Thermodynamic Characterization of Single Mismatches Found in Naturally Occurring RNA[†]

Amber R. Davis and Brent M. Znosko*

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

Received July 3, 2007; Revised Manuscript Received August 15, 2007

ABSTRACT: Many naturally occurring RNA structures contain single mismatches. However, the algorithms currently used to predict RNA structure from sequence rely on a minimal set of data for single mismatches, most of which occur rather infrequently in nature. As a result, several approximations and assumptions are used to predict the stability of RNA duplexes containing the most common single mismatches. Therefore, the relative frequency of single mismatches was determined by compiling and searching a database of 955 RNA secondary structures. Thermodynamic parameters for duplex formation, derived from optical melting experiments, are reported for 28 oligoribonucleotides containing frequently occurring single mismatches. These data were then combined with previous data to construct a dataset of 64 single mismatches, including the 30 most common in the database. Because of this increase in experimental thermodynamic parameters for single mismatches that occur frequently in nature, more accurate free energy calculations have resulted. To improve the prediction of the thermodynamic parameters for duplexes containing single mismatches that have not been experimentally measured, single mismatch-specific nearest neighbor parameters were derived. The free energy of an RNA duplex containing a single mismatch that has not been thermodynamically characterized can be calculated by: $\Delta G^\circ_{37,\text{single mismatch}} = \Delta G^\circ_{37,\text{mismatch nt}} + \Delta G^\circ_{37,\text{mismatch−NN interaction}} + \Delta G^\circ_{37,\text{AU/GU}}$. Here, $\Delta G^\circ_{37,\text{mismatch}}$ is −0.4, −2.1, and −0.3 kcal/mol for A•G, G•G, and U•U mismatches, respectively; $\Delta G^\circ_{37,\text{mismatch−NN interaction}}$ is 0.7, −0.5, 0.4, −0.4, and −1.0 kcal/mol for 5′YRR3′/3′RRY5′, 5′RYY3′/3′YYR5′, 5′YYR3′/3′RYY5′, 5′YRY3′/3′RYR5′, and 5′RRY3′/3′YYR5′ mismatch-nearest neighbor combinations, respectively, when A and G are categorized as purines (R) and C and U are categorized as pyrimidines (Y); and $\Delta G^\circ_{37,\text{AU/GU}}$ is a penalty of 1.2 kcal/mol for replacing a G-C base pair with either an A-U or G-U base pair. Similar predictive models were also derived for $\Delta H^\circ_{\text{single mismatch}}$ and $\Delta S^\circ_{\text{single mismatch}}$. These new predictive models, in conjunction with the reported thermodynamics for frequently occurring single mismatches, should allow for more accurate calculations of the free energy of RNA duplexes containing single mismatches and, furthermore, allow for improved prediction of secondary structure from sequence.

Although Watson−Crick pairs are the most common RNA secondary structure motif, many nucleotides in a given RNA are located in motifs other than canonical base pairs, such as internal loops, hairpin loops, and bulges. Single mismatches (or 1 × 1 internal loops) occur when two canonical pairs are separated by a single non-canonical pair. A single mismatch in RNA may occur when there is an error during DNA replication, and this error is passed on to RNA during transcription when RNA polymerase inserts the wrong nucleotide during transcription, or when one of the nucleotides in an RNA canonical pair is mutated or edited. In addition, single mismatches may occur naturally in RNA and may have evolved to serve a particular structural or functional role.

Single mismatches have been found to occur and to serve integral structural and/or functional roles in several types of RNA. For example, single mismatches provide signals for recognition and cleavage by *Escherichia coli* ribonuclease III (*1*, *2*) and are prevalent in and important for the function of miRNA (*3*). A single mismatch is necessary for the recognition and interaction of the influenza A viral promoter sequence with its RNA-dependent RNA polymerase proteins (*4*). Single mismatches have been found to be conserved within several viral genomes, including HIV-1 (*5*), cricket paralysis (*6*), and the turnip yellow mosaic (*7*) viruses and may provide methods to control the replication and translation processes of these viral genomes (*8*). In addition, West Nile, Dengue-3, and Yellow fever viruses were predicted to contain single mismatches which are important for structure and/or function of these viruses (*9*). Trinucleotide repeat expansion diseases contain a series of single mismatches separated by two base pairs and account for approximately 20 neurological diseases, such as Huntington's disease (*10*), Kennedy's disease (*10*), Friedreich's ataxia (*10*), and myotonic dystrophy (*11*). The expansion in the number of

[†] Partial funding for this project was provided by the Saint Louis University College of Arts and Sciences, Saint Louis University Department of Chemistry, a Saint Louis University Summer Research Award (B.M.Z.), and the Saint Louis University Beaumont Faculty Development Fund (B.M.Z.).

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

10.1021/bi701311c CCC: $37.00
Published on Web 10/24/2007

trinucleotide repeats results in a complex, higher order structure that leads to a toxic build-up of the repeat-containing RNAs in the nucleus (*12*). A better understanding of the stability and structure of RNA biomolecules containing single mismatches may aid in the development of useful and effective therapies for several bacterial and viral infections as well as diseases in which single mismatches play an important structural or functional role (*13*). The demand for a quick and accurate method to predict RNA secondary structure from sequence is important to better (1) understand structure−function relationships, (2) understand secondary and tertiary interactions, and (3) design pharmaceutical agents for both bacterial and viral agents that contain various structural motifs.

Mathews et al. (*14*−*16*), Zuker (*17*), and Hofacker (*18*) have developed algorithms to predict RNA secondary structure from sequence. These algorithms have been developed into the computer programs *RNAstructure* (*14*−*16*) and *mfold* (*17*) and into the Vienna RNA software package. These algorithms use the method of free energy minimization to predict secondary structure from sequence. In this method, a sequence of interest is folded into all possible secondary structure conformations. For each conformation, the free energy parameters of all secondary structure motifs (experimental or predicted) in that conformation are added to give a total free energy for that conformation. The total free energies for all of the possible secondary structure conformations are compared. The conformation with the lowest free energy is predicted to be the predominant species in solution. These programs have been quite influential. The original article describing *mfold* (*17*) has been cited over 1190 times (*19*), and the article describing the underlying algorithm (*20*) has been cited 1227 times (*19*). In addition, *RNAstructure* has been downloaded over 12,000 times (Mathews, D. H., personal communication, February 27, 2006), and the original article describing this program (*15*) has been cited over 1390 times (*19*). Also, the lowest free energy structure predicted by *RNAstructure* contains ∼73% of known base pairs (*15*). However, the lack of experimental parameters for naturally occurring, non-Watson−Crick regions, such as single mismatches, is a major limitation of these current algorithms (*14*−*17*).

RNAstructure (*14*−*16*) and *mfold* (*17*) both use two different methods to assign free energy parameters to non-Watson−Crick regions. If a particular motif has been thermodynamically characterized, the experimental free energy parameter is assigned. If a motif has not been thermodynamically characterized, these programs use a predictive model to assign a free energy parameter. Because only a few studies have investigated single mismatches (*21*−*25*), the thermodynamic contribution of most single mismatches is calculated by the following (*14*−*16*):

$$\Delta G^\circ_{37,\text{single mismatch}} = \Delta G^\circ_{37,\text{loop initiation}} + \Delta G^\circ_{37,\text{AU/GU closure}} + \Delta G^\circ_{37,\text{type of loop/first pair}} \quad (1)$$

Here, $\Delta G^\circ_{37,\text{loop initiation}}$ is the free energy of initiation of a single mismatch (0.5 kcal/mol), $\Delta G^\circ_{37,\text{AU/GU closure}}$ is the penalty for replacing a G-C nearest neighbor with either an A-U or G-U nearest neighbor (0.7 kcal/mol), and $\Delta G^\circ_{37,\text{type of loop/first pair}}$ is the bonus for a G•G mismatch (−2.6 kcal/mol) and for a 5′RU/3′YU stack in a single mismatch (−0.4 kcal/mol), where R is A or G, and Y is U or C in an A-U or G-C pair. Data for loops of various sizes, including a minimal set of single mismatch data, were used to derive these parameters. Therefore, an increase in the number and type of single mismatches thermodynamically measured would result in (1) a larger dataset of mismatch thermodynamics in which updated parameters can be derived or (2) a large and diverse enough dataset of single mismatch data to derive single mismatch-specific thermodynamic parameters. Either one of these results would likely increase the accuracy of secondary structure prediction.

Although previous thermodynamic studies have investigated single mismatches (*21*−*25*), experimental values for the majority of the most frequently occurring single mismatches are not available. However, it is likely that the scientists that use *RNAstructure* and *mfold* are interested in the secondary structures of naturally occurring mismatch sequences, such as those involved with the bacteria, viruses, and diseases mentioned previously. Therefore, the thermodynamic contribution of single mismatches that occur most often in nature and the approximation of those single mismatches that occur less often may produce more accurate predictions of secondary structure from sequence. A database of 955 RNA secondary structures was compiled and searched to determine the relative frequencies of single mismatches. The data reported here provide experimental values for the 30 most frequently occurring single mismatches in the database. By using this data and that of previous studies (*21*−*24*), single mismatch-specific nearest neighbor parameters have been derived.

