RNA Helix Thermodynamics: The End Game

Jeffrey Zuber[1,†], Susan J. Schroeder[2,†], Hongying Sun[3,4], Douglas H. Turner[5], David H. Mathews[3,4,6,*]
[1]Alnylam Pharmaceuticals, Inc., Cambridge, MA 02142, USA, [2]Department of Chemistry and Biochemistry, and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA, [3]Department of Biochemistry & Biophysics, University of Rochester, Rochester, NY 14642, USA, [4]Center for RNA Biology, University of Rochester, Rochester, NY 14642, USA, [5]Department of Chemistry, University of Rochester, Rochester, NY 14627, USA, and [6]Department of Biostatistics & Computational Biology, University of Rochester, Rochester, NY 14642, USA.
*To whom correspondence should be addressed. Tel: +1 585 275 1734; Email: David_Mathews@urmc.rochester.edu
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

ABSTRACT

Nearest neighbor parameters for estimating the folding stability of RNA secondary structures are in widespread use. For helices, current parameters penalize terminal AU base pairs relative to terminal GC base pairs. We curated an expanded database of helix stabilities determined by optical melting experiments. Analysis of the updated database shows that terminal penalties depend on the sequence identity of the adjacent penultimate base pair. New nearest neighbor parameters that include this additional sequence dependence accurately predict the measured values of 271 helices in an updated database with a correlation coefficient of 0.982. This refined understanding of helix ends facilitates fitting terms for base pair stacks with GU pairs. Prior parameter sets treated $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$ separately from other $\begin{smallmatrix}5'\mathrm{GU}\\3'\mathrm{UG}\end{smallmatrix}$ stacks. The improved understanding of helix end stability, however, makes the separate treatment unnecessary. Introduction of the additional terms was tested with three optical melting experiments. The average absolute difference between measured and predicted free energy

changes at 37° C for these three duplexes containing terminal adjacent AU and GU pairs improved from 1.38 to 0.27 kcal/mol. This confirms the need for the additional sequence dependence in the model.

INTRODUCTION

Over 80% of the human genome is transcribed into RNA, but less than 3% of the RNA codes for proteins (1,2). Functions for most RNA in the biosphere are still being discovered but already include catalysis (3), control of transcription, translation and expression (4-6), templating for synthesis of DNA (7) and RNA (8), recognition of sites for modification and editing (9-11) and sometimes combining such functions (12). RNA is the genomic material for many viruses, including human pathogens such as SARS and SARS-CoV-2, influenza, HIV, Ebola, and Hepatitis C. RNA can also be the basis for vaccines against some of these viruses. For example, mRNA vaccines show 94-95% effectiveness against SARS-CoV-2 infections (13).

RNA sequence determines the base pairing and 3D structure as well as function of the RNA. Prediction of secondary structure, i.e. the canonical set of Watson-Crick-Franklin (WCF) and GU base pairs, from sequence is a first step in predicting 3D structure (14) and in finding RNAs with common structures and functions (15,16). Some RNA, such as riboswitches, have more than one structure, and the ability to change structure is critical to function (5).

Secondary structure can be predicted from one or more sequences by minimizing free energy change for folding, ΔG°, often augmented with information from chemical mapping and/or sequence comparison. Usually, about half the nucleotides in an RNA are canonically paired (17). GU pairs play important roles in RNA structure and function as sites for binding metal ions (18,19), therapeutics (20), proteins, or metabolites (21).

A database of thermodynamic measurements for helices with canonical pairs and model non-canonical motifs forms the foundation for folding free energy predictions of RNA structure. These data are then fit to a nearest neighbor (NN) model to estimate parameters that can be used to predict folding stabilities of any RNA secondary structure (22). Hallmarks of the NN model are that each stability increment depends on local sequence and that total stability is the sum of the increments.

The model and parameters for approximating stabilities of WCF base-paired helixes have not changed substantially since 1998 (23,24). Individual parameters for nearest neighbors containing at least one GU pair, however, were revised on the basis of new data (24). In that revision, the penalty of 0.45 kcal/mol applied to terminal AU pairs and previously assumed for terminal GU pairs (23), was found unnecessary for GU pairs. Expansion of the database for duplexes, particularly those with terminal GU pairs (25) and the data presented here, make possible more extensive considerations of terminal effects on base pair stability. In particular, the data allow expansion of the model to include six new parameters specific for terminal nearest neighbors. Surprisingly, stabilities of terminal GU and AU pairs depend on whether the neighboring pair is a GC, AU, or GU pair. Incorporating this effect in the NN model also produces significant

revision of parameters for internal $\begin{smallmatrix}5'GU\\3'UG\end{smallmatrix}$ and $\begin{smallmatrix}5'AG\\3'UU\end{smallmatrix}$ nearest neighbors. With these changes, $\begin{smallmatrix}5'GGUC\\3'CUGG\end{smallmatrix}$ fits the NN model rather than being an outlier as considered previously (26). Thus, the new model presented here will be especially important for predicting structures containing GU pairs.

It is not surprising that GU pairs are more idiosyncratic than WCF pairs. Guanine has more hydrogen bonding groups and a larger dipole moment than other bases (27). Base stacking and hydrogen bonding that stabilize GU pairs can vary depending on local context, including position in a helix. Base stacking depends on interactions with both bases of a nearest neighbor. GU pairs can adopt different hydrogen bonded configurations and stacking interactions (Figure 1). In Figure 1A, the terminal UG pair is in a $\begin{smallmatrix}5'UG\\3'GU\end{smallmatrix}$ nearest neighbor and has a single hydrogen bond while the penultimate GU pair has two hydrogen bonds. Conformation of the terminal UG pair may be influenced by solvent interactions or crystal contacts through stacking interactions with the terminal UG pair of an adjacent molecule. In Figure 1B, the $\begin{smallmatrix}5'UG\\3'GU\end{smallmatrix}$ nearest neighbor is flanked by WCF GC pairs on both sides in the middle of the helix, i.e. $\begin{smallmatrix}5'...GUGC...\\3'...CGUG...\end{smallmatrix}$. In that context, both the UG and GU pair have a hydrogen bond from the G carbonyl to the U imino proton, a bifurcated hydrogen bond between the U carbonyl at C2 and the G imino and amino protons, and extensive cross-strand guanine stacking. In contrast, when the sequence is reversed, i.e. 5'...CGUG... compared to 5'...GUGC..., in the self-complementary duplex (Figure 1C), each GU pair has a single bifurcated hydrogen bond, and there is no cross-strand stacking.

Reported thermodynamic stabilities for the internal NN GU stacks in Figures 1B and C reflect the different nucleotide configurations and effect of considering

terminal effects. The ΔG°_{37} of $\begin{smallmatrix}5'UG\\3'GU\end{smallmatrix}$ and $\begin{smallmatrix}5'GU\\3'UG\end{smallmatrix}$ motifs are -0.38 and -0.19 kcal/mol, respectively, in the new model as compared to -0.57 and +0.72 kcal/mol in a prior model (24). The new parameters also include an increment of -0.74 kcal/mol for terminal consecutive GU pairs, i.e. a GU end on a GU pair, like those shown in Figure 1A. In contrast, previous models did not add favorable folding free energy for this sequence motif. In many NMR structures with terminal GU and AU pairs, these pairs show more dynamic behavior relative to internal pairs. This is observed as broad imino proton resonances and fewer NOE restraints (25,28-32). Thus, sequence orientation, stacking interactions, hydrogen bonding, and nucleotide dynamics are important factors in the structure and stabilities of GU pairs. They are now more accurately accounted for in the new thermodynamic parameter set. This should improve predictions of secondary structure from sequence.

MATERIALS AND METHODS

Optical Melting Experiment Database

For this analysis, optical melting experiments were compiled through an extensive literature review (24,25,28,33-61). Enumeration of all melting experiments included in this analysis is available in Supplemental Table 1 and in the spreadsheet provided in the supplementary materials. Experiments are included if only unmodified nucleotides are present and buffer has 1 M Na+ with pH between 6.5 and 7.5. Additionally, values of ΔH° determined with van't Hoff plots and curve fitting of absorption vs temperature had to agree within 15% (42), which is consistent with an approximate 2-state transition.

Most melting experiments for generating NN parameters have used 1 M Na^{+}. This was initially chosen because available duplexes had many AU pairs and

therefore needed high salt to melt in convenient temperature ranges (62,63) . The most important result from thermodynamic studies is the relative sequence dependence of nearest neighbor stability. This is expected not to depend on salt conditions because there is no site binding of Na^+ to fully base paired RNA. Local concentrations of mobile cations around large folded RNAs, however, depend on the local charge density of phosphate groups. Manning developed a first order cation condensation model that predicts local concentrations of cations around RNA do not depend on bulk concentrations (64). For A-form double helical and single strand RNA, respectively, the local "ion atmosphere" is predicted to have 1.7 M and 0.4 M of M^+ ions (65). They respectively neutralize 0.8 and 0.6 of the phosphate charge. Mg^{2+} will also bind as part of the ion atmosphere and neutralize 0.9 and 0.8, respectively of backbone charge. More detailed computations and experiments agree qualitatively with expectations from Manning theory (28,51,57,66-71).

Feature Correlations

Feature correlations were calculated for each model using the R statistical programming language (72). The resulting correlation matrices were then visualized with the *R corrplot* library (73).

