ABSTRACT: Thermodynamic parameters for GU pairs are important for predicting the secondary structures of RNA and for finding genomic sequences that code for structured RNA. Optical melting curves were measured for 29 RNA duplexes with GU pairs to improve nearest neighbor parameters for predicting stabilities of helices. The updated model eliminates a prior penalty assumed for terminal GU pairs. Six additional duplexes with the 5′GG/3′UU motif were added to the single representation in the previous database. This revises the $\Delta G^\circ$ for the 5′GG/3′UU motif from an unfavorable 0.5 kcal/mol to a favorable −0.2 kcal/mol. Similarly, the $\Delta G^\circ$ for the 5′UG/3′GU motif changes from 0.3 to −0.6 kcal/mol. The correlation coefficients between predicted and experimental $\Delta G^\circ$, $\Delta H^\circ$, and $\Delta S^\circ$ for the expanded database are 0.95, 0.89, and 0.87, respectively. The results should improve predictions of RNA secondary structure.

The explosion of biological data in the genomics era has filled databanks with large amounts of genetic information. Understanding of these data and making correlations are vital for maximally advancing the fields of biology and medicine. This necessitates accurate methods in bioinformatics and computational chemistry. One important area that bioinformatics and computational chemistry address is the prediction of RNA secondary structure.

RNA participates in a variety of cellular functions involving gene expression and regulation. RNA typically folds in a hierarchical way.1,2 Base pairs form to generate motifs such as helices and loops. Higher order interactions between these features result in three-dimensional structures. On that basis, knowledge of secondary structure is critical for the prediction of tertiary structure. Secondary structure prediction algorithms utilizing experimental thermodynamic data4–9 have relied on nearest neighbor models.10–13 Finding regions of genome sequences that code for structured RNA often also relies on nearest neighbor models.14–16 Because RNA molecules and their reverse complements can fold similarly, the thermodynamics of GU pairs provides information about the reading direction because their complement, CA, forms less stable base pairs.17

Prediction of GU pairs is also important because they are the most common non-Watson–Crick pair and have functions in a wide variety of RNAs. For example, GU pairs are found within two helical regions and at the junction of a helix and multibranch loop in eukaryotic 5S rRNA.18,19 A GU pair in the third position of the acceptor stem in tRNAAla20 distorts helix geometry21 and is important in Escherichia coli for recognition by alanine aminocyl tRNA synthetase.22–24 Local helix geometry due to a conserved GU pair may also be important for binding of a yeast intron with hPrp8 or L32 protein.25,26 The U-rich tail of guide RNAs bind to a purine-rich region in unedited pre-mRNA to generate recurring 5′AGA/3′UUU motifs that may help RNA editing proteins bind to the major groove.27,28 The S′ leader of HIV-1 can switch between helices containing GU pairs to promote translation or packaging of its genome.29

GU pairs expose the exocyclic amino group of guanine in the minor groove, presenting a unique site for hydrogen bonding to facilitate function and molecular recognition. For example, a GU amino group at the splice site in the Tetrahymena thermophila group I intron helps bind and align the splice site30–32 and stabilize the transition state of the splicing reaction.33 Sponer et al. reported a common tertiary interaction involving a GU pair, where the exocyclic NH2 of the G and the 2′OH of the U form hydrogen bonds, respectively, with the 2′OH and carbonyl oxygen of a cytidine in a GC pair of another helix.34

GU pairs can be metal ion binding sites.35–39 Colmenarejo and Tinoco observed that Co(NH3)62+ preferably binds to 5′GG/3′UG and 5′GG/3′UU over 5′UG/3′GU pairs, whereas Mg(H2O)62+ binds tightest to 5′UG/3′GU.40 This preference may explain why 5′UG/3′GU is the most prevalent tandem GU...
motif in rRNA. The propensity for binding metal ions allows design of sequences that bind heavy metals to facilitate solving of X-ray structures.11,37,38

Prediction of GU pairs often relies on a nearest neighbor model for folding stability. The database of RNA sequences from which GU nearest neighbor parameters were derived12 is relatively small, however, compared to that for Watson–Crick nearest neighbors.13 To expand the database, optical melting experiments were carried out on 29 oligoribonucleotide duplexes. Linear regression analysis on the expanded database provides a revised set of individual nearest neighbor (INN) parameters,38 which are reported herein. The parameters provide stability increments for internal and single terminal GU pairs. Stability increments for additional terminal GU pairs have been reported by Nguyen and Schroeder.43

**MATERIALS AND METHODS**

**Design of Oligonucleotides.** Oligonucleotides were designed to expand the previous database12 to provide all possible combinations of base pair triplets containing GU pairs flanked by Watson–Crick pairs in different orientations (Table 1) and to have a substantial representation of each nearest neighbor containing a GU pair. An additional six sequences containing the 5′GG/3′UU doublet provided nine new representations for that motif, which had only one representation. Care was taken to select self-complementary sequences that do not favorably form alternative secondary structures, such as hairpins or loops.

**Synthesis and Purification of Oligoribonucleotides.** Sequences for the following oligoribonucleotide duplexes were taken to select self-complementary sequences that do not favorably form alternative secondary structures, such as hairpins or loops.

**UV Melting.** RNA duplexes with concentrations from 10^{-6}–10^{-3} M were melted in 0.5 mM Na₂EDTA, 1 M NaCl, and 20 mM sodium cacodylate, pH 7, which maintains a stable pH over a wide temperature range.35 Absorbance at 280 nm, typically from 15 to 80 °C, was measured on a Beckman Coulter DU 640 spectrophotometer.

**NMR Experiments.** Spectra were acquired on a Varian Inova 500 or 600 MHz spectrometer. The buffer for NMR was 80 mM NaCl, 18.8 mM NaH₂PO₄, 1.16 mM Na₂HPO₄, 0.02 mM NaCl, 60 mM NaCl, pH 6.0, to which 15 µL of D₂O was added to provide a lock signal. One-dimensional 1H spectra were acquired with the water 1H signal suppressed with a binomial 1:1 shaped pulse.46 Two-dimensional 1H–1H NOESY and 1H–2H TOCSY spectra were acquired with the water signal suppressed by a WATERGATE-type pulse sequence with flipback.47,48 Two-dimensional 1H–2H NOESY spectra for r(AUGCGU)₂ were also measured in D₂O.

Spectra were processed with NMRPipe49 and resonances were assigned with SPARKY.50 Proton chemical shifts were referenced to a temperature-dependent water chemical shift, δᵣ: 7.83 − T/96.9

\[ \delta = 7.83 - T/96.9 \]

where T is temperature in Kelvin.51 The internal reference standard for water was 2,2-dimethylsilapentate-5-sulfonic acid.

**Melting Data Analysis.** Melting curves for each duplex were fit to a two-state model with MeltWin 3.552 to derive values for ΔH° and ΔS°. The melting temperature, Tₘ, was plotted against ln(CT/α) to provide another measure of ΔH° and ΔS°:

\[ 1/T_M = (R / ΔH°) \times \ln(C_T / \alpha) + \Delta S° / ΔH° \]

Here R is the gas constant (1.987 cal K⁻¹ mol⁻¹), C_T is the total concentration of strands, and α is 1 for self-complementary duplexes and 4 for non-self-complementary duplexes. Sequences were added to the database if ΔH° values derived from averaging fits of melting curves agreed within 15% with those derived from eq 2, consistent with the two-state model.

**Linear Regression to Fit Nearest Neighbor Parameters.** Nearest neighbor thermodynamic parameters were obtained with a regression function reported by Xia et al.11 Matrix calculations were performed with R53 and independently verified with Mathematica 8.054 and Octave.55 All three software packages yielded nearly identical results.

Terms representing free energy contributions from non-GU nearest neighbors, that is, helix initiation (ΔGᵣ/init), symmetry (ΔGᵣ/sym), terminal AU pairs (ΔGᵣ/term AU), and Watson–Crick nearest neighbors (ΔGᵣ/ (WC NN)),11 were subtracted from the free energy found from the Tₘ⁻¹ vs ln(CT/α) plots (ΔGᵣ/duplex) to provide an experimental free energy attributable to the GU components of each duplex:

\[ ΔGᵣ/ (GU component) = \sum n_ij ΔGᵣ/ (GU NN) \]

\[ j \]

\[ = ΔGᵣ/ (duplex) - ΔGᵣ/ (sym) - ΔGᵣ/ (init) - m_ij \]

\[ \times ΔGᵣ/ (term AU) - \sum n_ij ΔGᵣ/ (WC NN) \]

\[ j \]

\[ (3) \]

where i and j are labels for each different duplex and INN parameter, respectively, NN stands for nearest neighbor parameter, and mᵢ is the number of terminal AU pairs. For example,

\[ ΔGᵣ/37/ (GU component) = ΔGᵣ/37 \left( \frac{CGGCGGUGUGGGG}{GCUUUGUGGGG} \right) - 3ΔGᵣ/37 \left( \frac{5'C3'}{3'G5'} \right) - ΔGᵣ/37/ (init) \]

\[ (4) \]
Here, $\Delta G_{\text{INN}}^o$ (GU component) contains four $5\text{‘}G\text{‘}G/3\text{‘}\text{C}$ nearest neighbors and two $5\text{‘}G\text{‘}G/3\text{‘}U$ nearest neighbors. Values for Watson–Crick nearest neighbors from Xia et al.\textsuperscript{11} were used because experimental measurements on 22 duplexes not included in the fitting by Xia et al.\textsuperscript{11} are predicted within experimental error (Supporting Information). Making the new GU parameters consistent with the Xia et al.\textsuperscript{11} parameters provides compatibility with loop parameters derived with Xia et al.\textsuperscript{11} nearest neighbor parameters and allows easy adoption by programs using those parameters.