MATERIALS AND METHODS

Compiling and Searching a Database for Single Mismatches. A database of secondary structures that includes 151,503 nucleotides and 43,519 base pairs consisting of 22 small subunit rRNAs (*26*), 5 large subunit rRNAs (*27*, *28*), 309 5S rRNAs (*29*), 484 tRNAs (*30*), 91 signal recognition particles (*31*), 16 RNase P RNAs (*32*), 25 group I introns (*33*, *34*), and 3 group II introns (*35*) was assembled. The database was searched for single mismatches, and the number of occurrences for each type of mismatch was tabulated. In this work, G-U pairs were considered to be canonical base pairs. For example, $\left[{5'\text{CUUG}3' \atop 3'\text{GGUC}5'}\right]$ is considered to be a U•U single mismatch, not a 2 × 2 internal loop.

Design of Sequences for Optical Melting Studies. Sequences of mismatches and nearest neighbors were designed to represent those found most frequently in the database described above. Mismatches were placed in the center of the duplex. All duplexes contained the same stem except for the nearest neighbors adjacent to the mismatch. Duplexes were designed to have melting temperatures between 35 and 55 °C and to minimize the possible formation of either hairpin structures or various undesired duplexes. Moreover, terminal G-C pairs were selected to prevent end fraying during optical melting experiments.

RNA Synthesis and Purification. Oligonucleotides were ordered from the Keck Lab at Yale University (New Haven, CT) or from Azco BioTech, Inc. (San Diego, CA). The synthesis and purification of the oligonucleotides followed standard procedures that were described previously (*36*).

Concentration Calculations and Duplex Formation. Concentrations of the single-stranded oligoribonucleotides were calculated using Beer's Law. The concentration of each individual strand was calculated from an absorbance measured at 280 nm and the single-strand extinction coefficient, which was calculated using *RNACalc* (*37*). To ensure that the absorbance was between 0.2 and 2.0, the samples were diluted. Furthermore, the absorbances of the oligoribonucleotides were measured at 80 °C to disrupt any single-strand folding. Individual single-strand concentrations were used to mix equal molar amounts of non-self-complementary strands to form a duplex containing a single mismatch. It has been shown that small mixing errors of single-stranded, non-self-complementary strands do not appreciably affect the resulting thermodynamic parameters (*23*).

Optical Melting Experiments. Optical melting experiments were performed in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA (pH 7.0). Melting curves (absorbance vs temperature) were obtained using a heating rate of 1 °C/min from 10 to 90 °C on a Beckman-Coulter DU800 spectrometer with a Beckman-Coulter high-performance temperature controller. The absorbance was measured at 280 nm. To allow for a concentration range >50-fold, a melt scheme that consisted of at least nine concentrations was used.

Determination of Thermodynamic Parameters for Duplexes. The obtained melting curves were fit to a two-state model using *Meltwin* (*38*). The two-state model assumes linear sloping baselines and temperature-independent $\Delta H°$ and $\Delta S°$ values (*38*, *39*). Thermodynamic parameters were calculated using T_M values at different concentrations according to Borer et al. (*40*):

$$T_M^{-1} = (2.303R/\Delta H°)\log(C_T/4) + (\Delta S°/\Delta H°) \quad (2)$$

Here, R is the gas constant, 1.987 cal/mol·K. For transitions that conform to the two-state model, $\Delta H°$ values from the two methods generally agree within 10%, which indicates that the two-state model is a good estimate of the transition (*41*, *42*). To calculate the Gibb's free energy change at 37 °C, the following equation was used:

$$\Delta G°_{37} = \Delta H° - (310.15\ \text{K})\,\Delta S° \quad (3)$$

Determination of the Contribution of Single Mismatches to Duplex Thermodynamics. The total free energy change for duplex formation can be approximated by a nearest neighbor model (*43*) that is the sum of energy increments for helix initiation, nearest neighbor interactions between base pairs, and the single mismatch contribution. For example,

$$\Delta G°_{37}\begin{bmatrix}5'\text{GACCUGCUG}3'\\3'\text{CUGGUUGAC}5'\end{bmatrix} = \Delta G°_{37,\text{i}} + \Delta G°_{37}\begin{bmatrix}\text{GA}\\\text{CU}\end{bmatrix} + \Delta G°_{37}\begin{bmatrix}\text{AC}\\\text{UG}\end{bmatrix} + \Delta G°_{37}\begin{bmatrix}\text{CC}\\\text{GG}\end{bmatrix} + \Delta G°_{37,\text{single mismatch}} + \Delta G°_{37}\begin{bmatrix}\text{GC}\\\text{UG}\end{bmatrix} + \Delta G°_{37}\begin{bmatrix}\text{CU}\\\text{GA}\end{bmatrix} + \Delta G°_{37}\begin{bmatrix}\text{UG}\\\text{AC}\end{bmatrix} \quad (4)$$

Here, $\Delta G°_{37,\text{i}}$ is the free energy change for duplex initiation, 4.09 kcal/mol (*43*); $\Delta G°_{37,\text{single mismatch}}$ is the free energy contribution from the single mismatch, and the remainder of the terms are individual nearest neighbor values (*43*). Therefore, rearranging eq 4 can solve for the contribution of the single mismatch to duplex stability:

$$\Delta G°_{37,\text{single mismatch}} = \Delta G°_{37}\begin{bmatrix}5'\text{GACCUGCUG}3'\\3'\text{CUGGUUGAC}5'\end{bmatrix} - \Delta G°_{37,\text{i}} - \Delta G°_{37}\begin{bmatrix}\text{GA}\\\text{CU}\end{bmatrix} - \Delta G°_{37}\begin{bmatrix}\text{AC}\\\text{UG}\end{bmatrix} - \Delta G°_{37}\begin{bmatrix}\text{CC}\\\text{GG}\end{bmatrix} - \Delta G°_{37}\begin{bmatrix}\text{GC}\\\text{UG}\end{bmatrix} - \Delta G°_{37}\begin{bmatrix}\text{CU}\\\text{GA}\end{bmatrix} - \Delta G°_{37}\begin{bmatrix}\text{UG}\\\text{AC}\end{bmatrix} \quad (5)$$

Here, $\Delta G°_{37}[\begin{smallmatrix}5'\text{GACCUGCUG}3'\\3'\text{CUGGUUGAC}5'\end{smallmatrix}]$ is the value determined by optical melting experiments; $\Delta G°_{37,\text{i}}$ is the free energy change for duplex initiation, 4.09 kcal/mol (*43*); and $\Delta G°_{37,\text{single mismatch}}$ is the free energy contribution of the mismatch. More explicitly,

$$\begin{aligned}\Delta G°_{37,\text{single mismatch}} &= -13.04 - 4.09 - (-2.35) - (-2.24) - (-3.42) - (-2.11) - (-2.08) - (-2.11)\\ &= -2.82\ \text{kcal/mol} \quad (6)\end{aligned}$$

Previous studies (*23*, *24*) used reference duplexes to calculate $\Delta G°_{37,\text{single mismatch}}$ values; therefore, these values were used when available. However, as a result of the various nearest neighbor combinations used in this work, a reference strand would have been required for each single mismatch studied. Thus, nearest neighbor parameters were used as an alternative to the reference duplexes. The values of $\Delta H°_{\text{single mismatch}}$ and $\Delta S°_{\text{single mismatch}}$ were calculated in a similar manner.

Linear Regression and Single Mismatch Thermodynamic Parameters. Data collected for 28 duplexes in this study were combined with previously published data for 49 single mismatches (*21*−*24*). Of the 77 total duplexes, seven melted in a non-two-state manner and were not included in trends, averages, or linear regression. In addition, data from the following sequences were significantly different from what was predicted: $[\begin{smallmatrix}5'\text{CAGCUGGUC}3'\\3'\text{GUCGUUCAG}5'\end{smallmatrix}]$, $[\begin{smallmatrix}5'\text{CGAUCGC}3'\\3'\text{GCUUGCG}5'\end{smallmatrix}]$ (*24*), $[\begin{smallmatrix}5'\text{CGAUCAC}33'\\3'\text{GCUUGUG}5'\end{smallmatrix}]$ (*24*), $[\begin{smallmatrix}5'\text{CAGAAUGUC}3'\\3'\text{GUCUGGCAG}5'\end{smallmatrix}]$, $[\begin{smallmatrix}5'\text{CUCUCUC}3'\\3'\text{GAGUGAG}5'\end{smallmatrix}]$ (*24*), $[\begin{smallmatrix}5'\text{GAGGAGAG}'\\3'\text{CUCUGCUC}5'\end{smallmatrix}]$ (*21*), $[\begin{smallmatrix}5'\text{GAGGUGAG}3'\\3'\text{CUCAGCUC}5'\end{smallmatrix}]$ (*21*), and $[\begin{smallmatrix}5'\text{UGACACUCA}3'\\3'\text{ACUGAGAGU}5'\end{smallmatrix}]$ (*23*). Further investigation of the first five duplexes listed here revealed that self-complementary duplexes (resulting in the formation of AA or BB duplexes) may compete with the bimolecular association of the two different strands (resulting in AB duplexes). Further investigation of the last three duplexes listed here revealed that bimolecular association of the two strands forms a suboptimal secondary structure that is only 0.1 kcal/mol less favorable than the predicted free energy for the duplex containing the single mismatch. The formation of the desired duplex with a single mismatch could not be confirmed; therefore, the data for these eight duplexes were not included in trends, averages, and linear regression. Because two duplexes melted in a non-two state manner and had competing structures, the thermodynamic parameters for 64 (40 reported previously and 24 reported here) single mismatches were included in the linear regression used to derive single mismatch-specific nearest neighbor parameters. Three parameters consisting of a total of nine variables were used for linear regression: (1) a mismatch parameter containing variables for an A·G, G·G, or U·U mismatch; (2) a stacking parameter containing variables for $[\begin{smallmatrix}5'\text{YRR}3'\\3'\text{RRY}5'\end{smallmatrix}]$, $[\begin{smallmatrix}5'\text{RYY}3'\\3'\text{YYR}5'\end{smallmatrix}]$, $[\begin{smallmatrix}5'\text{YYR}3'\\3'\text{RYY}5'\end{smallmatrix}]$, $[\begin{smallmatrix}5'\text{YRY}3'\\3'\text{RYR}5'\end{smallmatrix}]$, and

$\left[{}^{5'\mathrm{RRY}3'}_{3'\mathrm{YYR}5'}\right]$ stacking combinations, when cytosine and uracil are classified as pyrimidines (Y), and adenine and guanine are classified as purines (R); and (3) a parameter for an A-U/G-U closure. The calculated experimental contribution of the single mismatch to duplex stability was used as a constant when doing linear regression. To simultaneously solve for each variable, the LINEST function of *Microsoft Excel* was used for linear regression. Many combinations of variables were tried, but this combination of variables produced a model that agreed closely with the experimental data and had error values that were comparable to those of the *RNAstructure* algorithm (*14*−*16*).