Fitting Linear Models

Parameter models were fit using measured ΔG°_{37} and ΔH° values for each optical melting experiment. For the fit of WCF stacking parameters, the theoretical contribution of RT ln(2) due to 2-fold symmetry of self-complementary duplexes, was subtracted from the experimentally measured duplex ΔG°_{37} (23,74). For the fit of GU stacking parameters, the contributions due to sequence symmetry and the WCF stacks from each duplex with any GU base pairs were subtracted from measurements. The calculated ΔG°_{37} and ΔH° are then used to fit

linear models in the R statistical programming language using the base function *lm (72)*. ΔS° values for the nearest neighbor parameters are calculated from the ΔG°_{37} and ΔH° values.

To estimate uncertainty in NN parameter values, a covariation analysis was used to account for the dependencies (due to the nested nature of the regressions) and correlation (due to a base pair appearing in up to two neighboring stacks) between parameters (75) (76). To perform covariation analysis, the optical melting data were resampled within experimental error (ΔH_{σ}= 12% ΔH° and $\Delta S^{\circ}_{\sigma}$= 13.5% ΔS° (23)). The resampling was performed with the *mvrnorm* function from the R MASS library (77), which preserves the observed correlation between ΔH° and ΔS° (ρ= 0.9996 (23)). The updated experimental values are then used to recalculate multiple sets of model parameters. The sets of model parameters are then used to calculate average values for each parameter as well as covariation (75,76). The standard errors of regression, which neglect the correlations and the effect of nested regressions, for the NN parameters can be found in Supplemental Table 2.

Leave-One-Out Analysis

To assess the impact of any one experimental value on the fit models, models were fit in which each experimental value was individually excluded from the fitting data. The root mean square deviations (RMSDs) in parameter values were calculated from the model fit to the full data set to measure the impact of excluding each individual experimental value.

Optical Melting Experiments

Optical melting experiments were conducted following standard protocols described in (78). Oligonucleotides were purchased from Integrated DNA Technologies including purification with standard desalting procedures and

assessment of purity by mass spectrometry. Oligonucleotides were dissolved in milliQ water, and the absorbance at 260 nm at 80 °C was measured. The appropriate amount of oligonucleotide was dried in a speed vac and resuspended in standard melting buffer of 1 M NaCl, 20 mM sodium cacodylate, pH 7, and 0.5 mM Na_2EDTA. Optical melting experiments were conducted in a Beckman DU800 UV-Vis spectrometer with a custom sample holder and cuvettes at 0.1 cm and 1.0 cm path lengths. Absorbance vs. temperature was measured at 260 and 280 nm. Data was analyzed at 280 nm with Meltwin software (51).

Stacking Term Counts

An archive of RNA sequences of known secondary structure {Sloma, 2016 #1348} was analyzed to count the number of occurrences of each stacking parameter. A Python script was used to parse each structure into individual helices and then to parse each helix into component stacking and helix end parameters.

Results

AU End Parameters Depend on Penultimate Pair

Prior work demonstrated that multiple GU terminal base pairs impact the stability of helical duplexes (25), and this motivated a reexamination of the treatment of helix ends. New terms to account for the end of a helix were introduced into the NN model. This was done by including a parameter for an AU terminal pair on an AU penultimate pair (not accounting for orientation of the two pairs and therefore applying to $\begin{smallmatrix}5'\ldots AU\\3'\ldots UA\end{smallmatrix}$, $\begin{smallmatrix}5'\ldots AA\\3'\ldots UU\end{smallmatrix}$, $\begin{smallmatrix}5'\ldots UU\\3'\ldots AA\end{smallmatrix}$, or $\begin{smallmatrix}5'\ldots UA\\3'\ldots AU\end{smallmatrix}$ helix ends) and a parameter for an AU terminal pair on a CG penultimate pair (applying to $\begin{smallmatrix}5'\ldots GU\\3'\ldots CA\end{smallmatrix}$, $\begin{smallmatrix}5'\ldots GA\\3'\ldots CU\end{smallmatrix}$, $\begin{smallmatrix}5'\ldots CU\\3'\ldots GA\end{smallmatrix}$, or $\begin{smallmatrix}5'\ldots CA\\3'\ldots GU\end{smallmatrix}$ helix ends). These parameters are applied in

addition to the base pair stacking parameter for these end stacks. This model is consistent with the observation that the thermodynamic stability bonus conferred by multiple GU terminal pairs is largely independent of orientation of the GU pairs (25).

Comparisons between the updated model and those used in the 1998 and 2004 NN models are in Table 1. Fitting to the model with the modified helix end parameters resulted in only moderate changes to the WCF stacks. The parameters for intermolecular initiation and the individual NN stacks were all within error of the 1998 parameters for both ΔG°_{37} and ΔH°. The only significant change was that an AU terminal pair on an AU penultimate pair is more favorable compared to the 1998 and 2004 models.

The updated model shows excellent correlations between predicted and measured values of ΔG°_{37} (R^2=0.9823, Figure 2) and ΔH° (R^2=0.8877, Supplemental Figure 1). The correlations between the model feature frequencies are modest and are mostly limited to expected correlations between stacks that can extend on each other (Supplemental Figure 2). The predicted folding ΔG°_{37}were within 0.5 kcal/mol of the measured value for 86.4% of the experiments (Supplemental Figure 3). Predicted ΔH° are within 5 kcal/mol of the measured value for 76% of the experiments (Supplemental Figure 3).

The impact of each optical melting experiment was determined by fitting the NN parameters on a data set that excluded that individual experiment and comparing the resulting parameter values to those fit on the full data set. The root mean squared deviations (RMSDs) in ΔG°_{37} and ΔH° parameter values for these leave-one-out (LOO) data sets can be seen in Supplemental Figure 4. No one individual experiment heavily impacted the parameter values. The biggest

impacts were RMSDs of 0.0363 kcal/mol in ΔG°_{37} and 0.3346 kcal/mol in ΔH°, substantially smaller than uncertainty in the parameter values.

GU Stacking Parameters

A similar model for terminal AU and GU stacks was used when fitting duplexes with GU base pairs. The model requires terms for an AU end with a penultimate GU pair, a terminal GU pair with a penultimate AU pair, a terminal GU pair with a penultimate GC pair, and a GU pair with a penultimate GU pair. The orientation of the two pairs is not considered.

Prior GU stack NN parameter sets treated $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$ as a special, non-nearest neighbor case. When results from a fitting model including a parameter for the non-nearest neighbor quadruplet $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$, however, were compared to results for a model not including that parameter, the other parameter values were all within uncertainty of each other. Additionally, for each duplex containing $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$, predicted ΔG°_{37} and ΔH° values from each model were also close to each other and to predicted ΔG°_{37} and ΔH° values from the 2012 model (24) (Supplemental Table 3). Additionally, the addition of the special $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$ parameter results in almost identical R^2 value for the fit (0.9256 vs 0.9267) (Figure 3 and Supplemental Figure 12). Evidently, a special, non-nearest neighbor parameter is not needed in the updated model when end effects for terminal AU and GU nearest neighbors are accounted for.

For the GU internal stacking NN parameters, the most substantial change is for the $\begin{smallmatrix}5'\mathrm{GU}\\3'\mathrm{UG}\end{smallmatrix}$ stack, where the ΔG°_{37} changed from +0.72 kcal/mol to -0.19 kcal/mol between the previous (24) and new models. An additional increment of

-0.74 is added for a terminal $\frac{5'GU}{3'UG}$ stack. The second most substantial change is for the $\frac{5'...AG}{3'...UU}$ stack, which went from having a ΔG°_{37} contribution of -0.35 kcal/mol (24) to +0.02 kcal/mol, but with an additional end increment of -0.31 kcal/mol.

The updated model shows good correlation between predicted and measured values for folding ΔG°_{37} (R^2=0.9256, Figure 3) and ΔH° (R^2=0.7659, Supplemental Figure 5). Correlations between model feature frequencies are mostly limited to expected correlations between the GU on GU end feature and the three possible stacks that can form that end (Supplemental Figure 6). For folding ΔG°_{37}, 53.1% of experiments had predicted values within 0.5 kcal/mol of the measured value (85.7% were within 1 kcal/mol) (Supplemental Figure 7). For ΔH°, 57.1% of experiments had predictions within 5 kcal/mol of the measured value (79.6% were within 10 kcal/mol) (Supplemental Figure 7).

RMSDs in ΔG°_{37} and ΔH° parameter values for the LOO data sets can be seen in Supplemental Figure 8. As with the WCF stacking parameters, no one individual experiment heavily impacted the parameter values. The biggest impacts are RMSDs of 0.0641 kcal/mol in ΔG°_{37} and 0.8675 kcal/mol in ΔH°, smaller than the uncertainty in the parameter values.

Uncertainty in parameter values for the updated NN model presented in Tables 1A and 1B were determined from a covariation analysis, which randomly perturbed the experimental values within experimental uncertainty and calculated the covariance matrix from observed changes in parameter values. Covariances between parameters are presented in Supplemental Figures 9 and 10. Covariances were generally small, with the strongest interactions between the intermolecular initiation parameter and parameters for individual stacks. There are also weaker interactions between GU end terms and equivalent internal

GU stacks that can form that end term. For example, the GU on GU end parameter value is negatively correlated with the values for internal $\frac{5'GU}{3'UG}$, $\frac{5'UG}{3'GU}$, and $\frac{5'GG}{3'UU}$ stacks.