Each experimental duplex $\Delta G_{\text{INN}}^o$ was given an error limit of $\pm 4\%$ to account for systematic errors unless the percent difference between parameters found from $T_{M}^{-1}$ vs $\ln(C_T/a)$ and averaged curve fits was greater. For the seven latter cases, this percent difference was doubled to provide an error limit. Error limits for $\Delta H^o$ were assumed to be $12\%$.\textsuperscript{11} The symmetry contribution, $0.43 \text{ kcal/mol}$ in $\Delta G_{\text{INN}}^o$, has no error\textsuperscript{56} and was therefore subtracted from $\Delta G_{\text{INN}}^o$ of self-complementary duplexes before calculating the error limit.

The GU component free energies were placed into $M \times 1$ matrix $G$, where $M$ is the number of duplexes.

$$G = S \cdot G_{\text{NN}}$$

(5)

$S$ is an $M \times N$ matrix containing the counts of each nearest neighbor doublet in a duplex, where $N$ is the number of GU nearest neighbor parameters being fit. $G_{\text{NN}}$ is an $N \times 1$ matrix that contains the nearest neighbor parameters to be derived from $G$ and $S$.

The general law of error propagation was used to calculate the variances for each duplex.\textsuperscript{55,58} Multiplication of both sides of eq 5 by an $M \times M$ matrix, $\sigma -1$, containing the variances in the diagonals yielded error-weighted matrices from which thermodynamic parameters were derived.

$$G_\sigma = \sigma^{-1}G = \sigma^{-1}S \cdot G_{\text{NN}} = S_G \cdot G_{\text{NN}}$$

(6)

The values in $G_{\text{NN}}$ are thus $S_{\sigma}^{-1}G_\sigma$. The variances of each INN parameter are obtained with singular value decomposition (SVD) (ref 11, Supporting Information). Nearest neighbor parameters for $\Delta H^o$ were found through the same process, and $\Delta S^o$ parameters were calculated from $\Delta S^o = (\Delta H^o - \Delta C^o)/T_M$.

Nearest neighbor parameters for Watson–Crick pairs were obtained from fitting published data for 112 duplexes, which included the 90 duplexes that Xia et al. previously fit, and 22 additional duplexes (Supporting Information). The symmetry contribution, if present, was subtracted from each thermodynamic parameter derived from the $T_{M}^{-1}$ vs $\ln(C_T/a)$ plot. Matrix calculations were carried out as described above to generate $\Delta G_{\text{INN}}^o$ and $\Delta H^o$ for each nearest neighbor parameter, with all three software packages yielding similar results.

The $F$-test was used to test the hypothesis that a least-squares model can fit the dependence of $G_\sigma$ on $S_\sigma$ and $G_{\text{NN}}$.\textsuperscript{59,60} If the $F$-value is larger than the critical $F$-value for $N$ and $N - v$ degrees of freedom at the $5\%$ significance level, where $N$ is the number of duplexes and $v$ the number of nearest neighbor parameters, or if the $p$-value is less than 0.05, then the hypothesis that there is a dependence of $G_\sigma$ on $G_{\text{NN}}$ may be accepted.\textsuperscript{60}

The paired $t$-test was used to evaluate the significance of the differences between predictions of thermodynamic properties with the updated parameters and those reported by Mathews et al.\textsuperscript{12} and the difference between experimental values and predictions by each set of nearest neighbor parameters. The difference between each pair of a set with $b$ values of a variable, $X_b$ before and after treatment is defined as $\mu(X_b) = \mu(X_b) - \mu(X_b)$, where $X_b$ represents the response of $X_b$ to treatment.\textsuperscript{51} The null hypothesis states that $\mu(X_b) = 0$. To test this and the alternative hypothesis that $\mu(X_b) \neq 0$, the mean and standard deviation of the difference between each block of values is found.

$$\bar{X}_D = \frac{\sum_{i=1}^{b}(X_{1,i} - X_{2,i})}{b}$$

(7)

$$S_D^2 = \frac{\sum_{i=1}^{b}[(X_{1,i} - X_{2,i}) - \bar{X}_D]^{2}}{b - 1}$$

(8)

A $t$-ratio is defined as

$$t = \frac{\bar{X}_D}{S_D/\sqrt{b}}$$

(9)

If the $t$-ratio is greater than $t$-value for $(b - 1)$ degrees of freedom or less than its negative, then the null hypothesis is rejected at the 0.05 significance level.

For example, in using the paired $t$-test to evaluate how well experimental $\Delta G_{\text{INN}}^o$’s are predicted by nearest neighbor parameters, $b$ is the number of duplexes whose $\Delta G_{\text{INN}}^o$’s are being tested and $X_D$ is the difference between the predicted and experimental $\Delta G^o$ for each sequence.

The probability density function (PDF), $f(t)$, of the Student’s $t$-distribution was used as a measure of how significantly a given INN parameter contributes to the model,\textsuperscript{51,59} with smaller values of $f(t)$ indicating greater contribution,

$$f(t) = \frac{\Gamma(1/2(r+1))}{(\pi r)^{1/2}\Gamma(1/2r)[1 + t^2/r]^{-(r+1)/2}}$$

(10)

where $\Gamma$ is the gamma function, $r = N - v$ degrees of freedom and $t = \Delta G^o_{\text{INN}}(\text{NN})/\sigma(\text{NN})$, that is, the quotient of the free energy of the INN parameter over the estimate of its error. Calculations were carried out with R\textsuperscript{55} using the anova and t-test functions, and the critical $t$-value was determined with the qt function in R.

#### RESULTS

Table 2 lists results for duplexes in the database used for determination of nearest neighbor parameters for GU pairs. Most of the duplexes are six to eight base pair in length and have melting temperatures in the $30$–$70 \, ^\circ\text{C}$ range. For the 29 new duplexes reported here, the average difference between $\Delta G_{\text{INN}}^o$, $\Delta H^o$, and $\Delta S^o$ derived from $T_{M}^{-1}$ vs $\ln(C_T/a)$ plots and averaged curve fits are $2\%$, $7\%$, and $8\%$, respectively. Three duplexes, r(AUGCUU)$_2$, r(AUGCGU)$_2$, and r(GUCGUAC)$_2$, with $T_{M}$’s less than $25 \, ^\circ\text{C}$ that were included in the database of Mathews et al.\textsuperscript{12} were omitted from the new database. Determination of thermodynamics from optical melting curves is difficult when the $T_{M}$ is less than $25 \, ^\circ\text{C}$.