RESULTS

Database Searching. The database containing 955 RNA secondary structures and 151,503 nucleotides was searched for single mismatches. In this database, 3284 single mismatches were found, averaging about three occurrences for each sequence. Table 1 shows a summary of the database results obtained. The first set of data lists frequency and percent occurrence when the mismatch nucleotides and nearest neighbors are specified. Categorizing single mismatches in this fashion results in 182 types of mismatches in the database. The 30 mismatch types listed in the first data set (Table 1) account for 54% of the total number of mismatches found. The 152 types of mismatches not shown account for the remaining 46%; however, each type represents <1% of the total number of mismatches found. When categorized in this manner, previous studies account for only 28% of the total number of single mismatches found, but after adding the data reported here, this percentage increases to 63%. Similarly, previous studies thermodynamically characterized only nine types of mismatches in the top 30, but after adding the data reported here, all of the mismatches in the top 30 have been studied.

The second set of data (Table 1) lists frequency and percent occurrence when only the mismatch sequence is specified. Categorizing single mismatches in this fashion results in seven types of mismatches in the database, representing all possible types of single mismatches. When categorized in this manner, previous studies account for all types of mismatches; however, the current work has also characterized five of these seven types.

The third set of data (Table 1) lists the frequency and percent occurrence of 5′ and 3′ nearest neighbor combinations. Categorizing single mismatches in this fashion results in 21 types of nearest neighbor combinations in the database, representing all possible types of nearest neighbor combinations. When categorized in this manner, previous studies account for 58% of all nearest neighbor combinations, but after adding the data reported here, this percentage increases to 92%.

Thermodynamic Parameters. Table 2 shows the thermodynamic parameters of duplex formation that were obtained from fitting each melting curve to the two-state model and from the van't Hoff plot of T_M^{-1} versus $\log(C_T/4)$. Data for the duplexes containing the 30 most frequently occurring single mismatches in the database are shown in order of decreasing frequency. However, data for 42 duplexes are shown because one mismatch was melted more than once with the same stem sequence, and several duplexes were melted with different stem sequences. As described in Materials and Methods, several duplexes melted in a non-two-state manner, and several duplexes may have been influenced by competing structures. These duplexes are marked in Table 2.

Contribution of Single Mismatches to Duplex Thermodynamics. The contributions of the 42 single mismatches to duplex stability are listed in Table 3. These contributions are described in Materials and Methods and are further defined by eqs 5 and 6. An additional 35 single mismatches were added to the 42 from Table 2, and the total list of 77 single mismatches can be found in Supporting Information (Table S1). These additional duplexes occur less frequently, and they have been thermodynamically characterized by previous studies (*21*−*24*).

Updated Model for Predicting the Thermodynamics of Single Mismatches. An updated model to predict secondary structure from sequence was derived by compiling data from previous works (*21*−*24*) and the data obtained from this work. Linear regression was then used to derive nearest neighbor parameters for predicting the contribution of a single mismatch to duplex thermodynamics. The free energy of an RNA duplex containing a single mismatch that has not been thermodynamically characterized can be calculated by the following equation:

$$\Delta G^\circ_{37,\text{single mismatch}} = \Delta G^\circ_{37,\text{mismatch nt}} + \Delta G^\circ_{37,\text{mismatch}-\text{NN interaction}} + \Delta G^\circ_{37,\text{AU/GU}} \quad (6)$$

Here, $\Delta G^\circ_{37,\text{mismatch}}$ is −0.4 ± 0.2, −2.1 ± 0.2, and −0.3 ± 0.2 kcal/mol for A•G, G•G and U•U mismatches, respectively; $\Delta G^\circ_{37,\text{mismatch}-\text{NN interaction}}$ is 0.7 ± 0.2, −0.5 ± 0.2, 0.4 ± 0.3, −0.4 ± 0.3, and −1.0 ± 0.4 kcal/mol for $\left[{}^{5'\mathrm{YRR}3'}_{3'\mathrm{RRY}5'}\right]$, $\left[{}^{5'\mathrm{RYY}3'}_{3'\mathrm{YYR}5'}\right]$, $\left[{}^{5'\mathrm{YYR}3'}_{3'\mathrm{RYY}5'}\right]$, $\left[{}^{5'\mathrm{YRY}3'}_{3'\mathrm{RYR}5'}\right]$, and $\left[{}^{5'\mathrm{RRY}3'}_{3'\mathrm{YYR}5'}\right]$ mismatch and nearest neighbor combinations, respectively, when A and G are categorized as purines (R), and C and U are categorized as pyrimidines (Y); and $\Delta G^\circ_{37,\text{AU/GU}}$ is a penalty of 1.2 ± 0.1 kcal/mol for replacing a G-C base pair with either an A-U or a G-U base pair. This newly proposed algorithm will only be used when experimental data are unavailable.

The free energy contributions of the 64 single mismatches were predicted using this model and compared to the experimental free energy values, resulting in a root-mean-square deviation of 0.47 kcal/mol. This deviation is slightly improved over the 0.67 kcal/mol root-mean-square deviation calculated using the current *RNAstructure* algorithm (*14*−*16*) for the same set of mismatches.

Equations similar to eq 6 can also be written for $\Delta H^\circ_{\text{single mismatch}}$ and $\Delta S^\circ_{\text{single mismatch}}$. All of the nearest neighbor parameters are listed in Table 4. A table showing the $\Delta G^\circ_{37,\text{single mismatch}}$ values for all possible combinations of single mismatches and nearest neighbors can be found in Supporting Information (Table S2).

DISCUSSION

RNA sequencing projects are generating an abundance of sequence information. In order to learn more about structure−function relationships of RNA and RNA−RNA, RNA−DNA, RNA−protein, and RNA−drug interactions and to improve rational drug design to target bacteria, viruses, and other diseases, many scientists are interested in secondary

Table 1: Summary of Database Search Results for Single Mismatches[a]

dataset 1			
mismatch with nearest neighbors			
mismatch[b]	freq[c]	%[d]	ref
G**U**C C**U**G	183	5.6	*e*
U**A**C A**G**G	157	4.8	*h*
C**U**G G**U**U	104	3.2	*f, h*
U**A**G A**G**C	97	2.9	*h*
A**U**C U**U**G	94	2.9	*e, h*
A**A**U U**G**G	89	2.7	*h*
A**A**C U**C**G	69	2.1	*h*
A**U**A U**U**U	62	1.9	*e*
C**A**G G**C**C	60	1.8	*h*
A**A**G U**C**C	54	1.6	*h*
C**A**C G**G**G	53	1.6	*h*
U**A**U A**G**A	53	1.6	*h*
A**G**G U**G**C	50	1.5	*h*
G**C**C C**U**G	48	1.5	*h*
C**A**C G**C**G	47	1.4	*e, h*
G**A**U C**C**A	45	1.4	*h*
G**U**G C**U**U	43	1.3	*f, h*
G**A**G C**C**C	42	1.3	*e*
C**U**C G**U**G	41	1.2	*e, h*
U**A**C G**G**G	40	1.2	*h*
U**A**G G**G**C	38	1.2	*f, h*
U**C**U A**U**A	38	1.2	*h*
G**A**C C**C**G	36	1.1	*h*
A**U**G U**U**C	36	1.1	*h*
G**A**C C**G**G	35	1.1	*h*
U**A**A A**G**U	35	1.1	*h*
U**A**A A**A**U	34	1.0	*h*
A**A**A U**C**U	34	1.0	*h*
A**A**C U**G**G	34	1.0	*h*
A**C**U U**U**A	34	1.0	*h*
previously	933	28.4	
new total	2064	62.9	

dataset 2			
mismatch			
mismatch[b]	freq[c]	%[d]	ref
A **G**	912	27.7	*e, f, h, i*
U **U**	792	240	*e, f, h*
A **C**	651	19.8	*e, f, h*
C **U**	410	12.5	*e, f, h*
A **A**	216	6.6	*e−g*
G **G**	214	6.5	*e, h*
C **C**	99	3.0	*e, f*
previously	3284	100.0	
new total	3284	100.0	

dataset 3				
5′ and 3′ adjacent base pairs				
closing	bp	freq[c]	%[d]	ref
G C	C G	340	10.4	*e, h*
G C	A U	320	9.7	*h*
A U	C G	303	9.2	*e, h*
C G	A U	271	8.3	*h*
A U	A U	268	8.2	*e, h*
C G	C G	264	8.0	*e, g, h. i*
A U	G C	253	7.7	*h*
C G	G U	201	6.1	*f−h*
C G	G C	148	4.5	*e, h*
G C	G U	146	4.5	*f, h*
A U	U G	128	3.9	*h*
A U	U A	116	3.5	*e, h*
U A	A U	113	3.4	*e, h*
C G	U G	88	2.7	*f*
U A	G U	77	2.3	
G C	U G	71	2.2	*f*
A U	G U	61	1.9	
G U	A U	54	1.6	
G U	G U	27	0.8	
U G	G U	24	0.7	
G U	U G	11	0.3	
previously		1899	57.8	
new total		3030	92.3	