Additional Melting Experiments Support the Model

Three duplexes, designed to test features of the new model for thermodynamic parameters with terminal AU pairs and penultimate GU pairs, were studied by optical melting. A significant difference between the two models occurs for the end parameter for an AU end on a penultimate GU pair. The new and previous (23) models use values of -0.72 kcal/mol and +0.45 kcal/mol at 37 °C, respectively. All three duplexes in Table 2 contain this motif, which was represented in only two duplex sequences in the database of optical melting experiments. The duplexes, $(5'UGUCGAUA)_2$ and $(5'AUAGCUGU)_2$ differ in orientation of the terminal AU pair stacking on the penultimate GU pair. Duplex $(5'AUUCGAGU)_2$ contains the motif $\frac{5'AG}{3'UU}$, which has values of -0.03 and -0.35 kcal/mol for the new and prior models, respectively. Table 2 compares predictions based on the new and prior model with the experimentally measured thermodynamic values for these three duplexes. The experimental results confirm improvement of the new model.

Discussion

The database of thermodynamic parameters forms the foundation for predictions of RNA structure and function in many widely used software suites (14,80-83,84 ,85-87). These RNA structure prediction programs enable design of mRNA vaccine sequences (13,88,89), analysis of metaproperties of transcriptomic changes in response to stress (90-92), determination of effects of nucleotide modifications on folding stability (93-95), discovery of accessible regions to target

with antisense DNA or siRNA (96-99), and rational design of small molecules targeting RNA (100-103). Curation and improvement of the RNA thermodynamic database facilitates hypothesis-driven RNA research in many fields and has significant impact on the RNA community. The progress reported here expands, compiles, and presents the thermodynamic NN parameters for WCF and GU pairs. Statistical significance of the new parameters is robust. Inclusion of helix-end effects for AU and GU pairs improves predictions of helices with these common motifs and resolves previously poorly understood terms for "special cases" of motifs containing GU pairs.

This work presents the next advance in development of a robust NN model for predicting RNA duplex stabilities. The NN model for ΔG° of RNA helixes composed of canonical pairs uses stacks of adjacent base pairs (104-106). This assumes that the total ΔG° and temperature dependence, ΔH°, for helix formation can be approximated by summing ΔG°s and ΔH°s assigned to nearest neighbors of canonical pairs. The experimental foundation for this approach was laid by Uhlenbeck and Martin in the Doty lab when they used optical melting to measure thermodynamics of duplex formation (62,63). As a postdoctoral fellow, Uhlenbeck and the Tinoco lab used biochemical methods to expand the database of sequences. Because WCF base pairing depends on strong, local hydrogen bonding and stacking interactions, a NN model was tested and found to fit the database. This suggested that the NN model would allow predictions for unmeasured sequences (105). In collaboration with related efforts in the Crothers lab, this led to original rules for predicting the thermodynamics of RNA folding (104).

Subsequent insights and research have continually improved success of the NN method. A rotational symmetry term was added to the model to account for the difference between duplexes formed by self- or non-self-complementary strands (42,74,107). Chemical synthesis on polymer supports allowed expansion of sequences available (41,42,108,109). Particularly important was addition of duplexes not beginning with multiple AU pairs and having melting temperatures near 37 °C, human body temperature. Analysis of the number of parameters allowed by the model (110,111) led to discovery that duplexes with the same nearest neighbors can have different thermodynamics depending on the terminal base pair (23). Applications of the method were expanded to larger RNAs by modifying dynamic programming algorithms to predict folding that optimized ΔG° (80,112,113) rather than base pairing (81). Recently, applications to even larger RNAs have become possible due to linearization of the algorithms (114,115).

NN parameters for canonical pairs have undergone substantial revisions over time, including treatment of end effects (23,24,26,42,105). GU terminal base pairs were initially assumed to be equivalent to AU terminal base pairs and were given the same penalty term (26). Expansion of the database and refitting of the model indicated that GU terminal base pairs do not require an end penalty (24). Measurements on helices with consecutive terminal GU pairs, however, revealed they are surprisingly more stable than predicted (25). Our updated NN model includes new parameters that account for end effects of both AU and GU pairs, including dependence on the penultimate pair.

Context-dependent variation of GU pair conformations (Figure 1) provides a structural rationale for treating terminal GU pairs differently. A fundamental assumption of a NN model is that strong local interactions dominate the energetic contributions determining conformation and stability for a particular nucleotide

sequence. This implies that a stack of two WCF pairs will have the same thermodynamic stability in the middle of a helix as at the end of a helix. This NN approximation is consistent with structures of WCF RNA helices and the regular periodic shape of an RNA double helix. The diversity of GU pair conformations and stabilities, however, introduces variation into WCF paired helices. The unique structures of GU pairs facilitate binding recognition and specificity for metal ions, RNA tertiary interactions, protein interactions and drug binding (21). The challenge is incorporating this functionally important and structurally diverse motif into a NN model.

Prior models attempting to combine GU and WCF pairs into one set of thermodynamic parameters always had a few unexplained exceptions. For example, the motif $\begin{smallmatrix}5'\mathrm{GU}\\3'\mathrm{UG}\end{smallmatrix}$ had one NN parameter value with an exception for the motif $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$, which had an extra bonus. Analysis of NMR structures and crystal structures of this motif, however, did not indicate a reason for this additional stability (51,52,116,117). In addition, the crystal structure of consecutive terminal GU pairs in (5'<u>GG</u>UGGCUG<u>UU</u>3')$_2$ had three slightly different helical conformations in the asymmetric unit but an overall remarkably A-form like structure that did not reveal a physical explanation for exceptional thermodynamic parameters (29). NMR studies of duplexes with consecutive terminal GU pairs usually showed broad resonances and few or weak NOEs in the final two GU pairs (25,29). Fluorescence and NMR studies have quantified different base pair dynamics in the middle and ends of helices for various types of base pairs (118-121). Consistent with this, the new NN model has increments for terminal nearest neighbors to distinguish them from internal nearest neighbors.

Interestingly, these increments are penalties +0.22 and +0.44 kcal/mol at 37 ^{o}C for an AU end pair on a penultimate AU or CG pair, respectively (Table 1). A

similar penalty of +0.45 kcal/mol has previously been attributed to the presence of one fewer hydrogen bond when duplexes with identical nearest neighbors have two terminal AU pairs rather than terminal GC pairs (23). In contrast, incremental bonuses of -0.31 to -0.74 kcal/mol are assigned to terminal nearest neighbors consisting of an AU and a GU pair or two GU pairs. This would be consistent with a ΔS° bonus due to increased base pair dynamics at the ends of helices. For example, equal populations of three conformations at the end of a helix would provide a ΔG° bonus of -RT ln (3) = -0.68 kcal/mol at 37 °C.

The NN approximation is essential for efficient dynamic programming approaches to computing the minimum ΔG° secondary structure for an RNA sequence (80,81). In prior models, the special cases for GU pairs required additional considerations in dynamic programming algorithm computations. A helix end occurs not only at the 5' and 3' ends of an RNA molecule but also at every junction, internal loop, hairpin loop, and mismatch or bulge in an RNA secondary structure. GU helix end pairs occur at in accepted RNA secondary structures at a rate of approximately 13 per 1000 bases in the sequence (Supplemental Figure 11). They occur at a much higher rate in predicted structural ensembles. Thus, consideration of GU pairs and special rules for positional dependence present a frequent step in the computations. The NN parameter model presented here improves predictions for sequences and structures with terminal AU and GU pairs and will also accelerate computation of the minimum free energy structure for any sequence.

For example, several terminal AU and GU motifs occur in the secondary structure for the Ψ packaging sequence in HIV-1 RNA (20) and the motif $\begin{matrix}5'\text{UUUU}\\3'\text{GAGG}\end{matrix}$ binds a novel drug. Each helix in the three-way junction that binds the drug has an AU or GU pair at the end, and the new NN parameters in this work would

estimate that the ΔG°_{37} for these three helices is at least 0.9 kcal/mol more stable than current predictions.

Another recent example is the SL3 helix that forms between the 5' and 3' ends of SARS-CoV-2. This helix has been identified experimentally (122) and computationally (123). One end of SL3 terminates in a $\begin{smallmatrix}5'\mathrm{UG}\\3'\mathrm{GU}\end{smallmatrix}$ nearest neighbor. Results in Table 1B assign a ΔG°_{37} of -0.38 -0.74 = -1.12 kcal/mol to this end, which is more stable than previous predictions.

Free energy predictions from nearest neighbors for RNAs secondary structures provide the base line for analysis of the stabilities of RNA interactions with drugs and proteins, and thus provide a foundational resource for RNA structure and function studies. Our future analyses will evaluate the impact of the new NN parameters on the thermodynamic parameters for mismatches, internal loops, bulges, and helix junctions. These loop motifs form many of the recognition sites for proteins, metal ions, and therapeutics.