With the Exclusion of the 5’GGUC/3’ CUG Motif, the Experimental Results Can Be Fit to a Nearest Neighbor Model. The results in Table 2 were fit to a nearest neighbor model for GU pairs after subtracting contributions from Watson–Crick nearest neighbors (eq 3). This method avoids conflating thermodynamic parameters for Watson–Crick pairs with the idiosyncrasies of GU pairs. Published thermodynamics for duplexes with all Watson–Crick pairs (Supporting Information) were used to test published parameters for
| Sequence | $T_m$ (°C) | $\Delta G^°$ (kcal/mol) | $\Delta H^°$ (kcal/mol) | $\Delta S^°$ (eu) | $T_m$ (°C) | $\Delta G^°$ (kcal/mol) | $\Delta H^°$ (kcal/mol) | $\Delta S^°$ (eu) | Ref. |
|----------|------------|--------------------------|------------------------|-----------------|------------|--------------------------|------------------------|-----------------|------|
| CCGGCGC | 7.44       | 47.00                    | 128.4                  | 38.6            | 7.46       | 47.00                    | 128.4                  | 38.6            | this work |
| CCGGGGC | 6.71       | 47.50                    | 126.2                  | 38.9            | 6.73       | 47.50                    | 126.2                  | 38.9            | this work |
| CCGGGGC | 6.94       | 48.00                    | 129.4                  | 39.2            | 6.96       | 48.00                    | 129.4                  | 39.2            | this work |
| CCGGGGC | 7.21       | 48.50                    | 132.6                  | 39.5            | 7.23       | 48.50                    | 132.6                  | 39.5            | this work |
| CCGGGGC | 7.44       | 49.00                    | 135.8                  | 39.8            | 7.46       | 49.00                    | 135.8                  | 39.8            | this work |
| CCGGGGC | 7.66       | 49.50                    | 139.0                  | 40.1            | 7.68       | 49.50                    | 139.0                  | 40.1            | this work |
| CCGGGGC | 7.88       | 50.00                    | 142.2                  | 40.4            | 7.90       | 50.00                    | 142.2                  | 40.4            | this work |
| CCGGGGC | 8.09       | 50.50                    | 145.4                  | 40.7            | 8.11       | 50.50                    | 145.4                  | 40.7            | this work |
| CCGGGGC | 8.29       | 51.00                    | 148.6                  | 41.0            | 8.31       | 51.00                    | 148.6                  | 41.0            | this work |
| CCGGGGC | 8.48       | 51.50                    | 151.8                  | 41.3            | 8.50       | 51.50                    | 151.8                  | 41.3            | this work |
| CCGGGGC | 8.66       | 52.00                    | 155.0                  | 41.6            | 8.68       | 52.00                    | 155.0                  | 41.6            | this work |
| CCGGGGC | 8.84       | 52.50                    | 158.2                  | 41.9            | 8.86       | 52.50                    | 158.2                  | 41.9            | this work |
| CCGGGGC | 9.01       | 53.00                    | 161.4                  | 42.2            | 9.03       | 53.00                    | 161.4                  | 42.2            | this work |
| CCGGGGC | 9.18       | 53.50                    | 164.6                  | 42.5            | 9.20       | 53.50                    | 164.6                  | 42.5            | this work |
| CCGGGGC | 9.34       | 54.00                    | 167.8                  | 42.8            | 9.36       | 54.00                    | 167.8                  | 42.8            | this work |
| CCGGGGC | 9.50       | 54.50                    | 171.0                  | 43.1            | 9.52       | 54.50                    | 171.0                  | 43.1            | this work |
| CCGGGGC | 9.66       | 55.00                    | 174.2                  | 43.4            | 9.68       | 55.00                    | 174.2                  | 43.4            | this work |
| CCGGGGC | 9.81       | 55.50                    | 177.4                  | 43.7            | 9.83       | 55.50                    | 177.4                  | 43.7            | this work |
| CCGGGGC | 9.96       | 56.00                    | 180.6                  | 44.0            | 9.98       | 56.00                    | 180.6                  | 44.0            | this work |

**Table 2. Thermodynamic Parameters for Duplex Formation in 1 M NaCl**

*a* $T_m$ vs log $C_T$ average of curve fits predicted

*b* $\Delta G^°$, $\Delta H^°$, $\Delta S^°$
Table 2. continued

| sequence<sup>b</sup> | $\Delta G^\circ_{\text{ref}}$ (kcal/mol) | $\Delta H^\circ$ (kcal/mol) | $\Delta S^\circ$ (eu) | $T_M^{-1}$ | $\Delta G^\circ_{\text{ref}}$ (kcal/mol) | $\Delta H^\circ$ (kcal/mol) | $\Delta S^\circ$ (eu) | $T_M^{-1}$ | ref |
|---------------------|-----------------|-----------------|-----------------|--------|-----------------|-----------------|-----------------|--------|-----|
| GGGUGGCC           | 9.72            | 73.40           | 206.9           | 55.0   | 9.75            | 74.20           | 208.0           | 55.0   | 104 |
| GGGUGGCC           | 13.10           | 87.20           | 238.8           | 65.9   | 13.30           | 90.30           | 248.4           | 65.4   | 203 |
| GGGUGGCC           | 8.30            | 78.30           | 225.9           | 47.6   | 8.30            | 77.40           | 222.8           | 47.7   | 207 |
| GUAUUGCG           | 7.30            | 50.30           | 138.7           | 47.4   | 7.40            | 57.90           | 162.8           | 46.7   | 201 |
| GUGGCGU/GGU        | 15.00           | 98.00           | 261.3           | 66.9   | 14.20           | 83.30           | 222.7           | 68.6   | 204 |
| GUGCG/UGU          | 6.05            | 69.10           | 203.3           | 38.7   | 6.08            | 64.70           | 188.9           | 39.0   | 203 |
| GUC/UGU            | 7.70            | 70.00           | 201.0           | 46.3   | 7.70            | 68.70           | 196.8           | 46.3   | 204 |
| UACGGCGU           | 9.70            | 51.70           | 135.4           | 63.1   | 10.00           | 54.70           | 144.1           | 63.6   | 207 |
| UACGGCGU           | 6.44            | 62.30           | 180.1           | 41.0   | 6.55            | 53.10           | 150.1           | 41.8   | 205 |
| UACGGCGU           | 8.40            | 46.90           | 124.2           | 56.0   | 8.60            | 45.00           | 117.1           | 58.7   | 206 |
| UACGGCGU           | 10.40           | 63.80           | 172.3           | 61.6   | 10.60           | 67.30           | 182.7           | 61.7   | 208 |
| UACGGCGU           | 6.20            | 44.60           | 124.0           | 40.5   | 6.20            | 49.20           | 138.9           | 40.1   | 209 |
| CAGGCGGAC/         | 9.43            | 98.95           | 288.6           | 46.4   | 9.39            | 95.19           | 276.7           | 46.6   | 210 |
| CAGGCGGAC/         | 12.31           | 77.02           | 208.6           | 66.2   | 11.59           | 66.84           | 178.2           | 67.1   | 211 |
| CAGGCGGAC/         | 8.70            | 92.33           | 269.7           | 47.5   | 8.54            | 84.54           | 245.0           | 47.9   | 212 |
| CAGGCGGAC/         | 11.60           | 87.90           | 246.0           | 55.9   | 11.70           | 90.10           | 252.7           | 55.9   | 213 |
| CAGGCGGAC/         | 6.99            | 76.76           | 225.0           | 42.6   | 7.00            | 76.35           | 223.5           | 42.6   | 214 |
| CAGGCGGAC/         | 7.88            | 90.52           | 266.5           | 44.7   | 7.77            | 84.47           | 243.7           | 44.9   | 215 |
| CAGGCGGAC/         | 8.78            | 88.20           | 256.2           | 48.3   | 8.49            | 80.30           | 231.4           | 48.5   | 216 |
| CAGGCGGAC/         | 11.55           | 101.66          | 290.5           | 56.0   | 11.65           | 101.34          | 289.2           | 56.4   | 217 |
| CAGGCGGAC/         | 14.76           | 102.42          | 282.6           | 64.1   | 14.18           | 94.72           | 259.7           | 64.2   | 218 |
| CAGGCGGAC/         | 10.15           | 97.81           | 282.7           | 51.8   | 9.97            | 92.86           | 267.3           | 52.0   | 219 |
| CAGGCGGAC/         | 8.82            | 86.57           | 250.6           | 48.9   | 8.72            | 82.81           | 238.9           | 48.9   | 220 |
| CAGGCGGAC/         | 9.83            | 93.86           | 270.9           | 51.3   | 9.40            | 83.92           | 240.3           | 51.4   | 221 |
| CAGGCGGAC/         | 9.87            | 98.93           | 280.7           | 48.1   | 9.79            | 93.52           | 270.0           | 48.2   | 222 |
| CAGGCGGAC/         | 13.59           | 115.71          | 329.3           | 59.8   | 13.56           | 113.36          | 321.8           | 60.2   | 223 |
| CAGGCGGAC/         | 4.78            | 62.63           | 186.4           | 32.8   | 4.60            | 72.40           | 218.6           | 32.5   | 224 |
| CAGGCGGAC/         | 8.90            | 58.60           | 160.1           | 55.0   | 9.20            | 66.70           | 185.6           | 54.2   | 225 |
| CAGGCGGAC/         | 8.60            | 69.60           | 196.7           | 50.4   | 8.50            | 66.90           | 188.5           | 50.5   | 226 |
| UCCAGGACG/         | 15.32           | 93.46           | 252.0           | 69.2   | 15.42           | 94.49           | 254.9           | 69.2   | 227 |

