due to non-two-state melts and/or large errors associated with the derived thermodynamic parameters.

^bFrequency of occurrence in the database described in Materials and Methods.

^cCalculated at 10⁻⁴ M oligomer concentration.

^dRef 38.

Table 3

Loop Contribution to Duplex Thermodynamics^a

Frequency ^b	Sequence	$\Delta H^{\circ}_{2 \times 3}$ (kcal/mol)	$\Delta S^{\circ}_{2 \times 3}$ (cal/K·mol)	$\Delta G^{\circ}_{37.2 \times 3}$ (kcal/mol)	Frequency ^b	Sequence	$\Delta H^{\circ}_{2 \times 3}$ (kcal/mol)	$\Delta S^{\circ}_{2 \times 3}$ (cal/K·mol)	$\Delta G^{\circ}_{37.2 \times 3}$ (kcal/mol)
90	GACGCCCCUG CUGCAAUGGAC	-8.0 -6.7	-36.1 -30.0	3.14 2.5	0	UGACUUCUCA ^c ACUGUUUGAGU	-17.7 -26.1	-59.9 -87.0	0.89 0.7
39	GUCAAACACAG CAGUCACGGUC	-12.2 -0.4	-50.2 -12.6	3.36 3.4	0	GAGCAGCGAC ^c CUCGGAAGCUG	-10.6 -6.7	-40.4 -30.0	1.91 2.5
31	GUCACACGAC CAGUAUACGUC	13.0 -0.4	-30.3 -12.6	3.56 3.4	0	GAGCGACGAC ^c CUCGAAAGCUG	-6.4 -11.6	-27.1 -43.4	1.94 1.8
18	GACGAUCCUG CUGCCAUCGAC	-20.8 -16.4	-71.9 -58.5	1.41 1.6	0	CCACGGCUC ^c GGUGAAAGAGG	-10.3 -11.6	-39.5 -43.4	1.99 1.8
18	GACGAUCCUG CUGCCGUCGUC	-12.4 -16.4	-45.5 -58.5	1.59 1.6	0	GAGCAACGAC ^c CUCGGAAGCUG	-9.4 -6.7	-37.2 -30.0	2.10 2.5
17	GACUAUCCUG CUGGGGUGGAC	-20.2 -10.1	-72.6 -41.1	2.31 2.5	0	UGACUUCUCA ^c ACUGCUUGAGU	-16.8 -16.4	-58.0 -58.5	1.26 1.6
13	GACUGGGCUG UGGGGAUGAC	23.0 10.5	64.3 22.0	2.95 3.5	0	GAGCAGCGAC ^c CUCGAAAGCUG	-8.7 -2.1	-35.5 -12.8	2.21 1.7
12	CUGUAUAGAC ^c GACGAAUUCUG	-13.3 -10.1	-52.6 -41.1	2.96 2.5	0	UCACUUCUGA ^c AGUGCUCGACU	-12.8 -6.7	-50.5 -30.0	2.90 2.5
10	GACGGAGCUG CUGCAAAACGAC	-16.6 -16.5	-58.1 -57.1	1.38 1.1	0	UCAGCCGUGA ^c AGUCAAUACAU	1.0 -6.7	-6.8 -30.0	3.14 2.5
10	GACCGCUCUG CUGGAAAAGAC	-0.7 -5.3	-11.8 -26.0	2.94 2.7	0	UGAGAAAGUCA ^e ACUCAAAACAGU	-3.9 -6.7	-17.8 -30.0	1.66 2.5
9	GACUAUCCUG CUGGAAUUGGAC	-12.7 -10.1	-49.6 -41.1	2.72 2.5	0	UGAGAAAGUCA ^c ACUCCGACAGU	-4.2 -6.7	-18.9 -30.0	1.67 2.5
7	GACAAACACAG CUGGAUUGUC	-4.1 5.9	-25.9 4.8	3.90 4.3	0	GAGCAGCGAC ^c CUCGAAAGCUG	-8.8 -6.7	-36.8 -30.0	2.57 2.5
6	GACACCACUG CUGUUAUUGAC	-13.0 5.9	-58.7 4.8	5.23 4.3	0	GAGCAACGAC ^c CUCGAAAGCUG	-2.1 -6.7	-15.8 -30.0	2.72 2.5
6	GUCACCGCUG CAGUUAACGAC	4.1 -0.4	-0.1 -12.6	4.11 3.4	0	UGACUUCUCA ^c ACUGCCUGAGU	-17.1 -16.4	-61.4 -58.5	1.93 1.6
6	GACUGGGCUG CUGGAGAUGAC	-8.9 1.0	-42.1 -8.6	4.21 3.6	0	GAGUAAACGAC ^f CUCGAAAGCUG	-17.3 -10.2	-61.9 -39.7	1.80 2.0