To illustrate a NN calculation to estimate helix stability, Figure 4 provides two example calculations. The first is the sequence (5'UGUCGAUA)$_2$, with experimental stability provided in Table 2. The second is 5'UAGGUCAG paired with 5'CUGGUCUA. This calculation illustrates that the $\begin{smallmatrix}5'\mathrm{GGUC}\\3'\mathrm{CUGG}\end{smallmatrix}$ motif, an outlier in prior nearest neighbor models, is now handled with nearest neighbor stacks. An Excel spreadsheet is provided with the Supplementary Materials to calculate user-inputted helical NN stabilities.

In summary, the updated NN model is consistent with previous parameters for WCF pairs, includes new parameters accounting for increased base pair dynamics at ends for helices ending in AU or GU pairs, improves predictions for duplexes with terminal AU or GU pairs, and resolves a prior exceptional

parameter for a specific GU motif. The model for the NN parameters has low uncertainty in ΔG° and ΔH° and low correlations between parameters. The statistically robust model maintains the physical basis that differences in hydrogen bonding, stacking, and nucleotide dynamics determine the sequence dependence of NN base stacks. The new thermodynamic parameters will help improve RNA structure prediction tools and facilitate discoveries in RNA biology, catalysis, and therapeutics.

Tables and Figures

Table 1A. The ΔG°_{37} and for ΔH° nearest neighbor parameters for helices composed of WCF pairs.

	New Model[§]			*1998 Model* [¶]	
Feature	ΔG°_{37} (kcal/mol)	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)	ΔH° (kcal/mol)
GC/CG	-3.46 ± 0.08	-16.52 ± 1.57	-42.13 ± 4.26	-3.42 ± 0.08	-14.88 ± 1.58
CC/GG	-3.28 ± 0.08	-13.94 ± 1.18	-34.41 ± 3.58	-3.26 ± 0.07	-13.39 ± 1.24
GA/CU	-2.42 ± 0.05	-13.75 ± 1.00	-36.53 ± 3.16	-2.35 ± 0.06	-12.44 ± 1.20
CG/GC	-2.33 ± 0.09	-9.61 ± 1.57	-23.46 ± 4.74	-2.36 ± 0.09	-10.64 ± 1.65
AC/UG	-2.25 ± 0.06	-11.98 ± 1.17	-31.37 ± 3.86	-2.24 ± 0.06	-11.40 ± 1.23
CA/GU	-2.07 ± 0.07	-10.47 ± 1.25	-27.08 ± 3.73	-2.11 ± 0.07	-10.44 ± 1.28
AG/UC	-2.01 ± 0.07	-9.34 ± 1.23	-23.66 ± 3.63	-2.08 ± 0.06	-10.48 ± 1.24
UA/AU	-1.29 ± 0.08	-9.16 ± 1.71	-25.40 ± 5.55	-1.33 ± 0.09	-7.69 ± 2.02
AU/UA	-1.09 ± 0.07	-8.91 ± 1.55	-25.22 ± 4.75	-1.10 ± 0.08	-9.38 ± 1.68
AA/UU	-0.94 ± 0.04	-7.44 ± 0.80	-20.98 ± 2.56	-0.93 ± 0.03	-6.82 ± 0.79
Initiation	+4.10 ± 0.24	+4.66 ± 3.85	1.78 ± 11.93	+4.09 ± 0.22	+3.61 ± 4.12
Symmetry	+0.43	0	-1.38	+0.43	0
AU End on AU	+0.22 ± 0.06	+4.36 ± 1.23	13.35 ± 3.83	+0.45 ± 0.04 [‡]	+3.72 ± 0.83 [‡]
AU End on CG	+0.44 ± 0.04	+3.17 ± 0.80	8.79 ± 2.50	+0.45 ± 0.04 [‡]	+3.72 ± 0.83 [‡]

Table 1B. The ΔG°_{37} and for ΔH° nearest neighbor parameters for stacks with GU pairs.

	New Model[§]			*2012 Model* *	
Feature	ΔG°_{37} (37 °C) (kcal/mol)	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (37 °C) (kcal/mol)	ΔH° (kcal/mol)
GC/UG	-2.23 ± 0.07	-14.73 ± 1.44	-40.32 ± 4.60	-2.15 ± 0.10	-11.09 ± 1.78
CU/GG	-1.93 ± 0.08	-9.26 ± 1.58	-23.64 ± 5.16	-1.77 ± 0.09	-9.44 ± 1.76
GG/CU	-1.80 ± 0.07	-12.41 ± 1.52	-34.23 ± 4.72	-1.80 ± 0.09	-7.03 ± 1.75
CG/GU	-1.05 ± 0.07	-5.64 ± 1.47	-14.83 ± 4.57	-1.25 ± 0.09	-5.56 ± 1.68
AU/UG	-0.76 ± 0.07	-9.23 ± 1.61	-27.32 ± 5.09	-0.90 ± 0.08	-7.39 ± 1.65
GA/UU	-0.60 ± 0.06	-10.58 ± 1.52	-32.19 ± 4.81	-0.51 ± 0.08	-10.38 ± 1.79
UG/GU	-0.38 ± 0.07	-8.76 ± 1.74	-27.04 ± 5.21	-0.57 ± 0.19	-12.64 ± 4.01
UA/GU	-0.22 ± 0.07	-2.72 ± 1.54	-8.08 ± 4.79	-0.39 ± 0.09	-0.96 ± 1.80
GG/UU	-0.20 ± 0.08	-9.06 ± 1.89	-28.57 ± 6.04	-0.25 ± 0.16	-17.82 ± 3.75
GU/UG	-0.19 ± 0.08	-7.66 ± 1.80	-24.11 ± 5.81	+0.72 ± 0.19	-13.83 ± 4.21
AG/UU	+0.02 ± 0.06	-5.10 ± 1.45	-16.53 ± 4.56	-0.35 ± 0.08	-3.96 ± 1.73
GGUC/CUGG	(-3.80 ±0.13)[†]	(-32.49 ±2.75)[†]	(-92.57 ±11.09)[†]	-4.12 ± 0.54	-30.80 ± 8.87
AU End on GU	-0.71 ± 0.15	5.16 ± 2.99	18.96 ± 9.15	+0.45 ± 0.04[¶]	+3.72 ± 0.83[¶]
GU End on CG	0.13 ± 0.08	3.91 ± 1.43	12.17 ± 4.34	0.00 ± 0.00 [◊]	0.00 ± 0.00 [◊]
GU End on AU	-0.31 ± 0.06	3.65 ± 1.37	12.78 ± 4.23	0.00 ± 0.00 [◊]	0.00 ± 0.00 [◊]
GU End on GU	-0.74 ± 0.08	6.23 ± 2.12	22.47 ± 6.65	0.00 ± 0.00 [◊]	0.00 ± 0.00 [◊]

[§]Uncertainty values were calculated from the experiment covariation analysis. Uncertainty values from standard errors of regression are listed in Supplemental Table 2. [¶]Parameters taken from (23). [‡]The 1998 model did not have separate values for each AU End variant. *Values taken from (24). [†]The new model does not have this parameter. The

shown value is the result of combining the NN stacks for that sequence. [◊]The 2012 model did not include a parameter for terminal GU base pairs.

Table 2: Melting buffer was 1 M NaCl, 20 mM sodium cacodylate pH 7, and 0.5 mM Na_2EDTA. *a)* Predicted ΔG°_{37} are calculated using Table 1 values for the new model parameters. *b)* Predicted ΔG°_{37} are calculated from the previous models described in (24) and (23). The duplex values in italics show borderline apparent 2-state behavior with a difference in ΔH° between the two analyses of 18.2%.

Duplex	New Model[a]	Prior Model[b]	Van't Hoff Plot Analysis				Curve Fit Analysis			
	$-\Delta G^{\circ}_{37\,^{\circ}C}$ (kcal/mol)	$-\Delta G^{\circ}_{37\,^{\circ}C}$ (kcal/mol)	$-\Delta G^{\circ}_{37\,^{\circ}C}$ (kcal/mol)	$-\Delta H^{\circ}$ (kcal/mol)	$-\Delta S^{\circ}$ (eu)	Tm (°C)[†]	$-\Delta G^{\circ}_{37\,^{\circ}C}$ (kcal/mol)	$-\Delta H^{\circ}$ (kcal/mol)	$-\Delta S^{\circ}$ (eu)	Tm (°C)[†]
$(5'UGUCGAUA)_2$	6.02 ± 0.31	4.22 ± 0.37	6.10 ± 0.01	59.32 ± 0.01	171.58 ± 0.01	34.8	6.11 ± 0.18	66.83 ± 3.42	195.78 ± 11.21	39.0
$(5'AUAGCUGU)_2$	6.33 ± 0.29	4.74 ± 0.37	5.81 ± 0.02	44.15 ± 0.01	123.63 ± 0.02	32.0	5.86 ± 0.26	47.88 ± 6.08	135.48 ± 19.18	38.2
$(5'AUUCGAGU)_2$	5.54 ± 0.26	4.14 ± 0.36	*5.32* ± 0.03	*34.58* ± 0.01	*94.33* ± 0.02	26.5	*5.10 ± 0.29*	*42.30 ± 2.98*	*119.94 ± 9.69*	*32.9*