<sup>a</sup>Duplexes with a 5′-GUC/3′-CGG motif are not listed because they were not used in fitting nearest neighbor parameters and no new sequences were measured.<sup>b</sup>Listed in order of length of the oligoribonucleotide and then alphabetically for sequences of the same length. All non-self-complementary sequences have a slash and only one sequence shown. GU base pairs are underlined.<sup>c</sup>Calculated at a total strand concentration of 1 × 10^{-4} M.
Table 3. INN Parameters for Canonical Base Pairs in 1 M NaCl without a Separate Parameter for Terminal GU Pairs

| INN | INN counts | ΔG°,37 (kcal/mol) | ΔG°,37 (kcal/mol) | ΔH° (kcal/mol) | ΔH° (kcal/mol) | ΔS° (eu) | ΔS° (eu) |
|-----|------------|------------------|------------------|--------------|--------------|----------|----------|
| 5'AA' | 2 | -0.93 (-0.96) | 0.03 (0.03) | -6.82 (-7.09) | 0.79 (0.77) | -19.0 (-19.8) | 2.5 (2.4) |
| 3'UU' | 2 | -1.10 (-1.09) | 0.08 (0.07) | -9.38 (-9.11) | 1.68 (1.56) | -26.7 (-25.8) | 5.2 (4.8) |
| 5'AU' | 2 | -1.33 (-1.39) | 0.09 (0.08) | -7.69 (-8.50) | 2.02 (1.86) | -20.5 (-22.9) | 6.3 (5.7) |
| 3'AS' | 2 | -2.08 (-2.07) | 0.06 (0.06) | -10.48 (-10.90) | 1.24 (1.15) | -27.1 (-28.5) | 3.8 (3.5) |
| 5'UC' | 2 | -2.11 (-2.11) | 0.07 (0.06) | -10.44 (-11.03) | 1.28 (1.18) | -26.9 (-28.8) | 3.9 (3.6) |
| 3'GC' | 2 | -2.24 (-2.27) | 0.06 (0.05) | -11.40 (-11.98) | 1.23 (1.12) | -29.5 (-31.3) | 3.9 (3.5) |
| 5'GC' | 2 | -2.35 (-2.39) | 0.06 (0.05) | -12.44 (-13.21) | 1.20 (1.05) | -32.5 (-34.9) | 3.7 (3.2) |
| 3'GC' | 2 | -2.36 (-2.38) | 0.09 (0.09) | -10.64 (-10.88) | 1.65 (1.54) | -26.7 (-27.4) | 5.0 (4.7) |
| 5'GC' | 2 | -3.26 (-3.31) | 0.07 (0.06) | -13.39 (-14.18) | 1.24 (1.07) | -32.7 (-35.0) | 3.8 (3.3) |
| 3'GC' | 2 | -3.42 (-3.46) | 0.08 (0.07) | -14.88 (-16.04) | 1.58 (1.33) | -36.9 (-40.6) | 4.9 (4.0) |
| 5'GU' | 8 | 0.72 | 0.19 | -13.83 | 4.21 | -46.9 | 13.0 |
| 3'UG' | 9 | -0.25 | 0.16 | -17.82 | 3.75 | -56.7 | 11.6 |
| 5'GG' | 22 | -0.35 | 0.08 | -3.96 | 1.73 | -11.6 | 5.3 |
| 3'US' | 18 | -0.39 | 0.09 | -0.96 | 1.80 | -1.8 | 5.5 |
| 5'UU' | 26 | -0.51 | 0.08 | -10.38 | 1.79 | -31.8 | 5.5 |
| 3'AG' | 10 | -0.57 | 0.19 | -12.64 | 4.01 | -38.9 | 12.3 |
| 5'UG' | 24 | -0.90 | 0.08 | -7.39 | 1.65 | -21.0 | 5.1 |
| 3'UG' | 26 | -1.25 | 0.09 | -5.56 | 1.68 | -13.9 | 5.1 |
| 5'CU' | 21 | -1.77 | 0.09 | -9.44 | 1.76 | -24.7 | 5.4 |
| 3'GC' | 24 | -1.80 | 0.09 | -7.03 | 1.75 | -16.8 | 5.4 |
| 5'GC' | 25 | -2.15 | 0.10 | -11.09 | 1.78 | -28.8 | 5.4 |
| 3'GC' | 3 | -4.12 | 0.54 | -30.80 | 8.87 | -86.0 | 23.7 |

Other Nearest Neighbor Parameters

initiation° 4.09 (4.23) 0.22 (0.20) 3.61 (6.40) 4.12 (3.56) -1.5 (6.99) 12.7 (10.9)
terminal AU penalty° 0.45 (0.43) 0.04 (0.04) 3.72 (3.85) 0.83 (0.77) 10.5 (11.04) 2.6 (2.4)
symmetry° 0.43 0.0 0 -1.4 0

“Values for ΔS° were derived from ΔS° = (ΔH° - ΔG°,37)/310.15. °Ref 12. †Ref calculated from nearest neighbor model. ‡Ref calculated from Watson-Crick nearest neighbors.

Watson-Crick pairs.11 Fitting the expanded database of 112 duplexes gave INN parameters within error of the values reported by Xia et al.11 (Table 3). Most free energy parameters did not change by more than 0.05 kcal/mol at 37 °C. Consequently, the GU component values were calculated from eq 3 by subtracting the previously published Watson-Crick thermodynamic values11 so that the GU parameters are consistent with the widely used Watson-Crick parameters. For fitting GU parameters, duplexes containing the motif, 5′GGUC/3′CUGG, were excluded from the regression due to its poor fit in the nearest neighbor model.12 For the other 70 duplexes, 12 GU INN parameters were initially derived by
Table 4. Probability Density Function (PDF) of Student’s t-Distribution for $\Delta G_{37}^o$ and $\Delta H^o$ for Each GU INN Motif without a Separate Parameter for Terminal GU Pairs$^a$

| motif      | PDF, $\Delta G_{37}^o$ | PDF, $\Delta H^o$ | motif      | PDF, $\Delta G_{37}^o$ | PDF, $\Delta H^o$ |
|------------|-------------------------|-------------------|------------|-------------------------|-------------------|
| 5'GU$^3$   | 5.8 × 10$^{-4}$         | 2.6 × 10$^{-3}$   | 5'AU$^3$   | 4.7 × 10$^{-16}$        | 6.1 × 10$^{-5}$   |
| 3'UG$^3$   | 1.2 × 10$^{-1}$         | 2.4 × 10$^{-5}$   | 3'UG$^3$   | 4.9 × 10$^{-20}$        | 2.4 × 10$^{-3}$   |
| 5'GG$^3$   | 8.7 × 10$^{-5}$         | 3.1 × 10$^{-2}$   | 5'CG$^3$   | 1.8 × 10$^{-27}$        | 2.7 × 10$^{-6}$   |
| 3'UU$^3$   | 1.0 × 10$^{-4}$         | 3.4 × 10$^{-1}$   | 3'GG$^3$   | 7.4 × 10$^{-28}$        | 2.8 × 10$^{-4}$   |
| 5'AG$^3$   | 5.9 × 10$^{-8}$         | 5.3 × 10$^{-7}$   | 3'CU$^3$   | 1.6 × 10$^{-29}$        | 1.0 × 10$^{-7}$   |
| 3'AG$^3$   | 5.6 × 10$^{-3}$         | 3.7 × 10$^{-3}$   | 5'GU$^3$   | 4.9 × 10$^{-8}$         | 2.4 × 10$^{-5}$   |
| 3'GG$^3$   | 1.2 × 10$^{-1}$         | 2.4 × 10$^{-5}$   | 3'CU$^3$   | 7.4 × 10$^{-28}$        | 2.8 × 10$^{-4}$   |

$^a$Refs 11 and 59. The S'GGUC/3CUGG motif is not included.

A-form conformation.$^{62}$ For r(AUGCGU)$_2$, however, the presence of broad on-diagonal peaks and exchange cross peaks in the aromatic region of the 2D spectra and of more imino resonances in a 1D spectrum than the number of imino protons in the sequence indicates the presence of alternate conformations. The presence of broad on-diagonal peaks, particularly for A1H8 and H2, in the aromatic region of the 2D spectra for r(AUGCGU)$_2$ also suggests multiple conformations at 0 °C.