Frequency ^b	Sequence	ΔH_{2-3}^c (kcal/mol)	ΔS_{2-3}^c (cal/K·mol)	$\Delta G_{37,2-3}^c$ (kcal/mol)	Frequency ^b	Sequence	ΔH_{2-3}^c (kcal/mol)	ΔS_{2-3}^c (cal/K·mol)	$\Delta G_{37,2-3}^c$ (kcal/mol)
5	GACUGGGCUG CUGGAGACGAC	8.2 -5.3	14.5 -26.0	3.76 2.7	0	GAGUGAUGAC ^f CUCGAAAGGCUG	-14.3 -8.8	-52.1 -35.7	1.85 2.2
3	CGACGAGCAG ^d GCUGAAGCGUC	-26.9 -16.5	-89.2 -56.8	0.77 1.1	0	GAGCAAUGAC ^f CUCGAAAGGCUG	-15.9 -10.2	-58.7 -39.7	2.26 2.0
2	CUGUGGACGA ^e GACGAGUUCU	1.2 1.0	-8.0 -8.6	3.66 3.6	0	GAGUAAACGAC ^f CUCGAAAGGCUG	-4.7 -0.4	-25.6 -12.6	3.20 3.4
0	GAGCGACGAC ^e CUCGAAAGGCUG	-18.0 -21.4	-59.0 -70.5	0.20 0.4	0	GAGUAAUUGAC ^f CUCGAAAGGCUG	0.6 -3.9	-8.3 -22.3	3.25 2.9
0	CCACGGCUC ^e GGUGAGAGAG	-14.0 -11.6	-49.5 -43.4	1.37 1.8	0	GAGUGAUGAC ^f CUCGAAAGGCUG	11.4 1.0	25.7 -8.6	3.35 3.6
0	GAGCAAACGAC ^e CUCGAAAGGCUG	-6.0 -16.5	-24.2 -57.1	1.48 1.1	0	GAGCAAUUGAC ^f CUCGAAAGGCUG	-0.5 -0.4	-13.1 -12.6	3.56 3.4
0	CGACGAGCAG ^e GCUGGAACGUC	-8.0 -2.1	-32.4 -12.8	2.04 1.7	0	GAGUAGUGAC ^f CUCGAAAGGCUG	15.3 5.9	35.9 4.8	4.05 4.3
0	CCUCUGCGGUGA ^e GGAGAAACCGC	-5.0 -5.3	-26.4 -26.0	3.18 2.7	0	GAGUAAUUGAC ^f CUCGAAAGGCUG	17.9 5.9	44.0 4.8	4.15 4.3

^aValues in italics are predicted values based on the model derived here.

^bFrequency of occurrence in the database described in Materials and Methods.

^cDerived from raw data available in Ref 20.

^dDerived from raw data available in Ref 7.

^eDerived from raw data available in Ref 21.

^fDerived from raw data available in Ref 38.

Table 4**2×3 Nucleotide Loop Nearest Neighbor Parameters at 37 °C^a**

	$\Delta H^\circ_{2\times 3}$ (kcal/mol)	$\Delta S^\circ_{2\times 3}$ (cal/K·mol)	$\Delta G^\circ_{37,2\times 3}$ (kcal/mol)
2×3 Loop Initiation ^a	-6.7 ± 2.1 (-3.6 ± 1.6) ^b	-30.0 ± 6.7	2.5 ± 0.1 (2.6 ± 0.1) ^b (2.6 ± 0.2) ^c
A-U or G-U Closure ^d	6.3 ± 1.5 (5.0 ± 0.7) ^b	17.4 ± 4.8	0.9 ± 0.1 (0.7 ± 0.1) ^b (0.7 ± 0.1) ^c
5'YA ^e 3'RG	n/a (-5.7 ± 3.8) ^b	n/a	n/a (-0.5 ± 0.2) ^b (-0.4 ± 0.2) ^c
5'RG ^e 3'YA	-9.8 ± 3.3 (-10.9 ± 2.7) ^b	-27.1 ± 10.7	-1.4 ± 0.2 (-1.2 ± 0.1) ^b (-1.1 ± 0.2) ^c
5'YG ^e 3'RA	-4.9 ± 2.5 (-8.6 ± 1.9) ^b	-13.4 ± 7.9	-0.7 ± 0.1 (-1.1 ± 0.1) ^b (-1.4 ± 0.1) ^c
G·G Mismatch ^f	-4.6 ± 4.9 (-9.0 ± 4.6) ^b	17.2 ± 15.7	-0.8 ± 0.3 (-0.7 ± 0.2) ^b (-0.7 ± 0.3) ^c
U·U Mismatch ^f	-9.7 ± 2.8 (-6.4 ± 2.5) ^b	-28.5 ± 9.1	-0.9 ± 0.2 (-0.4 ± 0.1) ^b (-0.3 ± 0.2) ^c

^aThese parameters apply to all 2×3 loops, regardless of sequence.^bNumbers in parentheses are values proposed in Ref 18. The initiation value listed here is the sum of the initiation value and asymmetry penalty from the original reference.^cNumbers in parentheses are values proposed in Ref 19. The initiation value listed here is the sum of the initiation value and asymmetry penalty from the original reference.^dThese parameters are applied per A-U or G-U closure.^eThese parameters are applied per each closing pair-first mismatch or closing pair-last mismatch combination.^fThese parameters are applied per each first/last mismatch.