[†] T_m values were calculated for a concentration of 1×10^{-4} M

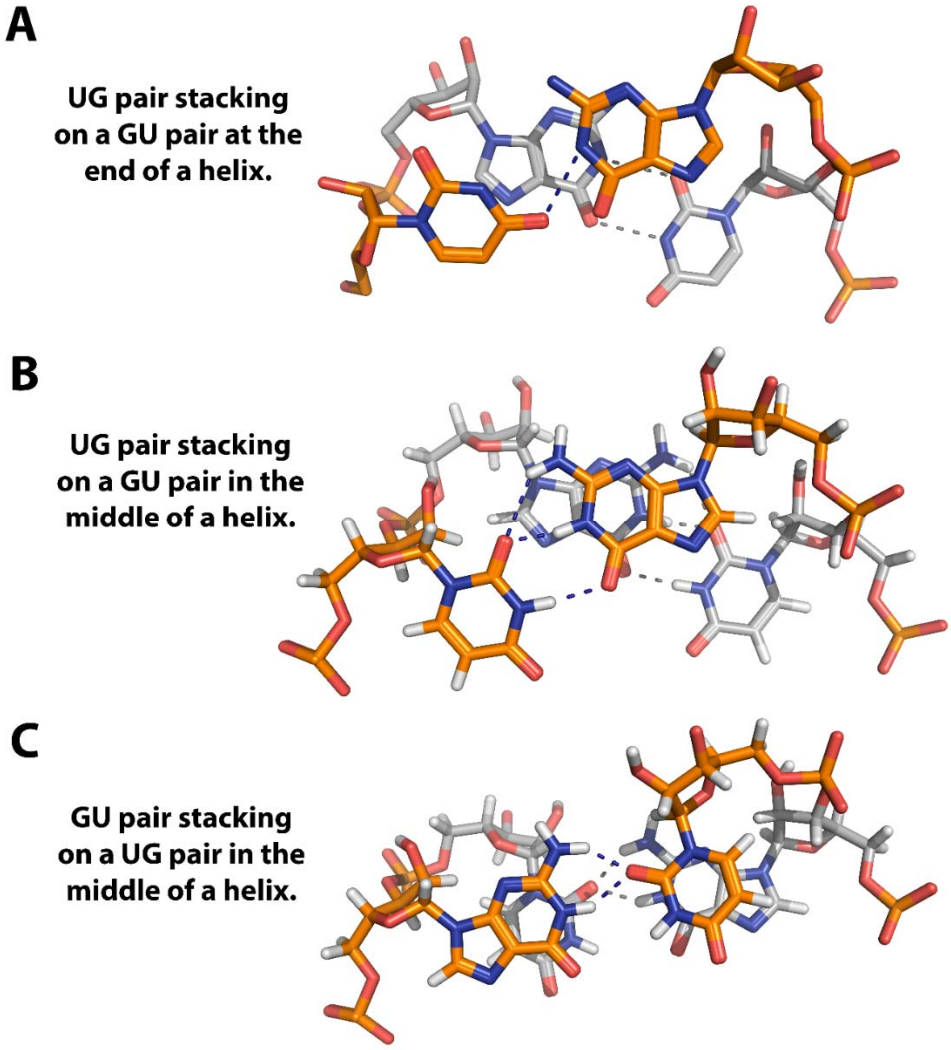


Figure 1: GU pair stacking and hydrogen bonding for three contexts of tandem GU pairs. A. X-ray crystal structure (1.4 Å resolution, R_{free} = 20.7%) with terminal stacking GU pairs in 5'**UG**CUCCUAGUACGUAAGGACCGGAG**UG**, PDB ID# 1MSY (124). Nucleotides in bold in the sequence are shown. The nucleotides with gold carbon atoms are in the forefront and these constitute the terminal base pair, while nucleotides with gray carbon atoms are in the back. The crystal packing has the terminal GU pairs of two molecules stacking on each other. B. NMR structure with internal UG pairs in (5'GAG**UG**CUC)$_2$, PDB ID# 1EKA (116). 28 unique NOE measurements define the orientation for these nucleotides. C. NMR structure with internal GU pairs in (5'GGC**GU**GCC)$_2$. PDB ID# 1EKD (116). 26 unique NOE measurements define the orientation for these nucleotides.

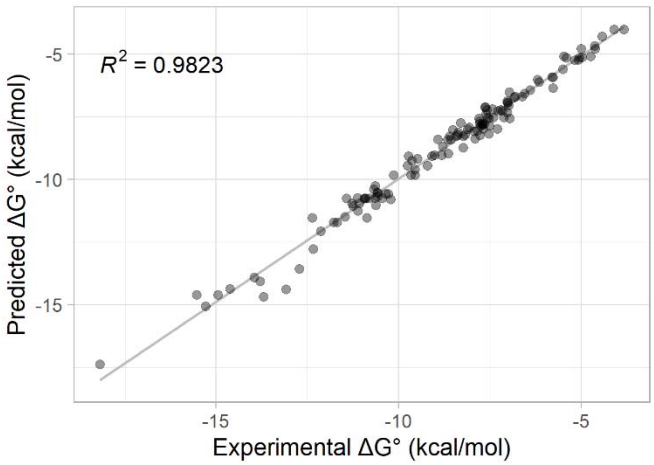


Figure 2: Correlation between predicted and measured ΔG°_{37} for duplexes with only WCF pairs. ΔG°_{37} values predicted from updated nearest neighbor parameters for duplexes composed solely canonical WCFW-C-F base pairs in Table 1A plotted against values determined from optical melting experiments.

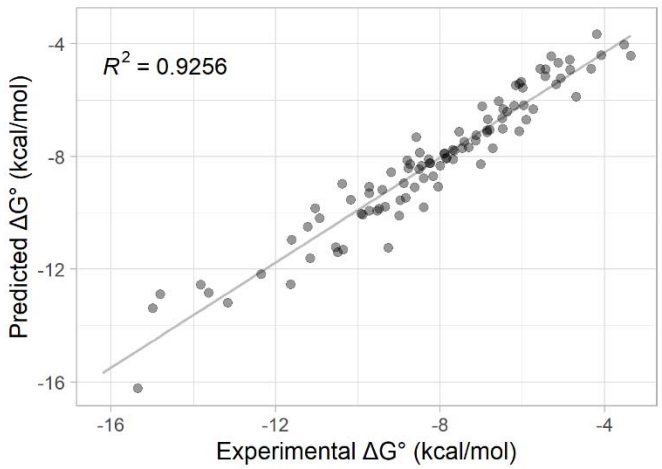


Figure 3: Correlation between predicted and observed ΔG°_{37} for duplexes with WCF and GU pairs. ΔG°_{37} values predicted from parameters in Table 1 plotted against values determined from optical melting experiments.

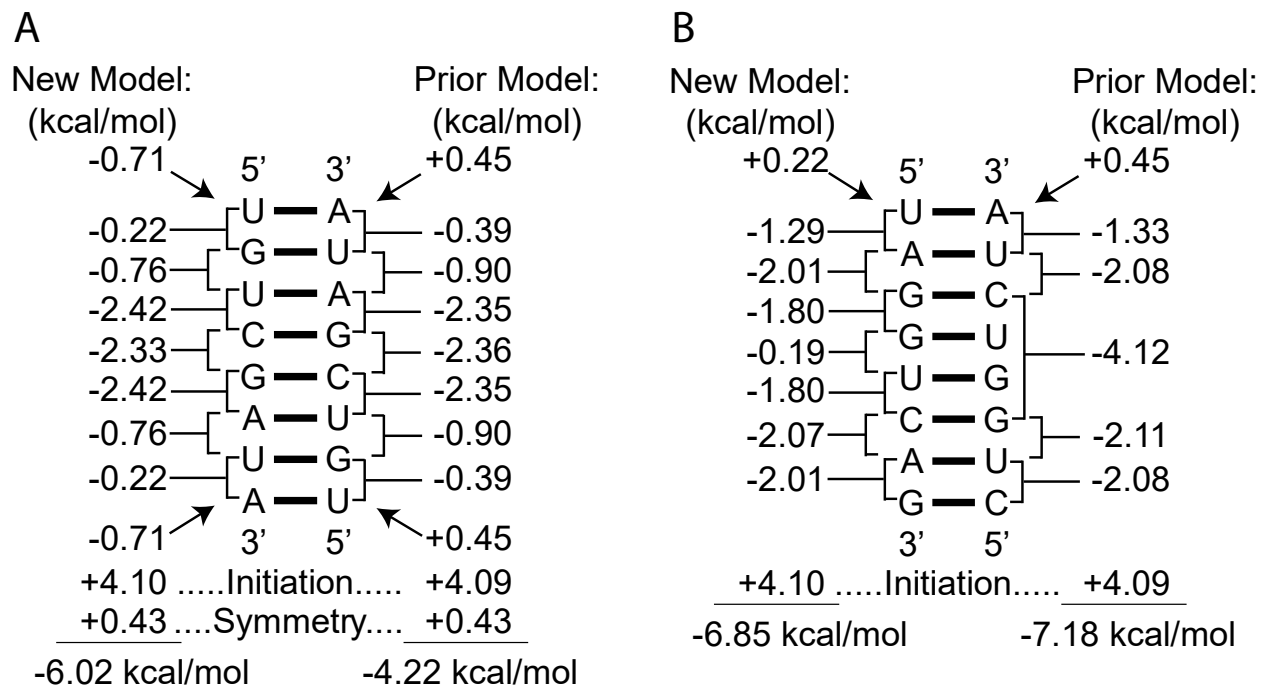


Figure 4. Example calculations of helical ΔG°_{37}. Panel A shows the stability calculation for (5'UGUCGAUA)$_2$, which is shown in Table 2 to have an experimentally determined ΔG°_{37} of -6.10 kcal/mol. This sequence is self-complementary and therefore the symmetry penalty is added. Panel B shows the stability calculation for 5'UAGGUCAG paired to 5'CUGGUCUA. This demonstrates the difference in treatment for the (GGUC)$_2$ motif. For both sequences, calculations are provided for the current parameters derived here and the previous parameters (23,24). The total stability is the sum of the stability increments.