[a] Not all combinations in dataset 1 are shown because of space limitations. For each set of sequences, the top strand is written 5′ to 3′, and the bottom strand is written 3′ to 5′. [b] Duplexes are written in alphabetical order by the loop nucleotide (A over G, not G over A). If the loop nucleotides are identical, then duplexes are written in alphabetical order by the nearest neighbors (CUG over GUU, not GUU over CUG). [c] Frequency of occurrence in database. [d] Percent out of 3284 mismatches, the total number of mismatches found in the database. [e] Ref *24*. [f] Ref *21*. [g] Ref *23*. [h] This work. [i] Ref *22*.

Table 2: Thermodynamic Parameters for Duplex Formation[a]

		analysis of melt curve fit/errors				analysis of T_m dependence/errors (ln plot)			
frequency[b]	sequence[c]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/K·mol)	$\Delta G°_{37}$ (kcal/mol)	T_m[d] (°C)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/K·mol)	$\Delta G°_{37}$ (kcal/mol)	T_m[d] (°C)
183	GC G**U**C CG[e] CG C**U**G GC	−66.5 ± 4.2	−187.2 ± 12.8	−8.45 ± 0.21	46.2	−62.6 ± 1.7	−175.3 ± 5.1	−8.24 ± 0.07	45.7
	GC G**U**C GC[e] CG C**U**G CG	−68.6 ± 4.0	−195.5 ± 12.5	−7.98 ± 0.16	47.8	−66.8 ± 1.3	−189.9 ± 4.0	−7.89 ± 0.03	47.6
	G**U**C CGCG[e] C**U**G GCGC	−57.6 ± 4.0	−155.9 ± 12.0	−9.26 ± 0.29	52.4	−56.9 ± 0.8	−153.8 ± 2.5	−9.25 ± 0.05	52.6
	C G**U**C CGG[e] G C**U**G GCC	−61.5 ± 3.1	−170.9 ± 9.7	−8.47 ± 0.11	47.1	−61.8 ± 1.5	−171.9 ± 4.6	−8.47 ± 0.05	47.0
157	GAC U**A**C CUG CUG A**G**G GAC	−82.7 ± 4.7	−232.9 ± 14.5	−10.46 ± 0.26	52.5	−88.8 ± 10.1	−251.8 ± 31.3	−10.67 ± 0.46	52.2
104	CU C**U**G CUC[f] GA G**U**U GAG	−68.6 ± 5.4	−201.1 ± 17.5	−6.27 ± 0.1	35.8	−67.0 ± 1.5	−195.9 ± 4.8	−6.21 ± 0.02	35.5
	CAG C**U**G GUC[h] GUC G**U**U CAG	−94.9 ± 7.3	−263.7 ± 22.0	−13.12 ± 0.51	60.1	−94.5 ± 7.2	−262.5 ± 21.7	−13.04 ± 0.48	60.0
97	GAC U**A**G CUG CUG A**G**C GAC	−76.4 ± 4.5	−217.4 ± 13.3	−8.95 ± 0.47	47.1	−76.8 ± 11.1	−218.9 ± 34.6	−8.93 ± 0.56	47.0
94	CG A**U**C GC[e, g, h] GC U**U**G CG	(−31.3)	(−81.5)	(−6.05)	(32.3)	(−57.2)	(−165.1)	(−6.03)	(34.3)
	CG A**U**C AC[e, h] GC U**U**G UG	−46.6 ± 11.7	−137.6 ± 38.7	−3.91 ± 0.34	20.5	−44.0 ± 2.8	−129.6 ± 9.3	−3.82 ± 0.10	19.0
	GAC A**U**C CUG CUG U**U**G GAC	−87.4 ± 2.3	−250.2 ± 7.1	−9.84 ± 0.10	49.2	−84.1 ± 1.6	−239.9 ± 5.0	−9.73 ± 0.06	49.2
89	CAG A**A**U GUC[h] GUC U**G**G CAG	−81.3 ± 12.9	−232.3 ± 40.6	−9.26 ± 0.47	47.8	−81.3 ± 16.4	−232.1 ± 50.8	−9.27 ± 0.87	47.8
69	CAG A**A**C GUC GUC U**C**G CAG	−84.6 ± 12.8	−242.7 ± 39.7	−9.29 ± 0.49	47.4	−83.9 ± 1.6	−240.8 ± 5.0	−9.23 ± 0.05	47.3
62	GC A**U**A CG[e, g] CG U**U**U GC	(−60.5)	(−181.1)	(−4.37)	(26.3)	(−48.1)	(−140.4)	(−4.58)	(24.9)
60	GAC C**A**G CUG CUG G**C**C GAC	−67.3 ± 13.8	−182.0 ± 42.8	−10.82 ± 0.56	58.1	−66.8 ± 8.2	−180.3 ± 24.9	−10.84 ± 0.54	58.4
	GAC C**A**G CUG CUG G**C**C GAC	−74.5 ± 13.5	−203.9 ± 40.9	−11.24 ± 0.85	57.9	−76.4 ± 3.3	−210.0 ± 9.9	−11.26 ± 0.19	57.5
54	GAC A**A**G CUG CUG U**C**C GAC	−71.2 ± 4.2	−201.2 ± 13.1	−8.77 ± 0.20	47.1	−69.5 ± 2.8	−196.0 ± 8.9	−8.71 ± 0.08	47.0
53	GAC C**A**C CUG[g] CUG G**G**G GAC	(−80.9)	(−225.3)	(−11.00)	(55.1)				
53	GAC U**A**U CUG CUG A**G**A GAC	−79.5 ± 7.8	−232.7 ± 24.6	−7.33 ± 0.22	40.1	−67.6 ± 3.9	−194.9 ± 12.7	−7.20 ± 0.06	40.1
50	GAC A**G**G CUG CUG U**G**C GAC	−84.0 ± 5.0	−237.9 ± 15.3	−10.22 ± 0.32	51.2	−85.1 ± 5.4	−241.4 ± 16.6	−10.23 ± 0.23	51.1
48	GAC G**C**C CUG CUG C**U**G GAC	−85.4 ± 2.6	−240.0 ± 7.7	−10.98 ± 0.19	54.1	−86.7 ± 2.5	−244.0 ± 7.7	−11.04 ± 0.12	54.0
47	CU C**A**C UC[e] GA G**C**G AG	−54.2 ± 3.2	−156.6 ± 10.4	−5.64 ± 0.11	32.0	−55.6 ± 2.1	−161.1 ± 6.7	−5.59 ± 0.03	31.9
	GAC C**A**C CUG CUG G**C**G GAC	−82.0 ± 10.4	−227.7 ± 31.7	−11.33 ± 0.64	56.3	−82.5 ± 4.1	−229.5 ± 12.4	−11.35 ± 0.23	56.2
45	GAC G**A**U CUG CUG C**C**A GAC	−87.5 ± 14.3	−252.3 ± 44.1	−9.23 ± 0.60	46.9	−87.2 ± 6.2	−251.6 ± 19.4	−9.18 ± 0.19	46.7
43	GA G**U**G GAG[f] CU C**U**U CUC	−67.9 ± 4.0	−198.6 ± 13.1	−6.29 ± 0.1	35.8	−70.9 ± 1.7	−208.4 ± 5.6	−6.21 ± 0.02	35.6
	GAC G**U**G CUG CUG C**U**U GAC	−83.4 ± 5.3	−241.7 ± 16.7	−8.48 ± 0.12	44.4	−83.7 ± 2.9	−242.7 ± 9.2	−8.48 ± 0.07	44.4
42	GA G**A**G AG[e] CU C**C**C UC	−42.3 ± 7.6	−120.0 ± 25.0	−5.04 ± 0.24	26.5	−40.9 ± 2.1	−115.9 ± 7.1	−4.93 ± 0.07	25.3
41	CU C**U**C UC[e, h] GA G**U**G AG	−63.3 ± 6.2	−187.3 ± 19.5	−5.22 ± 0.20	30.7	−54.9 ± 1.4	−160.1 ± 4.7	−5.21 ± 0.02	29.7
	CG C**U**C GC[e] GC G**U**G CG	−62.4 ± 2.7	−175.8 ± 8.4	−7.89 ± 0.14	43.9	−61.1 ± 1.7	−171.9 ± 5.3	−7.82 ± 0.05	43.7
	GAC C**U**G CUG CUG G**U**G GAC	−87.5 ± 3.5	−244.0 ± 10.6	−11.84 ± 0.21	57.0	−88.3 ± 2.7	−246.3 ± 8.2	−11.87 ± 0.15	57.0
40	CAG U**A**C GUC GUC G**G**G CAG	−83.3 ± 6.7	−242.0 ± 21.2	−8.28 ± 0.19	43.7	−87.0 ± 4.7	−253.7 ± 15.0	−8.37 ± 0.10	43.7
38	GAG U**A**G AG[f] CUC G**G**C UC	−68.4 ± 7.2	−200.8 ± 23.1	−6.14 ± 0.1	35.3	−63.8 ± 1.8	−186.0 ± 5.8	−6.09 ± 0.02	34.9
	CAG U**A**G GUC GUC G**G**C CAG	−56.5 ± 14.6	−154.3 ± 45.4	−8.59 ± 0.58	48.7	−62.6 ± 2.6	−174.2 ± 8.1	−8.57 ± 0.06	47.4
38	GAC U**C**U CUG CUG A**U**A GAC	−78.4 ± 8.4	−230.6 ± 26.6	−6.90 ± 0.23	38.5	−66.6 ± 5.2	−192.5 ± 16.9	−6.86 ± 0.11	38.5