Table 5 also lists duplexes that do not melt in a two-state manner. There are many possible reasons for this.$^{56,63-65}$ The average difference between experimental and predicted $T_m$ for these sequences is 10.0 °C, while the predicted free energy is, on average, within 1.33 kcal/mol of the experimental free energy (Table 5). Evidently, the INN model may provide useful predictions for non-two-state sequences even though $\Delta H^o$ from the van’t Hoff equation is erroneous.$^{64}$

The Expanded Database Improves Predictions of Duplex Stability. Using the previous parameters,$^{12}$ the correlation coefficients between experimental values for $\Delta G_{37}^o$, $\Delta H^o$, and $\Delta S^o$ and those predicted for the 70 duplexes in Table 2 are 0.89, 0.86, and 0.85, respectively. Comparisons of the values of $\Delta G_{37}^o$, $\Delta H^o$, and $\Delta S^o$ of the 70 duplexes as predicted with the previous parameters,$^{12}$ and those in Table 3 yielded, respectively, means of the differences of $-0.36$ kcal/mol, $-1.75$ kcal/mol, and $-4.5$ eu. The paired $t$-test gives $t$-values of $-3.386$, $-2.528$, and $-2.257$, respectively, which have absolute magnitudes greater than 1.995, indicating that the two sets of parameters differ with a significance level of 0.05 for 69 degrees of freedom.$^{61}$ Furthermore, the respective $p$-values of $1.2 \times 10^{-3}$, $1.4 \times 10^{-2}$, and $2.7 \times 10^{-2}$ are less than 0.05. This again indicates that the new parameters predict the thermodynamics of RNA duplexes significantly differently from those published previously.$^{12}$

Using the set of GU parameters in Table 3, the correlation coefficients between experimental values for $\Delta G_{37}^o$, $\Delta H^o$, and $\Delta S^o$ and those predicted for the 70 duplexes in Table 2 are 0.95, 0.89, and 0.87, respectively. Comparison of experimental values of $\Delta G_{37}^o$, $\Delta H^o$, and $\Delta S^o$ of the 70 duplexes and those predicted with the set of GU parameters in Table 3 yielded means of differences of $-0.05$ kcal/mol, 0.58 kcal/mol, and 2.1 eu and $t$-values of $-0.544$, $0.590$, and 0.686, respectively, which have absolute magnitudes less than 1.995. The corresponding $p$-values of 0.59, 0.56, and 0.50 are greater than 0.05. These results show that the thermodynamic properties predicted with
Table 5. Thermodynamic Parameters for Formation of Duplexes Not Included in Fitting Nearest Neighbor Parameters

| sequence b | $-\Delta G^\circ_{\text{seq}}$ (kcal/mol) | $-\Delta H^\circ$ (kcal/mol) | $-\Delta S^\circ$ (eu) | $T_m^*$ (°C) | $-\Delta G^\circ_{\text{seq}}$ (kcal/mol) | $-\Delta H^\circ$ (kcal/mol) | $-\Delta S^\circ$ (eu) | $T_m^*$ (°C) | ref |
|------------|---------------------------------|-----------------|-----------------|-------------|---------------------------------|-----------------|-----------------|-------------|----|
| AGGCUU     | 4.07 ± 0.05                      | 37.10           | 106.6           | 24.2        | Sequences with $T_m < 25^\circ$ C | 4.24            | 34.60           | 97.9         | 24.7         | 2.24            | 30.63           | 91.5         | 5.8         | 104 |
| AGGCUUp    | 4.22 ± 0.05                      | 24.50           | 65.4            | 19.5        | Non-Two-State Sequences         | 4.37            | 24.10           | 63.6         | 21.1         | 2.30            | 29.73           | 88.5         | 5.2         | 101 |
| GUCAGCAC/  | 4.20 ± 0.05                      | 47.30 ± 9.01    | 139.1 ± 30.1    | 22.2        |                                  | 3.80 ± 0.48     | 54.20 ± 11.09   | 162.3 ± 36.1 | 22.2         | 3.66            | 51.98           | 155.7        | 21.0        | this work |
| AGGCGUA     | 5.36                            | 49.70           | 143.0           | 35.0        |                                  | 5.54            | 36.10           | 98.5         | 35.8         | 2.32            | 24.63           | 71.9         | -0.1        | 104 |
| CAGG/GUC/   | 7.70 ± 0.14                      | 49.10 ± 4.46    | 133.5 ± 14.2    | 44.3        |                                  | 8.10 ± 0.14     | 60.00 ± 11.18   | 167.4 ± 35.6 | 45.1         | 3.52            | 59.57           | 173.0        | 33.8        | this work |
| GUGUGUGU/   | 7.93 ± 0.18                      | 56.21 ± 4.47    | 155.7 ± 14.4    | 44.9        |                                  | 8.05 ± 0.18     | 69.80 ± 6.64    | 199.1 ± 18.1 | 43.9         | 7.67            | 55.73           | 154.9        | 43.6        | this work |
| GGGUGACC    | 5.94                            | 65.80           | 193.1           | 38.3        |                                  | 5.99            | 49.60           | 140.6        | 39.0         | 6.54            | 61.72           | 177.9        | 41.4        | 104 |
| GAGCGUC                 | 6.57                            | 73.10           | 214.4           | 40.9        |                                  | 6.64            | 50.70           | 142.8        | 41.6         | 8.14            | 56.33           | 155.3        | 51.3        | 104 |
| GAAGGGCGGAG/   | 9.52 ± 0.24                      | 136.60 ± 11.90  | 409.7 ± 37.7    | 43.9        |                                  | 8.28 ± 0.56     | 59.27 ± 6.93    | 164.4 ± 21.0 | 46.4         | 10.88           | 111.46          | 324.3        | 49.6        | this work |
| GCUGUCGAG   | 13.22 ± 0.35                     | 141.60 ± 7.01   | 414.0 ± 21.5    | 54.4        |                                  | 10.39 ± 0.34    | 82.58 ± 14.35   | 232.7 ± 45.3 | 55.8         | 13.92           | 129.51          | 372.6        | 58.1        | this work |

The Nearest Neighbor Model Is Not Perfect. While the new INN parameters are not significantly different from the previous INN parameters used for constructing the model, the expanded database and the refined modeling have improved the predictions. For example, the calculated $T_m$ for the duplex GGCG/UGCGCU/GCUCUAGC is 56.4°C, which is closer to the experimental $T_m$ of 54.0°C. Moreover, the calculated $-\Delta G$, $-\Delta H$, and $-\Delta S$ values are in better agreement with the experimental values. This improvement suggests that the expanded database and the refined modeling provide more accurate predictions.

Figure 1. 1D and 2D nuclear Overhauser effects (NOEs) of GGGCG/GGCUC from 0 to 45°C. The NOEs were acquired at a temperature of 37°C.

No pattern is evident for the GCAU/UGCG/AGCG/CGAC/UCAC/CGCA/UCUA pairs with predicted $T_m$ values within 10°C of the experimental $T_m$. For these pairs, the predicted $-\Delta G$, $-\Delta H$, and $-\Delta S$ values are also in good agreement with the experimental values. This suggests that the expanded database and the refined modeling provide more accurate predictions for these pairs.

The expanded database provides improved modeling of the thermodynamics of GU pairs. The new INN parameters are not significantly different from the previous INN parameters used for constructing the model, but the expanded database and the refined modeling have improved the predictions. For example, the calculated $T_m$ for the duplex GGCG/UGCGCU/GCUCUAGC is 56.4°C, which is closer to the experimental $T_m$ of 54.0°C. Moreover, the calculated $-\Delta G$, $-\Delta H$, and $-\Delta S$ values are in better agreement with the experimental values. This improvement suggests that the expanded database and the refined modeling provide more accurate predictions. For these pairs, the predicted $T_m$ values are also in good agreement with the experimental $T_m$. This suggests that the expanded database and the refined modeling provide more accurate predictions for these pairs.

Additionally, the expanded database provides improved modeling of the thermodynamics of GU pairs. The new INN parameters are not significantly different from the previous INN parameters used for constructing the model, but the expanded database and the refined modeling have improved the predictions. For example, the calculated $T_m$ for the duplex GGCG/UGCGCU/GCUCUAGC is 56.4°C, which is closer to the experimental $T_m$ of 54.0°C. Moreover, the calculated $-\Delta G$, $-\Delta H$, and $-\Delta S$ values are in better agreement with the experimental values. This improvement suggests that the expanded database and the refined modeling provide more accurate predictions. For these pairs, the predicted $T_m$ values are also in good agreement with the experimental $T_m$. This suggests that the expanded database and the refined modeling provide more accurate predictions for these pairs.

Figure 2. 2D nuclear Overhauser effects (NOEs) of GGGCG/GGCUC from 0 to 45°C. The NOEs were acquired at a temperature of 37°C.

No pattern is evident for the GCAU/UGCG/AGCG/CGAC/UCAC/CGCA/UCUA pairs with predicted $T_m$ values within 10°C of the experimental $T_m$. For these pairs, the predicted $-\Delta G$, $-\Delta H$, and $-\Delta S$ values are also in good agreement with the experimental values. This improvement suggests that the expanded database and the refined modeling provide more accurate predictions for these pairs. It is evident that the expanded database provides improved modeling of the thermodynamics of GU pairs. The new INN parameters are not significantly different from the previous INN parameters used for constructing the model, but the expanded database and the refined modeling have improved the predictions. For example, the calculated $T_m$ for the duplex GGCG/UGCGCU/GCUCUAGC is 56.4°C, which is closer to the experimental $T_m$ of 54.0°C. Moreover, the calculated $-\Delta G$, $-\Delta H$, and $-\Delta S$ values are in better agreement with the experimental values. This improvement suggests that the expanded database and the refined modeling provide more accurate predictions. For these pairs, the predicted $T_m$ values are also in good agreement with the experimental $T_m$. This suggests that the expanded database and the refined modeling provide more accurate predictions for these pairs.
The expanded database allows preliminary testing of models beyond the nearest neighbor model. For example, terminal GU pairs could be considered separately and a base pair triplet model used for internal GU pairs. Comparison of measured values of ΔG°37 for terminal GU pairs with those predicted from the parameters in Table 3 give a standard deviation within ±0.3 kcal/mol at 37 °C (Supporting Information). For the 16 triplets, SWGY/3XUZ, with WX and YZ as Watson–Crick pairs, 12 measured ΔG°37(GU component) values are within 0.5 kcal/mol of the predicted values and the others within 1.0 kcal/mol (Supporting Information). The nearest neighbor model is apparently a reasonable approximation, and considerably more data would be required to develop a triplet model.