Supplemental Information to Accompany:

RNA Helix Thermodynamics: The End Game

Jeffrey Zuber[1,†], Susan J. Schroeder[2,†], Hongying Sun[3,4], Douglas H. Turner[5], David H. Mathews[3,4,6,*]
[1]Alnylam Pharmaceuticals, Inc., Cambridge, MA 02142, USA, [2]Department of Chemistry and Biochemistry, and Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA, [3]Department of Biochemistry & Biophysics, University of Rochester, Rochester, NY 14642, USA, [4]Center for RNA Biology, University of Rochester, Rochester, NY 14642, USA, [5]Department of Chemistry, University of Rochester, Rochester, NY 14627, USA, and [6]Department of Biostatistics & Computational Biology, University of Rochester, Rochester, NY 14642, USA.
*To whom correspondence should be addressed. Tel: +1 585 275 1734; Email: David_Mathews@urmc.rochester.edu
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Supplemental Table 1. Duplex sequences used for parameter fitting.

Sequence	ΔH (kcal/mol)	ΔS (eu)	ΔG (37) (kcal/mol)	Ref
CAAAAAAG GUUUUUUC	-53.83	-158.7	-4.61	(33)
CAAAAAAAG GUUUUUUUC	-59.78	-175.1	-5.47	(33)
CCGG/	-34.21	-95.6	-4.56	(34)
ACCGGU/	-59.5	-164.5	-8.48	(34)
CCGG/	-31	-86.5	-4.17	(34)
GGCC/	-35.79	-98.1	-5.36	(35)
GGCC/	-41.3	-116	-5.32	(35)
AGGCCU/	-52	-140	-8.58	(36)

AGGCCU/	-48.19	-128.4	-8.37	(36)
CGGCCG/	-54.12	-142.57	-9.90	(36)
UGGCCA/	-59.9	-164	-9.04	(36)
GCCGGC/	-62.72	-166	-11.24	(36)
UCCGGA/	-51.92	-142.3	-7.79	(36)
GCGC/	-30.48	-83.4	-4.61	(37)
GCGCGC/	-65.98	-178.5	-10.62	(37)
CGCGCG/	-54.51	-146.4	-9.10	(37)
UGGCCG/	-53	-143.3	-8.56	(38)
UCCGGG/	-47.7	-129.8	-7.44	(38)
GCCGGU/	-58.2	-158.1	-9.17	(38)
GGCGCU/	-56.4	-154.7	-8.42	(38)
CGCG/	-31.8	-91.6	-3.39	(39)
UGCGCA/	-51.54	-139.7	-8.21	(39)
AGCGCU/	-50.07	-135.7	-7.98	(39)
GGCGCC/	-67.78	-182	-11.33	(39)
AUGCGCGU/	-54.9	-147	-9.31	(40)
AUGCGUAU/	-46.8	-133.9	-5.27	(40)
AUGUGCAU/	-57.1	-164.2	-6.17	(40)
GCUGGC/	-59.1	-169.7	-6.47	(40)
GGCGUC/	-38.1	-107.8	-4.67	(40)

CUGCGG/	-41.4	-119.6	-4.31	(40)
CGGCUG/	-43.2	-121.4	-5.55	(40)
AUGCAU/	-41.7	-119.2	-4.73	(40)
AUGCGCAU/	-64.39	-174.8	-10.18	(40)
AUACGUAU/	-54.36	-154.2	-6.53	(40)
AUGUACAU/	-55.91	-159.3	-6.50	(40)
GCUAGC/	-59.13	-165.1	-7.92	(40)
GACGUC/	-58.06	-163.5	-7.35	(40)
CUGCAG/	-55.41	-155.7	-7.12	(40)
CAGCUG/	-51.55	-144.7	-6.67	(40)
AGAUAUCU/	-64.51	-186.8	-6.57	(41)
AUCUAGAU/	-59.89	-169.9	-7.20	(41)
AACUAGUU/	-54.62	-153	-7.17	(41)
AGUUAACU/	-52.42	-148.5	-6.36	(41)
ACUUAAGU/	-47.23	-132.4	-6.17	(41)
GAACGUUC/	-77	-218.3	-9.29	(41)
GUUCGAAC/	-74.19	-211	-8.75	(41)
UCUAUAGA/	-62.06	-177.7	-6.95	(41)
UAGAUCUA/	-60.15	-170.6	-7.24	(41)
GUCGAC/	-53.63	-150.1	-7.08	(41)
ACUAUAGU/	-59.21	-168.4	-6.98	(41)

UGAUCA/	-44.73	-128	-5.03	(41)
GCAUGC/	-62.34	-177.2	-7.38	(42)
GUGCAC/	-59.61	-167.5	-7.66	(42)
GUCUAGAC/	-76.02	-212.5	-10.11	(42)
GAUAUAUC/	-62.04	-180.4	-6.09	(42)
GUAUAUAC/	-63.35	-185.1	-5.94	(42)
GAGAGA CUCUCU	-62.05	-178.1	-6.81	(42)
AGAGAGAG UCUCUCUC	-73.66	-201.7	-11.10	(42)
AAUGCAUU/	-59.81	-169.7	-7.18	(43)
CAUGCAUG/	-73.67	-206.3	-9.69	(43)
GAUGCAUC/	-72.75	-201.9	-10.13	(43)
UAUGCAUA/	-67.73	-195	-7.25	(43)
UGCGCA/	-50.8	-137.6	-8.12	(43)
GAUGCAUU/	-62.9	-180.8	-6.82	(44)
UAUGCAUG/	-62.3	-180.1	-6.44	(44)
GCGGCG CGCCGC	-58.5	-155	-10.43	(45)
UGACCUCA ACUGGAGU	-76.09	-205.55	-12.34	(46)
CGCGCG/	-54.74	-146.77	-9.22	(46)
CACAG GUGUC	-40.2	-115.4	-4.41	(47)
AUGGUCAU/	-57.4	-167.6	-5.42	(48)
CCUGUAGG/	-71.1	-207.3	-6.81	(48)

CGGGUCCG/	-81.4	-226.4	-11.18	(48)
CGUUGACG/	-73.5	-214.6	-6.94	(48)
CUGGUCAG/	-70.5	-204.4	-7.11	(48)
GGAGUUCC/	-73.1	-214.9	-6.45	(48)
GGAUGUCC/	-73	-208.4	-8.36	(48)
GGCGUGCC/	-73.4	-206.9	-9.23	(48)
GUCGUGAC/	-69.1	-203.3	-6.05	(48)
GAGUUGAG CUCGGCUC	-73.9	-211.8	-8.21	(48)
AGUCGAUU/	-53.3	-152.6	-5.97	(48)
CGGAUUCG/	-72.6	-213	-6.54	(48)
CCUAGG/	-54.1	-149.1	-7.86	(48)
CGUACG/	-46.6	-133.1	-5.32	(48)
GGAUCC/	-53.7	-149.1	-7.46	(48)
GAGCUC/	-54.85	-151.94	-7.73	(49)
GCGUGC/	-46.18	-132.43	-5.11	(49)
GAGGUCUC/	-82.8	-238.7	-8.77	(50)
GAGCUC/	-62.3	-175.3	-7.93	(50)
GGUACC/	-54.9	-153.4	-7.32	(50)
GAGUGCUC/	-83	-237.4	-9.37	(51)
GGUUGACC/	-78.3	-225.9	-8.24	(51)
GGCUGGCC/	-87.2	-238.8	-13.14	(51)

CCAUGUGG/	-70.5	-202.1	-7.82	(52)
CCAGUUGG/	-61.1	-178.6	-5.71	(52)
CCAUGG/	-56.93	-159.9	-7.34	(52)
GAGGAG **CUCCUC**	-55.7	-152.2	-8.50	(53)
GAGUGGAG **CUCGUCUC**	-82.3	-234.1	-9.69	(53)
GAGUUGAG **CUCGGCUC**	-70	-199.2	-8.22	(53)
GAGGUGAG **CUCUGCUC**	-78.4	-228.2	-7.62	(53)
CGCG/	-33.31	-95.6	-3.66	(53)
ACGCA **UGCGU**	-45.4	-130.4	-4.96	(53)
AGCGA **UCGCU**	-46.31	-133	-5.06	(53)
GCACG **CGUGC**	-45.31	-126.2	-6.17	(53)
GCUCG **CGAGC**	-43.38	-120.1	-6.13	(53)
CACGUG/	-50.31	-141	-6.58	(53)
CCGCGG/	-60.79	-164.3	-9.83	(53)
GCAACG **CGUUGC**	-50.57	-140.5	-6.99	(53)
GCAUCG **CGUAGC**	-51.89	-143.9	-7.26	(53)
GCCGCG **CGGCGC**	-59.69	-157.4	-10.87	(53)
GCGCCG **CGCGGC**	-57.85	-151.3	-10.92	(53)
GCGCGG **CGCGCC**	-71.16	-192.7	-11.39	(53)
GCGUCG **CGCAGC**	-52.38	-140.6	-8.77	(53)
GCUACG **CGAUGC**	-58.02	-162.7	-7.56	(53)