Table 2 (Continued)

		analysis of melt curve fit/errors				analysis of T_m dependence/errors (ln plot)			
frequency[b]	sequence[c]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/K·mol)	$\Delta G°_{37}$ (kcal/mol)	T_m[d] (°C)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/K·mol)	$\Delta G°_{37}$ (kcal/mol)	T_m[d] (°C)
36	CUG C**C**G GAC GAC G**A**C CUG[g]	(−49.4)	(−128.5)	(−9.57)	(57.3)	(−76.2)	(−210.0)	(−11.03)	(56.5)
36	GAC A**U**G CUG CUG U**U**C GAC	−88.9 ± 11.5	−256.6 ± 36.2	−9.32 ± 0.36	47.1	−90.9 ± 5.9	−263.0 ± 18.6	−9.33 ± 0.17	46.9
35	GAC G**A**C CUG CUG C**G**G GAC	−70.2 ± 5.0	−193.5 ± 15.4	−10.19 ± 0.25	54.1	−72.9 ± 0.9	−201.9 ± 2.9	−10.27 ± 0.04	53.8
35	GAC U**A**A CUG CUG A**G**U GAC	−65.4 ± 4.0	−189.4 ± 13.0	−6.65 ± 0.05	37.6	−64.6 ± 1.4	−186.8 ± 4.4	−6.65 ± 0.01	37.6
34	GAC U**A**A CUG CUG A**A**U GAC	−69.4 ± 20.6	−203.5 ± 66.2	−6.25 ± 0.17	35.8	−69.4 ± 8.0	−203.8 ± 26.2	−6.19 ± 0.25	35.5
34	GAC A**A**A CUG CUG U**C**U GAC	−71.8 ± 6.4	−209.6 ± 20.6	−6.83 ± 0.06	38.3	−72.8 ± 2.8	−212.9 ± 9.2	−6.83 ± 0.03	38.3
34	GAC A**A**C CUG CUG U**G**G GAC	−77.2 ± 12.3	−219.4 ± 37.8	−9.16 ± 0.71	47.9	−75.2 ± 12.2	−213.1 ± 38.0	−9.10 ± 0.62	48.0
34	CAG A**C**U GUC GUC U**U**A CAG	−76.6 ± 13.4	−224.7 ± 42.4	−6.95 ± 0.27	38.7	−72.5 ± 5.4	−211.6 ± 17.5	−6.91 ± 0.10	38.6

[a] Measurements were made in 1.0 M NaCl, 10 mM sodium cacodylate, and 0.5 mM Na_2EDTA at pH 7.0. [b] Frequency of occurrence in the database described in Materials and Methods. [c] Single mismatch is identified by bold letters. The nearest neighbors and the mismatch are set apart for easy identification. The top strand of each duplex is written 5′ to 3′ and each bottom strand is written 3′ to 5′. [d] Calculated at 10^{-4} M oligomer concentration. [e] Ref *24*. [f]Ref *21*. [g] Data derived from non-two-state melts. [h] Duplexes that were not included in averages, trends, and the derivation of the predictive model because a bimolecular association of one of the strands with itself may be a competing structure.

and tertiary structures of RNA. Computer algorithms use thermodynamic parameters to predict secondary structure from sequence (*14*−*17*). Because these algorithms rely on thermodynamic parameters for every type of motif (Watson−Crick pairs, single mismatches, bulges, internal loops, hairpins, etc.), the accuracy of predicted secondary structures depends on the accuracy of the thermodynamic parameters for each motif. Although single mismatches are common in nature, relatively few studies have investigated their thermodynamics (*21*−*25*), resulting in thermodynamic parameters that rely on several approximations and assumptions. Additionally, the single mismatches studied previously (*21*−*24*) rarely occur in known secondary structures (Table 1). As a result, the stabilities of many commonly occurring single mismatches are based on assumptions and approximations.

In this study, the 30 most frequently occurring single mismatches have been thermodynamically characterized, providing experimental parameters for these single mismatches. These new data were combined with previous data and were used to derive single mismatch-specific nearest neighbor parameters to improve the current model used to predict the stability of RNA duplexes containing single mismatches and, therefore, improve secondary structure prediction from sequence.

Database Searching. The database compiled for this study contains 955 secondary structures from eight different kinds of RNAs. We have assumed that the number and variety of structures in this database approximates the number and types of single mismatches found in nature. It is important to note, however, that the results found from searching this database may be slightly skewed. For example, the most prevalent type of RNA found in this database was tRNA, accounting for roughly half of the structures in the database. Therefore, single mismatches prevalent in tRNA structures may be over-represented. However, the database is large enough and diverse enough to contain all possible single mismatches when considering the mismatch nucleotides, all possible nearest neighbor combinations, and 95% of all possible combinations of mismatch-nearest neighbor nucleotides.

It is clear from the first set of data in Table 1 that the pioneering experiments on single mismatches (*21*−*24*) provided results for only nine of the top 30 mismatch-nearest neighbor nucleotide combinations found most commonly in the database. The results reported here expand the database to include all combinations in the top 30. The first set of data in Table 1 provides some important results. For example, it is interesting to note that only seven of the top 30 single mismatches contain mismatches that have been previously considered to be stabilizing (G·G and 5′RU/3′YU) in single mismatches (*14*−*17*). Clearly, the presence or absence of stabilizing mismatches does not determine the frequency of occurrence for single mismatches.

The second set of data listed in Table 1 indicates that all possible single mismatches were found in the database. A·G and U·U mismatches are the most prevalent single mismatches. As mentioned previously, it is clear that stability does not always indicate relative frequency since G·G mismatches are shown to contribute the greatest degree of stability to a duplex while having the second smallest relative frequency of all single mismatches.

The third set of data in Table 1 shows the nearest neighbor combinations of single mismatches and their relative frequency. The most frequent nearest neighbor combination is $[^{\mathrm{GXC}}_{\mathrm{CXG}}]$, representing 10% of the total number of mismatches. G-U pairs occur much less frequently than A-U and G-C pairs; however, 27% of the mismatches in the database have at least one adjacent G-U pair. This percentage is similar to what was found for 1 × 2 internal loops; 30% of 1 × 2 loops in the database had at least one adjacent G-U pair (unpublished data). Both of these results are consistent with

Table 3: Contributions of 42 Single Mismatches to Duplex Thermodynamics[a]