One clear exception to the nearest neighbor model is multiple terminal GU pairs. Thus, the parameters in Table 3 cannot be used beyond the first terminal GU pair at a helix end. Parameters for additional terminal GU pairs have been published by Nguyen and Schroeder.

Imino Proton NMR Spectra of Several Duplexes Are Consistent with the Expected Base Pairing. To check for expected base pair, NMR imino proton spectra were measured for 12 duplexes. All had chemical shifts from 10 to 15 ppm (Figure 2(126,337),(916,353)). Chemical shifts for GH1 and UH3 of GU pairs were relatively upfield (10–12 ppm), consistent with expectations. Chemical shifts for UH3 in AU pairs resonated between 13 and 15 ppm and GH1 in GC pairs resonated from 12 to 13.5 ppm, as expected. The absence of an imino peak for a terminal base pair in r(CUGCUAG)2, indicates exchange with water. The G3-H1 and U7-H3 resonances of r(CUUGAUUCUAG)2 appear to overlap, as evident by the presence of a single large peak. These chemical shift signatures show that the RNA sequences form the expected duplexes.

**DISCUSSION**

GU pairs are the most common non-Watson–Crick base pairs in RNA structures. Thus, the thermodynamics of GU pairs are important for finding regions of RNA that are structured predicting the secondary structure or determining structure on the basis of chemical modification and/or NMR data. GU pairs can serve as binding sites for proteins or metal ions and participate in tertiary interactions. Thus, a better characterization of the thermodynamic properties of GU pairs can improve prediction of secondary and tertiary structure and help predict binding sites for metal ions and target sites for therapeutics. For example, GU pairs in group I introns can bind cations, including Mg2+, Co3+, and Os3+ Divalent metal ion binding by GU pairs, which have greater negative potential in the major groove than other base pairs, was postulated as important for activating RNA catalysis. Divalent ions that interact with a GU pair help catalyze splicing by group I and group II introns and cleavage by HDV ribozyme. Metal ion binding with RNA neutralizes negative potential, which may promote higher order RNA folding. The SGG/3′UU and S GU/3′UG motifs particularly contain greater negative potential in the major groove than their Watson–Crick counterparts.

The Database of Sequences for Determining GU Thermodynamic Parameters Was Expanded. Not including sequences containing the S′GUUC/3′CUUG motif, the database in Table 2 expands from 35 to 70 the duplexes used to fit nearest neighbor parameters for GU pairs. This expansion includes published data not included in the original database along with 29 new measurements (Table 2). Two of the original 35 duplexes were removed from the database because their melting temperatures were below 25 °C, which makes it difficult to analyze the melting curves. A third duplex, r(AUCUAGGU)ₙ, was omitted because two-state melting could not be confirmed. The expanded database contains GU pairs flanked by Watson–Crick pairs in all possible orientations. The new set of GU INN parameters were obtained with consideration for propagated errors from experiment and from Watson–Crick nearest neighbor parameters. Errors for the free energies of individual nearest neighbors were less than 0.2 kcal/mol for tandem GU pairs and 0.1 kcal/mol for other GU motifs. The S′GG/3′UU motif, which was previously represented by a single sequence, was added to the fitting. The favorable free energy of −0.25 ± 0.16 kcal/mol for S′GG/3′UU is in better agreement with the value of −0.5 kcal/mol used by Mathews et al. to optimize secondary structure prediction than with the previous single experimental measurement of 0.47 kcal/mol.

GU Pairs Are Generally Less Stable than GC and AU Pairs. The free energies of formation for many of the duplexes with GU pairs (Table 2) can be compared with the free energies when the U or G of the GU pairs is replaced with a C or A, respectively, to form GC or AU pairs (Table 6). Because many of the latter duplexes terminated with a 3′ phosphate, the comparisons assume that the 3′ phosphate has negligible effect on ΔG°37 at 1 M NaCl. Duplexes containing GC pairs in place of GU pairs are more stable at 37 °C by 1.8 ± 0.8 kcal/mol per GU pair (Table 6). This is presumably due to the presence of an additional hydrogen bond in GC pairs and unfavorable backbone distortion due to GU pairs. Terminal substitutions all have a less than average effect while internal substitutions have a larger than average effect, as expected if backbone distortion is less important for a terminal GU.

The effect of replacing GU with GC pairs can be compared to replacing AU pairs with GC pairs (Supporting Information). On average, replacing an AU pair with a GC pair stabilized a duplex by 1.5 ± 0.4 kcal/mol per AU pair. In this case, there was no apparent difference between terminal and internal substitutions. Duplexes containing AU pairs in place of GU pairs are more stable at 37 °C by 0.6 ± 0.7 kcal/mol per GU pair (Table 6). While the difference is zero within the standard deviation, in only 2 of 11 cases is the GU duplex more stable than the AU duplex and in both cases the difference is within the experimental error of 4%.

Unlike terminal AU pairs, no penalty for terminal GU pairs is required to account for base pair composition. The terminal AU penalty of 0.45 kcal/mol at 37 °C was considered to account for numbers of base pairing hydrogen bonds. Thus, the penalty for terminal GU pairs was assumed to be equal to that of AU pairs, consistent with wobble GU pairs at the end of a helix having two hydrogen bonds. When the terminal GU parameter was included in the reparameterization of GU nearest neighbor thermodynamic parameters, the free energy of each nearest neighbor parameter differed by no more than 0.01 kcal/mol from that calculated without it (Table 2 and Supporting Information). The lack of a terminal GU penalty may arise from the flexibility of a terminal GU pair which allows optimization of
hydrogen bonding and stacking interactions without incurring the energetic penalty associated with an interior GU distorting the backbone. Thus, flexibility of terminal GU’s may compensate for the difference between the free energy of formation of two and three hydrogen bonds in GU and GC pairs, respectively.

Figure 2. 1D imino NMR spectra for some RNA duplexes with GU pairs. Spectra were acquired at 0 °C for r(AGGCUU) and r(UUGCGU); 1 °C for r(UCCGUAA); 5 °C for others.
Tandem GU Pairs Have Structural Features That Correlate with Their Thermodynamic Properties. With the exception of $S'G\text{GU}/3'C\text{UG}$, the $S'U/G\text{U}3'G$ motif is more stable than $S'G/U/3'U$ (Table 3). Available structures show that $S'U/G\text{U}3'U$ contains interstrand stacking between the guanines,90–93 whereas $S'G/U/3'U$ does not.90 The favorability of the $S'U/G\text{U}3'U$ motif relative to the $S'G/U/3'U$ motif is consistent with molecular dynamics (MD) simulations91 that predict a one hydrogen bond $G\text{U}3'G$ pair92,93 predominates in duplexes containing the $S'G/U/3'U$ motif while a two hydrogen bond model predominates in duplexes containing the $S'U/G\text{U}3'U$ motif. There is also less overlap of negative potentials in $S'U/G\text{U}3'U$ than in $S'G/U/3'U$.91 In two different sequences containing the $S'U/G\text{U}3'U$ motif, there is also intrastrand stacking between each GU pair and its Watson–Crick neighbors.94 By comparison, the $S'G/U/3'U$ motif contains less overlap between the GU pairs and Watson–Crick purine neighbors, but has intrastrand stacking between the tandem GU pairs.94 Furthermore, the $S'U/G\text{U}3'U$ motif preserves the A-form of RNA more than $S'G/U/3'U$.96

The $S'G\text{GU}/3'C\text{UG}$ motif is an exception to the above generalizations. NMR spectra and modeling indicate that the GU pairs of r(GAGGUCCUC)2 contain two hydrogen bonds,52 whereas the GU pairs in r(GGCGUUCGC)2 contain only one hydrogen bond.98 This difference would contribute to the favorable free energy of $S'G\text{GU}/3'C\text{UG}$ compared to that of $S'G/U/3'U$ in other contexts, such as $S'G\text{GU}/3'C\text{UG}$. Pan et al. saw similar hydrogen-bonding scenarios in MD simulations.91 Additional stability for the $S'G\text{GU}/3'C\text{UG}$ motif may also be conferred from less overlap of its negative electrostatic potentials between a GC and GU pair than for its related motif, $S'G\text{GU}/3'C\text{UG}$.52 These patterns may explain the poor fit of nearest neighbor parameters for the $S'G/U/3'U$ motif when duplexes containing the $S'G\text{GU}/3'C\text{UG}$ motif are included in the fit. Alternatively, the extra stability of the $S'G\text{GU}/3'C\text{UG}$ motif over $S'G\text{CG}/3'\text{GU}$ may arise from poor cross-strand overlap between the U in a GU pair and the C in its neighboring GC pair in $S'G\text{CG}/3'\text{GU}$.97 Stacking interactions alone do not contribute to the stability of nearest neighbor motifs comprised of the same base pairs, however, as evident from the comparable stability of $S'U/G/3'U$ and $S'G/G/3'U$. This contrasts with the expectation that the free energy of $S'G/G/3'U$ is between the other tandem GU motifs because its base stacking is intermediate among them.98 Understanding the interactions responsible for the observed sequence dependence of thermodynamics presents a challenge to computational chemists.