GGCGCG CCGCGC	-63.53	-170.1	-10.77	(53)
GUGGUG CACCAC	-48.84	-132.7	-7.68	(53)
GUGUCG CACAGC	-50.88	-140.9	-7.18	(53)
UCAUGA/	-41.89	-121.2	-4.30	(53)
UCGCGA/	-48.94	-135.7	-6.85	(53)
UCUAGA/	-36.53	-101.8	-4.96	(53)
AAGGAGG UUCCUCC	-58.72	-158.6	-9.53	(53)
ACUGUCA UGACAGU	-52.24	-142.9	-7.92	(53)
AGUCUGA UCAGACU	-51.48	-141.8	-7.50	(53)
GACUCAG CUGAGUC	-64.11	-177.5	-9.06	(53)
GAGUGAG CUCACUC	-70.49	-196	-9.70	(53)
GUCACUG CAGUGAC	-57.81	-158.6	-8.62	(53)
ACCUUUGC UGGAAACG	-77.42	-215.3	-10.64	(53)
CGACGCAG GCUGCGUC	-70.45	-187.43	-12.32	(53)
CUCGCACA GAGCGUGU	-72.61	-195.09	-12.10	(53)
GGCUUCAA CCGAAGUU	-61.59	-165.7	-10.20	(53)
UCCUUGCA AGGAACGU	-70.27	-190.82	-11.09	(53)
UUCCGGAA/	-67.43	-182.6	-10.80	(53)
UUGCGCAA/	-62.17	-167.6	-10.19	(53)
UUGGCCAA/	-63.68	-169.8	-11.02	(53)
UUGUACAA/	-49.45	-137.8	-6.71	(53)

AAGGUUGGAA UUCCAACCUU	-75.84	-203.62	-12.69	(53)
CAUGCG GUACGC	-48.57	-134	-7.01	(53)
GAGCUG CUCGAC	-51.59	-142.2	-7.49	(53)
GCUGAG CGACUC	-55.85	-155.2	-7.71	(53)
GUGCAG CACGUC	-55.94	-155.6	-7.68	(53)
UAAGGUA AUUCCAU	-51.27	-142.9	-6.95	(53)
GAGAUCUC/	-75	-209.2	-10.12	(53)
GCCAUGGC/	-93.91	-254.2	-15.07	(53)
GCUGCGAC CGACGCUG	-86.18	-232.95	-13.93	(53)
UCCGCGCA AGGCGCGU	-81.15	-214.61	-14.59	(53)
GAGUGAG CUCGCUC	-64.2	-181.8	-7.81	(54)
GCGGCG CGCCGC	-61.5	-163.1	-10.91	(54)
GCGCGC/	-68.2	-185	-10.82	(54)
AAGGCCGGAA UUCCGGCCUU	-98.34	-258.54	-18.15	(28)
CCUCUGGUGA GGAGACCGCU	-93.46	-251.96	-15.31	(28)
AGGCCGGA UCCGGCCU	-84.09	-221.91	-15.26	(28)
GAGCCGAC CUCGGCUG	-80.13	-213.99	-13.76	(28)
UCACCUGA AGUGGACU	-71.51	-195.62	-10.84	(28)
CUGGUC GACCAG	-55.1	-151.7	-8.05	(55)
CCAGCGUCCU GGUUGUAGGA	-87.9	-246	-11.60	(56)
GAGGAC CUCCUG	-57.3	-157.1	-8.58	(56)

GCAUGUGC/	-72.4	-206.1	-8.48	(57)
GCAGUUGC/	-64.8	-190	-5.87	(57)
GCUGGUGC CGAUUACG	-69.4	-199.1	-7.65	(57)
GUCUAGAU/	-70	-201	-7.66	(58)
AUCUAGAU/	-59.1	-166.9	-7.34	(58)
GGCAUGCC/	-90.2	-244.1	-14.49	(59)
GGCUAGCC/	-76.2	-203	-13.24	(59)
GGCAAGGC CCGUUCCG	-72.8	-192.6	-13.07	(59)
GGCGGGGC CCGUUCCG	-76.5	-202.2	-13.79	(59)
UCCGCC AGGUGG	-57	-162.2	-6.69	(60)
GAGUUGAC CUCAACUG	-88.7	-251.8	-10.60	(60)
GUUGCAGU/	-45.2	-128.3	-5.41	(61)
UGCGAUCCUG/	-69.5	-201.6	-6.97	(61)
AGCGCU/	-49.5	-134.5	-7.78	(61)
UUGCAG/	-37.2	-106.5	-4.17	(61)
GUGCAU/	-47.5	-136.9	-5.04	(61)
UUUGCAGG/	-46.1	-133.1	-4.82	(61)
UUAUCGAUGG/	-61.1	-175	-6.82	(61)
GGAUGCAUUU/	-74.4	-214.2	-7.97	(61)
UGAUGCAUUG/	-75.4	-216.9	-8.13	(61)
GUAUCGAUGU/	-63.4	-180.2	-7.51	(61)

GUCGCGGU/	-57.8	-160.5	-8.02	(61)
GUGAUCGU/	-47.1	-132.5	-6.01	(61)
UGUCGAUG/	-68.8	-201.4	-6.34	(61)
GGCAUGUU/	-69.1	-201	-6.76	(61)
UGCUAGUG/	-74.9	-217.7	-7.38	(25)
GGUCGCGGUU/	-66.5	-186	-8.81	(25)
GUGAUGCAUUGU/	-93.7	-270.9	-9.68	(25)
UGUGGCCGUG/	-66.2	-178.3	-10.90	(25)
UGUCGGUG/	-57.3	-173.5	-3.49	(25)
UUGGCUGG/	-49.1	-141.7	-5.15	(25)
GGUGGCCGUU/	-66.4	-178.6	-11.01	(25)
UUGCUAGUGG/	-71.9	-206.5	-7.85	(25)
GUGGCUGU/	-52.3	-149.1	-6.06	(25)
GUGCUAGUGU/	-74	-211.3	-8.47	(25)
UUGUCGGUGG/	-64	-195.6	-3.33	(25)
GUGUCGGUGU/	-65.3	-197.5	-4.05	(25)
UGUGGCUGUG/	-55.5	-159.8	-5.94	(25)
UAUCGAUA/	-61.8	-176.2	-7.15	(25)
UUAUCGAUAA/	-82.3	-235.8	-9.17	(25)
GCGGGAC CGCUCUG	-45.2	-116.8	-8.97	(24)
CAGGGCUC GUCUCGAG	-62.8	-166.6	-11.13	(24)

CUGGCUAG/	-60.38	-171.8	-7.10	(24)
GACGCCAG CUGUGGUC	-63.8	-171.8	-10.52	(24)
GACGCGUU/	-62.2	-169.9	-9.51	(24)
GCAGCUGU/	-72.3	-199.8	-10.33	(24)
GGAGCUCU/	-66.57	-180.9	-10.46	(24)
GUAGCUAU/	-50.3	-138.7	-7.28	(24)
GUCGGGCC CAGCUCGG	-96	-261.3	-14.96	(24)
UACCGGUG/	-51.7	-135.4	-9.71	(24)
UCACGUGG/	-46.9	-124.2	-8.38	(24)
UGACGUCG/	-63.8	-172.3	-10.36	(24)
UUACGUAG/	-44.6	-124	-6.14	(24)
CAGAGGAGAC GUCUUUUCUG	-98.95	-288.6	-9.44	(24)
CAGCGCGUUG/	-77.02	-208.6	-12.32	(24)
CAGUCGAUUG/	-92.33	-269.7	-8.68	(24)
CCGAAUUUGG/	-76.76	-225	-6.98	(24)
CGGAAUUUCG/	-90.52	-266.5	-7.87	(24)
CGGAUAUUCG/	-88.2	-256.2	-8.74	(24)
CGGGCGUUCG/	-101.66	-290.5	-11.56	(24)
CGGUGCAUCG/	-102.42	-282.6	-14.77	(24)
CUGGAUUCAG/	-97.81	-282.7	-10.13	(24)
GAGAGCUUUC/	-86.57	-250.6	-8.85	(24)

GAGGAUCUUC/	-93.86	-270.9	-9.84	(24)
GAGUGGAGAG CUCAUUUCUC	-96.93	-280.7	-9.87	(24)
GGUUCGGGCC/	-115.71	-329.3	-13.58	(24)
GUGAAUUUAC/	-62.63	-186.4	-4.82	(24)
GUGUGCAUAC/	-58.6	-160.1	-8.94	(24)
GUUAGCUGAC/	-69.6	-196.7	-8.59	(24)

Supplemental Table 2

A) WCF helical parameters annotated with errors estimated as standard errors of regression. These error estimates are not as good as those in Table 1 derived by perturbing the input experimental data.