		ΔH° (kcal/mol)			ΔS° (cal/K·mol)		ΔG°_{37} (kcal/mol)		
frequency[b]	sequence[c]	measured	single mismatch-specific model[d]	Mathews et al. model[e]	measured	single mismatch-specific model[d]	measured	single mismatch-specific model[d]	Mathews et al. model[e]
183	G**U**C[f] C**U**G	−19.6	−18.1 (1.5)	−13.9 (5.7)	−60.9	−55.8 (5.1)	−0.75	−0.8 (0.0)	0.1 (−0.9)
	G**U**C[f] C**U**G	−13.5	−18.1 (4.6)	−13.9 (0.4)	−41.8	−55.8 (14.0)	−0.48	−0.8 (0.3)	0.1 (0.6)
	G**U**C[f] C**U**G	−12.1	−18.1 (6.0)	−13.9 (1.8)	−33.3	−55.8 (22.5)	−1.79	−0.8 (1.0)	0.1 (1.9)
	G**U**C[f] C**U**G	−18.0	−18.1 (0.1)	−13.9 (4.1)	−53.3	−55.8 (2.5)	−1.30	−0.8 (0.5)	0.1 (1.4)
157	U**A**C A**G**G	−23.8	−8.7 (15.1)	−5.5 (18.3)	−74.5	−30.8 (43.7)	−0.64	0.8 (1.4)	1.2 (1.8)
104	C**U**G[g] G**U**U	−12.2	−12.2 (0.1)	−5.5 (6.7)	−42.7	−43.0 (0.3)	1.07	1.3 (0.2)	1.2 (0.3)
	C**U**G[f] G**U**U	−26.4	−12.2 (14.2)	−5.5 (20.9)	−75.9	−43.0 (32.9)	−2.82	1.3 (4.1)	1.2 (4.0)
97	U**A**G A**G**C	−10.3	−8.5 (1.8)	−5.5 (4.8)	−37.4	−32.2 (5.2)	1.26	1.5 (0.2)	1.2 (0.1)
94	A**U**C[f, h, i] U**U**G	−16.2	−19.9 (3.7)	−8.9 (3.7)	−54.0	−65.6 (11.6)	−0.49	0.4 (0.9)	0.8 (1.3)
	A**U**C[f,i] U**U**G	4.5	−19.9 (24.4)	−8.9 (13.4)	−18.2	−65.6 (47.4)	1.12	0.4 (0.7)	0.8 (0.3)
	A**U**C U**U**G	−19.1	−19.9 (0.8)	−8.9 (10.2)	−62.8	−65.6 (2.8)	0.33	0.4 (0.1)	0.8 (0.5)
89	A**A**U[i] U**G**G	−22.1	−10.5 (11.6)	−0.5 (21.6)	−68.6	−40.6 (28.0)	−0.82	2.0 (2.8)	1.9 (2.7)
69	A**A**C U**C**G	−19.7	−21.3 (1.7)	−5.5 (14.1)	−64.1	−69.5 (5.4)	0.17	0.2 (0.0)	1.2 (1.0)
62	A**U**A[f, h] U**U**U	−4.3	−16.2 (11.2)	−3.9 (0.4)	−18.9	−59.2 (40.3)	−1.50	2.1 (3.6)	1.5 (3.0)
60	C**A**G G**C**C	2.6	0.0 (2.6)	−10.5 (13.1)	6.8	0.0 (6.8)	0.53	0.0 (0.5)	0.5 (0.0)
	C**A**G G**C**C	−7.0	0.0 (7.0)	−10.5 (3.5)	−22.9	0.0 (22.9)	0.11	0.0 (0.1)	0.5 (0.4)
54	A**A**G U**C**C	−3.0	−1.8 (1.2)	−5.5 (2.5)	−14.7	−9.8 (4.9)	1.51	1.2 (0.3)	1.2 (0.3)
53	C**A**C[h] G**G**G	−13.0	−6.9 (6.1)	−10.5 (2.5)	−42.4	−21.0 (21.5)	0.21	−0.4 (0.2)	0.5 (0.7)
53	U**A**U A**G**A	−3.5	−10.5 (7.0)	−0.5 (3.0)	−17.8	−40.6 (22.8)	1.92	2.0 (0.1)	1.9 (0.0)
50	A**G**G U**G**C	−18.6	−20.0 (1.4)	−13.4 (5.2)	−60.1	−61.6 (1.5)	−0.01	−0.9 (0.9)	−1.4 (1.4)
48	G**C**C C**U**G	−21.5	−5.5 (16.0)	−10.5 (11.0)	−67.1	−16.2 (50.9)	−0.73	−0.5 (0.2)	0.5 (1.3)
47	C**A**C[f] G**C**G	−13.4	−7.7 (5.7)	−10.5 (2.9)	−41.6	−23.4 (18.2)	−0.35	−0.4 (0.1)	0.5 (0.9)
	C**A**C G**C**G	−14.6	−7.7 (6.9)	−10.5 (4.1)	−46.6	−23.4 (23.2)	−0.14	−0.4 (0.3)	0.5 (0.6)
45	G**A**U C**C**A	−23.0	−21.3 (1.7)	−5.5 (17.5)	−74.9	−69.5 (5.4)	0.22	0.2 (0.0)	1.2 (1.0)
43	G**U**G[g] C**U**U	−16.6	−14.4 (2.2)	−8.9 (7.7)	−55.5	−49.4 (6.10)	0.67	0.9 (0.2)	0.8 (0.10)
	G**U**G C**U**U	−19.3	−14.4 (4.9)	−8.9 (10.4)	−66.0	−49.4 (16.6)	1.08	0.9 (0.2)	0.8 (0.3)
42	G**A**G[f] C**C**C	1.4	0.0 (1.4)	−10.5 (11.9)	3.6	0.0 (3.6)	0.31	0.0 (0.3)	0.5 (0.2)
41	C**U**C[f, h] G**U**G	−12.6	−12.6 (0.0)	−13.9 (1.3)	−40.6	−39.6 (1.0)	0.03	−0.3 (0.3)	0.1 (0.1)
	C**U**C[f] G**U**G	−13.0	−12.6 (0.4)	−13.9 (0.9)	−41.5	−39.6 (1.9)	−0.16	−0.3 (0.1)	0.1 (0.3)
	C**U**C G**U**G	−20.4	−12.6 (7.8)	−13.9 (6.5)	−63.4	−39.6 (23.8)	−0.66	−0.3 (0.3)	0.1 (0.8)
40	U**A**C G**G**G	−22.6	−8.7 (13.9)	−5.5 (17.1)	−77.0	−30.8 (46.2)	1.19	0.8 (0.4)	1.2 (0.0)
38	U**A**G[g] G**G**C	−9.0	−8.5 (0.5)	−5.5 (3.5)	−32.8	−32.2 (0.6)	1.19	1.5 (0.3)	1.2 (0.0)
	U**A**G G**G**C	4.5	−8.5 (13.0)	−5.5 (10.0)	8.50	−32.2 (40.7)	1.89	1.5 (0.4)	1.2 (0.7)
38	U**C**U A**U**A	−2.5	−3.6 (1.1)	−0.5 (2.0)	−15.4	−19.6 (4.2)	2.26	2.4 (0.1)	1.9 (0.4)

Table 3 (Continued)

		ΔH° (kcal/mol)			ΔS° (cal/K·mol)		ΔG°$_{37}$ (kcal/mol)		
frequency[b]	sequence[c]	measured	single mismatch-specific model[d]	Mathews et al. model[e]	measured	single mismatch-specific model[d]	measured	single mismatch-specific model[d]	Mathews et al. model[e]
36	G**A**C[h] C**C**G	−11.0	−19.5 (8.5)	−10.5 (−0.5)	−33.1	−59.7 (26.6)	−0.72	−1.0 (0.3)	0.5 (1.2)
36	A**U**G U**U**C	−24.4	−14.4 (10.0)	−8.9 (15.5)	−81.7	−49.4 (32.3)	0.89	0.9 (0.0)	0.8 (0.1)
35	G**A**C C**G**G	−7.7	−6.9 (0.8)	−10.5 (2.8)	−25.0	−21.0 (4.1)	0.04	−0.4 (0.5)	0.5 (0.5)
35	U**A**A A**G**U	−1.6	−10.3 (8.6)	−0.5 (1.1)	−12.7	−42.0 (29.3)	2.36	2.7 (0.3)	1.9 (0.5)
34	U**A**A A**A**U	−6.4	−3.4 (3.0)	−0.5 (5.9)	−29.7	−21.0 (8.7)	2.82	3.1 (0.3)	1.9 (0.9)
34	A**A**A U**C**U	−9.8	−3.6 (6.2)	−0.5 (9.3)	−39.0	−19.6 (19.4)	2.21	2.4 (0.2)	1.9 (0.3)
34	A**A**C U**G**G	−10.2	−8.7 (1.5)	−5.5 (4.7)	−36.0	−30.8 (5.2)	0.96	0.8 (0.2)	1.2 (0.2)
34	A**C**U U**U**A	−8.5	−9.1 (0.6)	−0.5 (8.0)	−34.7	−35.8 (1.1)	2.24	1.9 (0.3)	1.9 (0.3)

[a] Calculations were based on the data obtained from T_M^{-1} vs ln (C_T/4) plots. Differences between the measured values and the predicted values are shown in parentheses. [b] Frequency of occurrence in the database is described in Materials and Methods. [c] Single mismatch is identified by bold letters. The top strand of each duplex is written 5′ to 3′, and each bottom strand is written 3′ to 5′. [d] Values derived from the proposed model of this work. [e] Values predicted by models published by Mathews et al. (*14*−*16*). [f] Ref *24*. [g] Ref *21*. [h] Data derived from non-two-state melts. [i] Duplexes that were not included in averages, trends, and the derivation of the predictive model because a bimolecular association of one of the strands with itself may be a competing structure.

Table 4: Nearest Neighbor Parameters for Single Mismatches at 37 °C

	ΔH° (kcal/mol)	ΔS° (cal/K·mol)	ΔG°$_{37}$ (kcal/mol)
mismatch parameters			
A·G	−6.9 ± 2.4	−21.0 ± 7.9	−0.4 ± 0.2
G·G	−18.2 ± 3.0	−51.8 ± 9.6	−2.1 ± 0.2
U·U	−12.6 ± 2.5	−39.6 ± 8.2	−0.3 ± 0.2
mismatch-NN interaction parameters[a]			
YRR **RRY**	0.2 ± 2.6	−1.4 ± 8.5	0.7 ± 0.2
RYY **YYR**	−5.5 ± 3.2	−16.2 ± 10.2	−0.5 ± 0.2
YYR **RYY**	2.2 ± 3.8	5.7 ± 12.4	0.4 ± 0.3
YRY **RYR**	−7.7 ± 4.2	−23.4 ± 13.5	−0.4 ± 0.3
RRY **YYR**	−19.5 ± 5.2	−59.7 ± 16.9	−1.0 ± 0.4
A-U or G-U closure parameters[b]			
	−1.8 ± 1.3	−9.8 ± 4.1	1.2 ± 0.1

[a] The top strand is written 5′ to 3′, and the bottom strand is written 3′ to 5′. Adenine and guanine are classified as purines (R), and cytosine and uracil are classified as pyrimidines (Y). The pairs on the left and right are the adjacent nearest neigbors, and the pair in the center is the single mismatch. [b] These parameters are applied per A-U or G-U closure.

a previous study that showed G-U base pairs occur frequently at the loop−helix junctions of small and large subunit RNA (*44*).