GU Pairs of RNA Are More Stable than GT Pairs of DNA. Comparison of $\Delta G^{\text{rr}}$ values for GT nearest neighbors in DNA74 with those measured for GU nearest neighbors show that GT nearest neighbors are on average 0.84 ± 0.36 kcal/mol less stable than their GU counterparts. The extra stability of GU relative to GT is also evident from comparisons of $\Delta G^{\text{rr}}$ (GU or GT component) for duplexes containing comparable triplet motifs (Table 7). This may reflect a possible hydrogen bond between the amino group of guanine and the O2' of uracil,99 which is not possible with DNA. MD simulations utilizing residual dipolar coupling (RDC) restraints suggest that the $S'T\text{G}/3'G$ motif contains a bifurcated hydrogen bond90 similar to that in the $S'G/U/3'U$ motif.98,99 Another difference between GT and GU is that the $S'G\text{TC}/3'C\text{CG}$ motif fits the nearest neighbor parameters for the $S'G/T/3'T$ motif better than their respective uracil-substituted RNA motifs.74 Consistent with the relative stabilities of GT and GU nearest neighbors, component free energies of GT pairs in duplexes are

### Table 6. Free Energy Differences When GU Pairs Are Replaced with AU or GC Pairs

| GC duplex | ref | $\Delta G_{\text{AU}}$ (GC duplex) | GU duplex | ref | $\Delta G_{\text{GU}}$ (GU duplex) | $\Delta G_{\text{per GU pair}}$ (kcal/mol) |
|-----------|-----|---------------------------------|-----------|-----|-----------------------------------|---------------------------------------------|
| CCGCGCG | 11  | 9.84                            | CUGCGCG | 101 | 4.31                             | 2.77                                        |
| CCGCGCGGp | 110 | 9.90                             | CCGCGCG | 101 | 5.55                             | 2.18                                        |
| CGCGCGGOp | 110 | 9.90                             | CGCGCGGp | 110 | 8.56                             | 0.67                                        |
| CUGCGAG | 111 | 7.11                             | GUGCAG | 43  | 4.20                             | 1.46                                        |
| CCGCGGp | 110 | 11.20                            | GCGCGp | 110 | 9.17                             | 1.02                                        |
| GCGCGp | 112 | 11.33                            | GCGCGp | 112 | 8.42                             | 1.46                                        |
| GCGCGGp | 112 | 11.33                            | GCGCGGp | 112 | 4.67                             | 3.33                                        |
| GUCGAC | 111 | 7.65                             | GUCGAC | 111 | 5.10                             | 1.28                                        |
| AUGCAGAp | 101 | 10.17                           | AUGCAGAp | 101 | 5.27                             | 2.45                                        |
| AUGCAGAp | 113 | 9.67                             | AUGCAGAp | 113 | 6.44                             | 1.62                                        |
| AUGCAGAp | 113 | 10.12                            | AUGCAGAp | 113 | 6.82                             | 1.65                                        |
| GCAGCGC | 114 | 13.87                            | GCAGCGC | 114 | 10.30                            | 1.79                                        |

Average: 1.80 ± 0.76

| AU duplex | ref | $\Delta G_{\text{AU}}$ (AU duplex) | GU duplex | ref | $\Delta G_{\text{GU}}$ (GU duplex) | $\Delta G_{\text{per GU pair}}$ (kcal/mol) |
|-----------|-----|---------------------------------|-----------|-----|-----------------------------------|---------------------------------------------|
| ACCGCG | 115 | 8.51                            | CCGCGCG | 102 | 9.17                             | -0.33                                       |
| AGCGG | 112 | 7.99                            | GCGCGG | 102 | 8.42                             | -0.22                                       |
| CAGCGGp | 111 | 6.68                            | CCGCGG | 111 | 5.55                             | 0.57                                        |
| CGCGGp | 111 | 7.11                            | CCGCGG | 111 | 4.31                             | 1.40                                        |
| GAGCGC | 116 | 7.35                            | GCGCGG | 116 | 4.67                             | 1.34                                        |
| CCGCGAp | 88  | 7.99                            | GCGCGGp | 102 | 7.44                             | 0.28                                        |
| CUCACUC/ | 11  | 9.71                            | CUCCGUC/ | 117 | 7.78                             | 1.93                                        |
| AAUGCAGAp | 113 | 7.18                            | AAUGCAGAp | 106 | 6.82                             | 0.18                                        |
| AUAGCAGAu | 101 | 6.53                             | AUAGCAGAu | 101 | 5.27                             | 0.63                                        |
| AGCGCGAp | 101 | 10.17                           | AGCGCGAp | 101 | 9.31                             | 0.43                                        |
| UAGCGCAUAp | 113 | 7.27                           | UAGCGCAUAp | 106 | 6.44                             | 0.42                                        |

Average: 0.60 ± 0.70
consistently less favorable than those of GU pairs flanked by the same Watson–Crick pairs (Table 7).

Table 7. Component Free Energies of GU and GT Pairs

| Sequences       | $\Delta G^{37}$ (kcal/mol) | Experimental (kcal/mol) | Predicted (kcal/mol) |
|-----------------|----------------------------|-------------------------|----------------------|
| GCGU         | −5.11                      | −2.79                   | −1.78                |
| GACCCTGCAC/   | −7.17                      | −0.40                   | 0.20                 |
| AUUGCU Gu †  | −5.27                      | −3.07                   | −2.54                |
| CACCCTGTAACG/ | −8.94                      | −0.90                   | −0.30                |
| CTTTGATCTAG/e | −6.10                      | −0.55                   | −0.15                |
| CUGCCUC/      | −8.22                      | −3.45                   | −3.65                |
| GACCTGGGAC/   | −7.91                      | −1.40                   | −0.30                |
| GUGCUAG       | −7.10                      | −4.04                   | −4.32                |
| CTTTGATCTAG/e | −5.89                      | −1.20                   | 0.20                 |
| GAGUGUC       | −9.40                      | −5.06                   | −4.87                |
| GAGATCCTCC     | −7.66                      | −2.20                   | −0.70                |
| GACGU         | −5.90                      | 0.64                    | 0.02                 |
| GGGUUC        | −6.43                      | 0.27                    | 0.02                 |
| GCGAATGTCGC/  | −6.87                      | 1.40                    | 2.60                 |
| GAAUGUC       | −8.40                      | −1.86                   | −2.37                |
| GGAUGUC       | −8.39                      | −1.69                   | −2.37                |
| GCGAATGTCGC/e | −7.98                      | 0.10                    | 0.70                 |
| CAGUCGUAUC/e  | −8.70                      | −0.97                   | −1.25                |
| GTACATGTCAC/  | −7.78                      | −1.20                   | 0.80                 |
| CGATGATGTC/e  | −7.71                      | 0.35                    | 0.80                 |
| GAGCGUUC      | −8.82                      | −1.06                   | −1.72                |
| GAGAGCTTCC     | −6.96                      | 1.70                    | 2.10                 |
| GAGAGCUUC      | −9.83                      | −1.93                   | −2.12                |
| CATGAGGTCAC/  | −8.57                      | −0.90                   | 0.40                 |
| GAGUGGAGAG     | −9.87                      | −0.78                   | −1.15                |
| ACGTGAGAG/e   | −4.61                      | 0.30                    | 1.50                 |
| GUAGAUAAUC    | −4.78                      | −1.86                   | −1.80                |
| GUAGAGAGC     | −8.60                      | −1.06                   | −1.80                |
| CGTGAGCAGTCG/ | −8.19                      | 0.70                    | 1.50                 |
| CGTTGACAGTCG/e| −7.86                      | 1.10                    | 1.50                 |
| GUGUGAUAC     | −8.90                      | −1.30                   | −2.58                |
| CGTGCGTAGC/e  | −8.42                      | 0.20                    | 1.00                 |
| CGTTGTCTAGA/TC| −9.40                      | 0.20                    | 1.00                 |

*Data for DNA duplexes were referenced from ref 74. † Listed in order of length of the oligoribonucleotide and then alphabetically for sequences of the same length. All nonself-complementary sequences have a dash and only one sequence shown. GU and GT base pairs are underlined. ‡ Component free energies were divided by 2. § Calculated with free energies in Table 3. ¶ Marginally non-two-state. |
(10) Borre, P. N., Dengler, B., Tinoco, I., and Uhlenbeck, O. C. (1974) Stability of ribonucleic acid double-stranded helices. J. Mol. Biol. 86, 843–853.