Thermodynamic Contributions of Single Mismatches to Duplex Thermodynamics. An examination of the data listed in Tables 3 and S1 (Supporting Information) indicates a large variance in the obtained thermodynamic parameters. Single mismatch contributions to enthalpy, entropy, and free energy changes range from −25.1 to 4.5 kcal/mol, −81.7 to 8.5 cal/(K·mol), and −2.58 to 2.82 kcal/mol, respectively (Tables 3 and S1 (Supporting Information)).

Of the 77 single mismatches in Tables 3 and S1 (Supporting Information), 19 stabilize the duplex, 45 destabilize the duplex, and 13 were omitted (as described in Materials and Methods). There does seem to be a correlation between the number of G-C pairs directly adjacent to the mismatch and the free energy contribution to duplex stability. For example, the 29 single mismatches with two adjacent G-C pairs contribute an average of −0.5 kcal/mol to duplex stability. The 28 mismatches with one G-C adjacent base pair and the seven mismatches with no adjacent G-C base pairs contribute an average of 1.1 and 2.2 kcal/mol to duplex stability, respectively.

It is interesting to note that there are nine single mismatches (mismatch nucleotides plus adjacent nearest neighbors) that were studied in at least two different stem sequences (Tables 3 and S1(Supporting Information)). The differences in contributions of the mismatches to duplex stabilities can vary with the stem sequence and/or position of the mismatch within the stem. For example, the free energy contribution of $[^{5'\mathrm{CUG}3'}_{3'\mathrm{GUC}5'}]$ only differed by 0.1 kcal/mol when it was placed in the center of two different stems; however, $[^{5'\mathrm{AUC}3'}_{3'\mathrm{UUG}5'}]$ differed by 0.8 kcal/mol when it was placed in the center of two different stems. Similarly, the free energy contribution of $[^{5'\mathrm{GGC}3'}_{3'\mathrm{CGG}5'}]$ differed by only 0.1 kcal/mol when it was moved from the center of a duplex to an off-center position of a different stem; however, $[^{5'\mathrm{GAC}3'}_{3'\mathrm{CAG}5'}]$ differed by 2.1 kcal/mol when it was moved from the center of a duplex to an off-center position of a different stem. Thus, changing the identity of the base pairs not directly adjacent to the mismatch and/or the location of the mismatch within the duplex can make a substantial difference in the free energy contribution of the mismatch. Similar non-nearest neighbor effects have been observed previously for single mismatches (*24*, *45*), bulges (*46*), and inosine−uridine pairs (*36*). It has recently been reported that the accuracy of RNA secondary structure prediction by free energy minimization is limited by non-nearest neighbor effects (*47*). Since non-nearest neighbor effects may be complicated to interpret and to include in algorithms such as *RNAstructure* and *mfold*, non-nearest neighbor effects were ignored here, and data were treated as if stability relied only upon immediate nearest neighbors. We are, however, currently investigating the role of non-nearest neighbor effects.

Examination of the single mismatch free energy values (Tables 3 and S1 (Supporting Information)) gives the following order of mismatch stability (from most stabilizing to least stabilizing): G•G (range of −2.6 to 0 kcal/mol, average of −1.8 kcal/mol) > U•U (range of −1.8 to 1.7 kcal/mol, average of 0.3 kcal/mol) > A•C (range of −0.4 to 2.2 kcal/mol, average of 0.6 kcal/mol) > C•U (range of −0.7 to 2.3 kcal/mol, average of 0.9 kcal/mol) > A•A (range of −1.3 to 2.8 kcal/mol, average of 1.1 kcal/mol) ≈ A•G (range of −0.6 to 2.4 kcal/mol, average of 1.1 kcal/mol) ≈ C•C (range of 0.5 to 1.7 kcal/mol, average of 1.1 kcal/mol). From the ranges observed for every mismatch, it is evident that the identity of the single mismatch nucleotides alone does not determine the stability of single mismatches. The nearest neighbors and the interactions between the single mismatch nucleotides and the nearest neighbors also contribute to single mismatch stability. For example, the stability of a $\left[\begin{smallmatrix}5'\mathrm{GGC}3'\\3'\mathrm{CGG}5'\end{smallmatrix}\right]$ single mismatch may result from hydrogen bonding between the G•G mismatch, three hydrogen bonds between both G-C adjacent base pairs, and the interactions between the mismatch nucleotides and the adjacent base pairs. Conversely, the lack of stability of a $\left[\begin{smallmatrix}5'\mathrm{UAA}3'\\3'\mathrm{AAU}5'\end{smallmatrix}\right]$ single mismatch may result from a lack of hydrogen bonding between the A•A mismatch, only two hydrogen bonds between both A-U adjacent base pairs, and the lack of interactions between the mismatch nucleotides and the adjacent base pairs.

Single Mismatch-Specific Model for Predicting Thermodynamics of Single Mismatches. RNAstructure uses two methods to calculate the free energy contribution of single mismatches (*14*−*16*). If a single mismatch has been measured thermodynamically, *RNAstructure* uses the measured value or average of measured values. Previous to this study, only nine of the 30 most frequent single mismatches found in the database had been measured thermodynamically (Table 1, first set of data). Now, the 30 most frequently occurring single mismatches in the database have experimental data. Since most scientists are likely interested in single mismatches that do occur in nature, these single mismatches now have experimental data, and scientists no longer have to rely on a model based on several approximations and assumptions for these single mismatches.

For those single mismatches that do not have experimental numbers, *RNAstructure* approximates the free energy contribution as described in the Introduction and shown by eq 1. These parameters were derived from the smaller dataset of single mismatch data and from data for internal loops of various sizes. Since the number of single mismatches with experimental values has significantly increased with this work, the previous and new single mismatch data were combined to derive single mismatch-specific nearest neighbor parameters. Parameters were chosen to account for the identity of the single mismatch nucleotides, the identity of the nearest neighbors, and the interaction between the single mismatch nucleotides and the nearest neighbors. The new model to predict the free energy contribution of single mismatches to duplex stability is described in eq 6 and Tables 4 and S2 (Supporting Information).

The $\Delta G^{\circ}_{37,\mathrm{mismatch\ nt}}$ parameter accounts for the identity of the single mismatch nucleotides. This parameter assigns a bonus to A•G (−0.4 kcal/mol), G•G (−2.1 kcal/mol), and U•U (−0.3 kcal/mol) mismatches (all other mismatches are assumed to contribute no favorable or unfavorable contributions to duplex stability). From the results of previous studies (*16*, *48*), these mismatches have been considered stabilizing mismatches. The *RNAstructure* model (*14*−*16*), however, does not assign a bonus to A•G single mismatches, but it does assign a bonus of −2.6 kcal/mol for G•G mismatches and a bonus of −0.4 kcal/mol to U•U mismatches only if the mismatch is adjacent to a 5′R-Y nearest neighbor (where R is A or G, and Y is C or U in an A-U or G-C base pair). The bonus for U•U mismatches proposed here is within experimental error of *RNAstructure*’s U•U bonus; however, it is important to note that the proposed U•U bonus is independent of the nearest neighbors and is applied to all U•U mismatches. These A•G, G•G, and U•U mismatches likely involve more and/or stronger hydrogen bonds between the mismatch nucleotides than do A•A, A•C, C•C, and C•U mismatches.

The $\Delta G^{\circ}_{37,\mathrm{mismatch-NN\ interaction}}$ parameter accounts for the interaction between the mismatch nucleotides and the nearest neighbors. This parameter assigns a bonus to $\left[\begin{smallmatrix}5'\mathrm{RYY}3'\\3'\mathrm{YYR}5'\end{smallmatrix}\right]$ (−0.5 kcal/mol), $\left[\begin{smallmatrix}5'\mathrm{YRY}3'\\3'\mathrm{RYR}5'\end{smallmatrix}\right]$ (−0.4 kcal/mol), and $\left[\begin{smallmatrix}5'\mathrm{RRY}3'\\3'\mathrm{YYR}5'\end{smallmatrix}\right]$ (−1.0 kcal/mol) and a penalty to $\left[\begin{smallmatrix}5'\mathrm{YRR}3'\\3'\mathrm{RRY}5'\end{smallmatrix}\right]$ (0.7 kcal/mol) and $\left[\begin{smallmatrix}5'\mathrm{YYR}3'\\3'\mathrm{RYY}5'\end{smallmatrix}\right]$ (0.4 kcal/mol), when A and G are categorized as purines (R) and when C and U are categorized as pyrimidines (Y). All other combinations of single mismatch nucleotides and nearest neighbors are assumed to contribute no favorable or unfavorable contributions to duplex stability. *RNAstructure* does not include a parameter to account for the interaction between the mismatch nucleotides and the nearest neighbors. The *RNAstructure* 5′RU/3′YU bonus (−0.4 kcal/mol) (*14*−*16*) described above accounts for both the identity of the single mismatch nucleotides and the interaction between the mismatch nucleotides and the nearest neighbors. These five particular combinations of mismatch nucleotides and nearest neighbors must be arranged in the duplex in such a way to optimize stacking and/or minimize helix distortion in the case of the combinations with a bonus and in such a way to make stacking unfavorable and/or distort the helix in the case of the combinations with a penalty.