(11) Xia, T. B., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X. Q., 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, 14719–14732.

(12) 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.

(13) Turner, D. H. (2000) Conformational changes, in Nucleic Acids: Structures, Properties, and Functions (Bloomfield, V. A., Crothers, D. M., Tinoco, J., L., Eds.) pp 259–334, University Science Books, Herndon, VA.

(14) Washietl, S., Hofacker, I. L., and Stadler, P. F. (2005) Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. U.S.A. 102, 2454–2459.

(15) Uzilov, A., Keegan, J., and Mathews, D. H. (2006) Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change. BMC Bioinformatics 7, 173.

(16) Gruber, A. R., Neuböck, R., Hofacker, I. L., and Washietl, S. (2007) The RNAs web server: prediction of thermodynamically stable and evolutionarily conserved RNA structures. Nucleic Acids Res. 35, W335–W338.

(17) Reiche, K., and Stadler, P. F. (2007) RNAstr: reading direction of structured RNAs in multiple sequence alignments. Algorithm. Mol. Biol. 2, 6.

(18) White, S. A., Nilges, M., Huang, A., Brunger, A. T., and Moore, P. B. (1992) NMR analysis of helix-I from the SS RNA of Escherichia coli. Biochemistry 31, 1610–1621.

(19) Szyniszewski, M., Barciszewska, M. Z., Erdmann, V. A., and Barciszewski, J. (2000) An analysis of G-U base pair occurrence in eukaryotic 5S RNAs. Mol. Biol. Evol. 17, 1194–1198.

(20) Sprinzl, M., and Vassilenko, K. S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. 33, D139–D140.

(21) Limmer, S., Reif, B., Ott, G., Arnold, L., and Sprinzl, M. (1996) NMR evidence for helix geometry modifications by a G-U wobble base pair in the acceptor arm of E-coli tRNA(Ala). FEBS Lett. 385, 15–20.

(22) Hou, Y. M., and Schimmel, P. (1988) A simple structural feature recognizes an RNA G:U juxtaposition. Science 240, 793–796.

(23) McClain, W. H., and Foss, K. (1988) Changing the identity of a terminal GU pairs stabilize RNA helices. Biochemistry 27, 1824–1828.

(24) Bevilacqua, P. C., and Golden, B. L. (2010) A 1.9 Å crystal structure of the 5S RNA of hexammine. J. Mol. Biol. 390, 119–135.

(25) Pyle, A. M., Moran, S., Strobel, S. A., Chapman, T., Turner, D. H., and Cech, T. R. (1994) Replacement of the conserved G-C pair at the cleavage site of the tetrahymena ribozyme decreases binding, reactivity, and fidelity. Biochemistry 33, 13856–13863.

(26) Strobel, S. A., and Cech, T. R. (1995) Minor groove recognition of the conserved G-U pair at the tetrahymena ribozyme reaction site. Science 267, 675–679.

(27) Strobel, S. A., and Cech, T. R. (1996) Exocyclic amine of the conserved G-U pair at the cleavage site of the Tetrahymena ribozyme contributes to 5′-splice site selection and transition state stabilization. Biochemistry 35, 1201–1211.

(28) Konarska, M. M. (1999) The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5′ splice site G. Mol. Biol. Evol. 17, 1816.

(29) Leung, S. S., and Koslowsky, D. J. (2001) Interactions of mRNAs and gRNAs involved in trypanosome mitochondrial RNA editing: structure probing of an mRNA bound to its cognate gRNA. Nucleotides Nucleic Acids 28, 424–434.

(30) Knitt, D. S., Narlikar, G. J., and Herschlag, D. (1994) Dissection of the role of the conserved G-U pair in group I RNA self-splicing. Biochemistry 33, 13864–13879.

(31) Pyle, A. M., Moran, S., Strobel, S. A., Chapman, T., Turner, D. H., and Cech, T. R. (1994) Replacement of the conserved G-U with a G-C pair at the cleavage site of the tetrahymena ribozyme decreases binding, reactivity, and fidelity. Biochemistry 33, 13856–13863.

(32) F. (2011) NMR detection of structures in the HIV-1 5′-leader RNA that regulate genome packaging. Science 334, 242–245.

(33) Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) Reliable prediction of noncoding RNAs.
(96) Masquida, B., and Westhof, E. (2000) On the wobble G·U and related pairs. RNA 6, 9–15.
(97) Jang, S. B., Hung, L. W., Jeong, M. S., Holbrook, E. L., Chen, X. Y., Turner, D. H., and Holbrook, S. R. (2006) The crystal structure at 1.5 Å resolution of an RNA octamer duplex containing tandem G·U basepairs. Biophys. J. 90, 4530–4537.
(98) Deng, J. P., and Sundaralingam, M. (2000) Synthesis and crystal structure of an octamer RNA r(guguuuac)/r(guaggcac) with G·G/U·U tandem wobble base pairs: comparison with other tandem G·U pairs. Nucleic Acids Res. 28, 4376–4381.
(99) Shi, K., Wahl, M. C., and Sundaralingam, M. (1999) Crystal structure of an RNA duplex r(GGGCGCUCC)2 with non-adjacent G·U base pairs. Nucleic Acids Res. 27, 2196–2201.
(100) Álvarez-Salgado, F., Desvaux, H., and Boulard, Y. (2006) NMR assessment of the global shape of a non-labelled DNA dodecamer containing a tandem of G·T mismatches. Magn. Reson. Chem. 44, 1081–1089.
(101) Sugimoto, N., Kierzek, R., Freier, S. M., and Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helices. Biochemistry 25, 5755–5759.
(102) Freier, S. M., Kierzek, R., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Free energy contributions of G·U and other terminal mismatches to helix stability. Biochemistry 25, 3209–3213.
(103) Testa, S. M., Disney, M. D., Turner, D. H., and Kierzek, R. (1999) Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: Implications for antisense design and targeting a group I intron. Biochemistry 38, 14214–14223.
(104) He, L., Kierzek, R., SantaLucia, J., Walter, A. E., and Turner, D. H. (1991) Nearest-neighbor parameters for G·U mismatches - SGU3/3UGS5 is destabilizing in the contexts CGUG/GUGC, UGU/AUGU, and AGGU/UUGU but stabilizing in GGUC/CUGG. Biochemistry 30, 11124–11132.
(105) 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.
(106) Sugimoto, N., Kierzek, R., and Turner, D. H. (1987) Sequence dependence for the energetics of terminal mismatches in ribonucleic acid. Biochemistry 26, 4559–4562.
(107) Ziomek, K., Kierzek, E., Biala, E., and Kierzek, R. (2002) The thermal stability of RNA duplexes containing modified base pairs placed at internal and terminal positions of the oligoribonucleotides. Biophys. Chem. 97, 233–241.
(108) Schroeder, S. J., and Turner, D. H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA. Biochemistry 40, 11509–11517.
(109) Schroeder, S. J., and Turner, D. H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. Biochemistry 39, 9257–9274.
(110) Freier, S. M., Sinclair, A., Neilson, T., and Turner, D. H. (1985) Improved free energies for G·C base-pairs. J. Mol. Biol. 185, 645–647.
(111) Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986) Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. U.S.A. 83, 9373–9377.
(112) Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., and Turner, D. H. (1986) Stability of XGCGCp, GCGCy, and XGCGCy helices: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. Biochemistry 25, 3214–3219.
(113) Sugimoto, N., Kierzek, R., and Turner, D. H. (1987) Sequence dependence for the energetics of dangling ends and terminal base pairs in ribooligonucleotides. Biochemistry 26, 4554–4558.
(114) Burkard, M. E., and Turner, D. H. (2000) NMR structures of r(GCAGGCGUGC)3 and determinants of stability for single guanosine-guanosine base pairs. Biochemistry 39, 11748–11762.
(115) 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.
(116) Kierzek, R., Caruthers, M. H., Longfellow, C. E., Swinton, D., Turner, D. H., and Freier, S. M. (1986) Polymer-supported RNA synthesis and its application to test the nearest-neighbor model for duplex stability. Biochemistry 25, 7840–7846.
(117) Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) Thermodynamics of single mismatches in RNA duplexes. Biochemistry 38, 14214–14223.