The $\Delta G^{\circ}_{37,\mathrm{AU/GU}}$ parameter accounts for the identity of the nearest neighbors. This parameter assigns a penalty (1.2 kcal/mol) for replacing a G-C nearest neighbor with an A-U or G-U nearest neighbor. The *RNAstructure* model (*14*−*16*) assigns a penalty of 0.7 kcal/mol for this same parameter. The value proposed here was derived from a dataset of solely single mismatch thermodynamics. The value proposed by the *RNAstructure* model was derived from a dataset of internal loops of various sizes. This difference in the datasets used for the derivation may account for the difference in the values.

A similar comparison can be made between the enthalpy parameters published previously (*14*) and those derived here. The enthalpy values for a G•G mismatch, for a U•U mismatch, and for an A-U/G-U closure are easiest to compare directly. All three of the values proposed here (Table 4) are statistically different from those proposed previously (*14*). Entropy parameters were not published previously but are derived here (Table 4).

By identifying the most common single mismatches found in RNA secondary structure and measuring their thermodynamics, many more frequently occurring single mismatches now have experimental thermodynamic parameters. Since the dataset of single mismatch data has grown significantly with the data presented here, single mismatch-specific nearest neighbor parameters were derived. Both the experimental data and the new predictive model can be used when predicting the stability of an RNA duplex that contains a single mismatch and can be incorporated into *RNAstructure* or *mfold* to improve secondary structure prediction from sequence.

ACKNOWLEDGMENT

We thank the reviewers for their insightful comments and suggestions. We also thank Mr. Matthew Alexander for the critical review of the data and the manuscript.

SUPPORTING INFORMATION AVAILABLE

A table listing the contributions of 77 single mismatches to duplex thermodynamics and a table listing the thermodynamic contributions for all possible single mismatch and nearest neighbor combinations at 37 °C. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Calin-Jageman, I., and Nicholson, A. W. (2003) Mutational analysis of an RNA internal loop as a reactivity epitope for *Escherichia coli* ribonuclease III substrates, *Biochemistry 42*, 5025−5034.
2. Saito, H., and Richardson, C. C. (1981) Processing of mRNA by ribonuclease III regulates expression of gene 1.2 of bacteriophage T7, *Cell 27*, 533−542.
3. Du, T., and Zamore, P. D. (2005) MicroPrimer: The biogenesis and function of microRNA, *Development 132*, 4645−4652.
4. Bae, S. H., Cheong, H. K., Lee, J. H., Cheong, C., Kainosho, M., and Choi, B. S. (2001) Structural features of an influenza virus promoter and their implications for viral RNA synthesis, *Proc. Natl. Acad. Sci. U.S.A. 98*, 10602−10607.
5. Huthoff, H., and Berkhout, B. (2002) Multiple secondary structure rearrangements during HIV-1 RNA dimerization, *Biochemistry 41*, 10439−10445.
6. Schüler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroeer, B., Mielke, T., Penczek, P. A., Westhof, E., and Spahn, C. M. T. (2006) Structure of the ribosome-bound cricket paralysis virus IRES RNA, *Nat. Struct. Mol. Biol. 13*, 1092−1096.
7. Wientges, J., Putz, J., Giege, R., Florentz, C., and Schwienhorst, A. (2000) Selection of viral RNA-derived tRNA-like structures with improved valylation activities, *Biochemistry 39*, 6207−6218.
8. Thunder, C., Witwer, C., Hofacker, I. L., and Stadler, P. F. (2004) Conserved RNA secondary structures in Flaviviridae genomes, *J. Gen. Virol. 85*, 1113−1124.
9. Shi, P.-Y., Brinton, M. A., Veal, J. M., Zhong, Y. Y., and Wilson, W. D. (1996) Evidence for the existence of a pseudoknot structure at the 3′ terminus of the Flavivirus genomic RNA, *Biochemistry 35*, 4222−4230.
10. Everett, C. M., and Wood, N. W. (2004) Trinucleotide repeats and neurodegenerative disease, *Brain 127*, 2385−2405.
11. Ranum, L. P. W., and Day, J. W. (2004) Myotonic dystrophy: RNA pathogenesis comes into focus, *Am. J. Hum. Gen. 74*, 793−804.
12. Pinheiro, P., Scarlett, G., Rodgers, A., Rodger, P. M., Murray, A., Brown, T., Newbury, S. F., and McClellan, J. A. (2002) Structures of CUG repeats in RNA: Potential implications for human genetic diseases, *J. Biol. Chem. 277*, 35183−35190.
13. Broda, M., Kierzek, E., Gdaniec, Z., Kulinski, T., and Kierzek, R. (2005) Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure, *Biochemistry 44*, 10873−10882.
14. Lu, Z. J., Turner, D. H., and Mathews, D. H. (2006) A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation, *Nucleic Acids Res. 34*, 4912−4924.
15. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure, *J. Mol. Biol. 288*, 911−940.
16. Mathews, D. H., Disney, M. D., Childs, J. C., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc. Natl. Acad. Sci. U.S.A. 101*, 7287−7292.
17. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res. 31*, 3406−3415.
18. Hofacker, I. L. (2003) Vienna RNA secondary structure server, *Nucleic Acids Res. 31*, 3429−3431.
19. *ISI Web of Knowledge [v3.0]*, Thomson Corporation, Stamford, CT. http://portal.isiknowledge.com/portal.cgi (accessed July 2007).
20. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule, *Science 244*, 48−52.
21. Xia, T. B., McDowell, J. A., and Turner, D. H. (1997) Thermodynamics of nonsymmetric tandem mismatches adjacent to G-C base pairs in RNA, *Biochemistry 36*, 12486−12497.
22. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C, and U.U mismatches in RNA internal loops: Evidence for stable hydrogen-bonded U.U and C.C.+ pairs, *Biochemistry 30*, 8242−8251.
23. Peritz, A. E., Kierzek, R., Sugimoto, N., and Turner, D. H. (1991) Thermodynamic study of internal loops in oligoribonucleotides: Symmetric loops are more stable than asymmetric loops, *Biochemistry 30*, 6428−6436.
24. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) Thermodynamics of single mismatches in RNA duplexes, *Biochemistry 38*, 14214−14223.
25. Bevilacqua, J. M., and Bevilacqua, P. C. (1998) Thermodynamic analysis of RNA combinatorial library contained in a short hairpin, *Biochemistry 37*, 15877−15884.
26. Gutell, R. R. (1994) Collection of small-subunit (16s- and 16s-like) ribosomal-RNA structures-1994, *Nucleic Acids Res. 22*, 3502−3507.
27. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) A compilation of large subunit (23s-like and 23s-like) ribosomal-RNA structures-1993, *Nucleic Acids Res. 21*, 3055−3074.
28. Schnare, M. N., Damberger, S. H., Gray, M. W., and Gutell, R. R. (1996) Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA, *J. Mol. Biol. 256*, 701−719.
29. Szymanski, M., Specht, T., Barciszewska, M. Z., Barciszewski, J., and Erdmann, V. A. (1998) 5S rRNA data bank, *Nucleic Acids Res. 26*, 156−159.
30. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes, *Nucleic Acids Res. 26*, 148−153.
31. Larsen, N., Samuelsson, T., and Zwieb, C. (1998) The signal recognition particle database (SRPDB), *Nucleic Acids Res. 26*, 177−178.
32. Brown, J. W. (1998) The ribonuclease P database, *Nucleic Acids Res. 26*, 351−352.
33. Damberger, S. H., and Gutell, R. R. (1994) A comparative database of group I intron structures, *Nucleic Acids Res. 22*, 3508−3510.
34. Waring, R. B., and Davies, R. W. (1984) Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing: A review, *Gene 28*, 277−291.

35. Michel, F., Umesono, K., and Ozeki, H. (1989) Comparative and functional-anatomy of group-Ii catalytic introns: A review, *Gene 82*, 5−30.
36. Wright, D. J., Rice, J. L., Yanker, D. M., and Znosko, B. M. (2007) Nearest neighbor parameters for inosine-uridine pairs in RNA duplexes, *Biochemistry 46*, 4625−4634.
37. McDowell, J. A. (1995) *RNA Calculations*, Version 1.1.
38. McDowell, J. A., and Turner, D. H. (1996) 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 35*, 14077−14089.
39. Petersheim, M., and Turner, D. H. (1983) Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp, *Biochemistry 22*, 256−263.
40. Borer, P. N., Dengler, B., Tinoco, I., and Uhlenbeck, O. (1974) Stability of ribonucleic-acid double-stranded helices, *J. Mol. Biol. 86*, 843−853.
41. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1990) Effects of GA mismatches on the structure and thermodynamics of RNA internal loops, *Biochemistry 29*, 8813−8819.
42. Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers 26*, 1601−1620.
43. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs, *Biochemistry 37*, 14719−14735.
44. Gautheret, D., Konings, D., and Gutell, R. R. (1995) GU base-pairing motifs in ribosomal-RNA, *RNA 1*, 807−814.
45. Siegfried, N. A., Metzger, S. L., and Bevilacqua, P. C. (2007) Folding cooperativity in RNA and DNA is dependent on position in the helix, *Biochemistry 46*, 172−181.
46. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides, *Biochemistry 29*, 278−285.
47. Mathews, D. H. (2006) Revolutions in RNA secondary structure prediction, *J. Mol. Biol. 359*, 526−532.
48. Schroeder, S., Kim, J., and Turner, D. H. (1996) G•A and U•U mismatches can stabilize RNA internal loops of three nucleotides, *Biochemistry 35*, 16105−16109.

BI701311C