	New Model		*1998 Model* †	
Feature	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)
GC/CG	-3.46 ± 0.07	-16.52 ± 0.84	-3.42 ± 0.08	-14.88 ± 1.58
CC/GG	-3.28 ± 0.05	-13.94 ± 0.66	-3.26 ± 0.07	-13.39 ± 1.24
GA/CU	-2.42 ± 0.05	-13.75 ± 0.64	-2.35 ± 0.06	-12.44 ± 1.20
CG/GC	-2.33 ± 0.08	-9.61 ± 1.00	-2.36 ± 0.09	-10.64 ± 1.65
AC/UG	-2.25 ± 0.06	-11.98 ± 0.76	-2.24 ± 0.06	-11.40 ± 1.23
CA/GU	-2.07 ± 0.06	-10.47 ± 0.76	-2.11 ± 0.07	-10.44 ± 1.28
AG/UC	-2.01 ± 0.06	-9.34 ± 0.78	-2.08 ± 0.06	-10.48 ± 1.24
UA/AU	-1.29 ± 0.09	-9.16 ± 1.10	-1.33 ± 0.09	-7.69 ± 2.02
AU/UA	-1.09 ± 0.08	-8.91 ± 1.01	-1.10 ± 0.08	-9.38 ± 1.68
AA/UU	-0.94 ± 0.04	-7.44 ± 0.52	-0.93 ± 0.03	-6.82 ± 0.79
Initiation	+4.10 ± 0.20	+4.66 ± 2.47	+4.09 ± 0.22	+3.61 ± 4.12
AU End on AU	+0.22 ± 0.06	+4.36 ± 0.76	+0.45 ± 0.04 ‡	+3.72 ± 0.83 ‡
AU End on CG	+0.44 ± 0.04	+3.17 ± 0.55	+0.45 ± 0.04 ‡	+3.72 ± 0.83 ‡

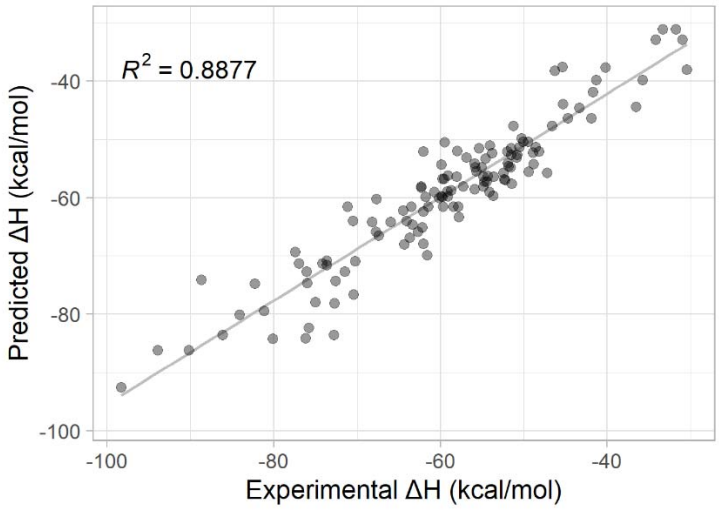
B) GU stack parameters annotated with errors estimated as standard errors of regression. These error estimates are not as good as those in Table 1 derived by perturbing the input experimental data.

	New Model		*2012 Model* ¶	
Feature	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)
GC/UG	-2.23 ± 0.12	-14.73 ± 1.32	-2.15 ± 0.10	-11.09 ± 1.78
CU/GG	-1.93 ± 0.13	-9.26 ± 1.45	-1.77 ± 0.09	-9.44 ± 1.76
GG/CU	-1.80 ± 0.11	-12.41 ± 1.28	-1.80 ± 0.09	-7.03 ± 1.75
CG/GU	-1.05 ± 0.12	-5.64 ± 1.37	-1.25 ± 0.09	-5.56 ± 1.68
AU/UG	-0.76 ± 0.13	-9.23 ± 1.50	-0.90 ± 0.08	-7.39 ± 1.65
GA/UU	-0.60 ± 0.11	-10.58 ± 1.26	-0.51 ± 0.08	-10.38 ± 1.79
UG/GU	-0.38 ± 0.15	-8.76 ± 1.75	-0.57 ± 0.19	-12.64 ± 4.01
UA/GU	-0.22 ± 0.13	-2.72 ± 1.47	-0.39 ± 0.09	-0.96 ± 1.80
GG/UU	-0.20 ± 0.15	-9.06 ± 1.71	-0.25 ± 0.16	-17.82 ± 3.75
GU/UG	-0.19 ± 0.16	-7.66 ± 1.85	+0.72 ± 0.19	-13.83 ± 4.21
AG/UU	+0.02 ± 0.12	-5.10 ± 1.38	-0.35 ± 0.08	-3.96 ± 1.73
GGUC/CUGG	(-3.79 ± 0.27)†	(-32.48 ± 3.16)†	-4.12 ± 0.54	-30.80 ± 8.87
AU End on GU	-0.71 ± 0.30	5.16 ± 3.42	+0.45 ± 0.04*	+3.72 ± 0.83*
GU End on CG	0.13 ± 0.14	3.91 ± 1.63	0.00 ± 0.00 ‡	0.00 ± 0.00 ‡
GU End on AU	-0.31 ± 0.14	3.65 ± 1.56	0.00 ± 0.00 ‡	0.00 ± 0.00 ‡
GU End on GU	-0.74 ± 0.16	6.23 ± 1.89	0.00 ± 0.00 ‡	0.00 ± 0.00 ‡

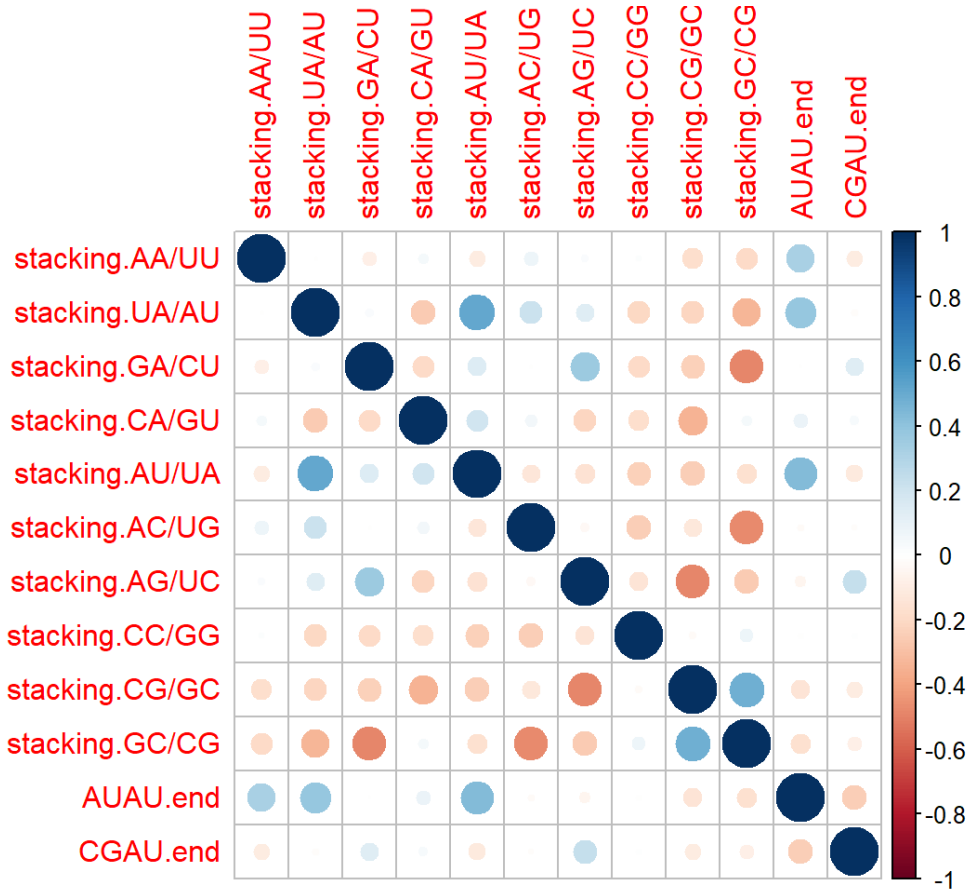
Supplemental Table 3: Comparison of predictions of models with and without 5'GGUC/3'CUGG parameter.

	Model without 5'GGUC/3'GGUC[§]		Model with 5'GGUC/3'GGUC[§]		2012 Model[¶]	
Duplex	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)	ΔG°_{37} (kcal/mol)	ΔH (kcal/mol)
AUGGUCAU UACUGGUA	-5.14 ± 0.18	-57.86 ± 3.88	-5.50 ± 0.21	-62.57 ± 5.28	-5.12 ± 1.23	-59.39 ± 5.31
CGGGUCCG GCCUGGGC	-10.48 ± 0.15	-74.92 ± 2.98	-10.82 ± 0.20	-79.41 ± 5.06	-10.84 ± 1.18	-75.25 ± 4.95
CUGGUCAG GACUGGUC	-7.42 ± 0.14	-67.44 ± 3.16	-7.78 ± 0.17	-71.97 ± 4.89	-7.98 ± 1.13	-69.03 ± 4.80
GAGGUCUC CUCUGGAG	-8.12 ± 0.14	-74.00 ± 2.89	-8.48 ± 0.18	-78.57 ± 4.90	-8.46 ± 1.11	-73.03 ± 4.77

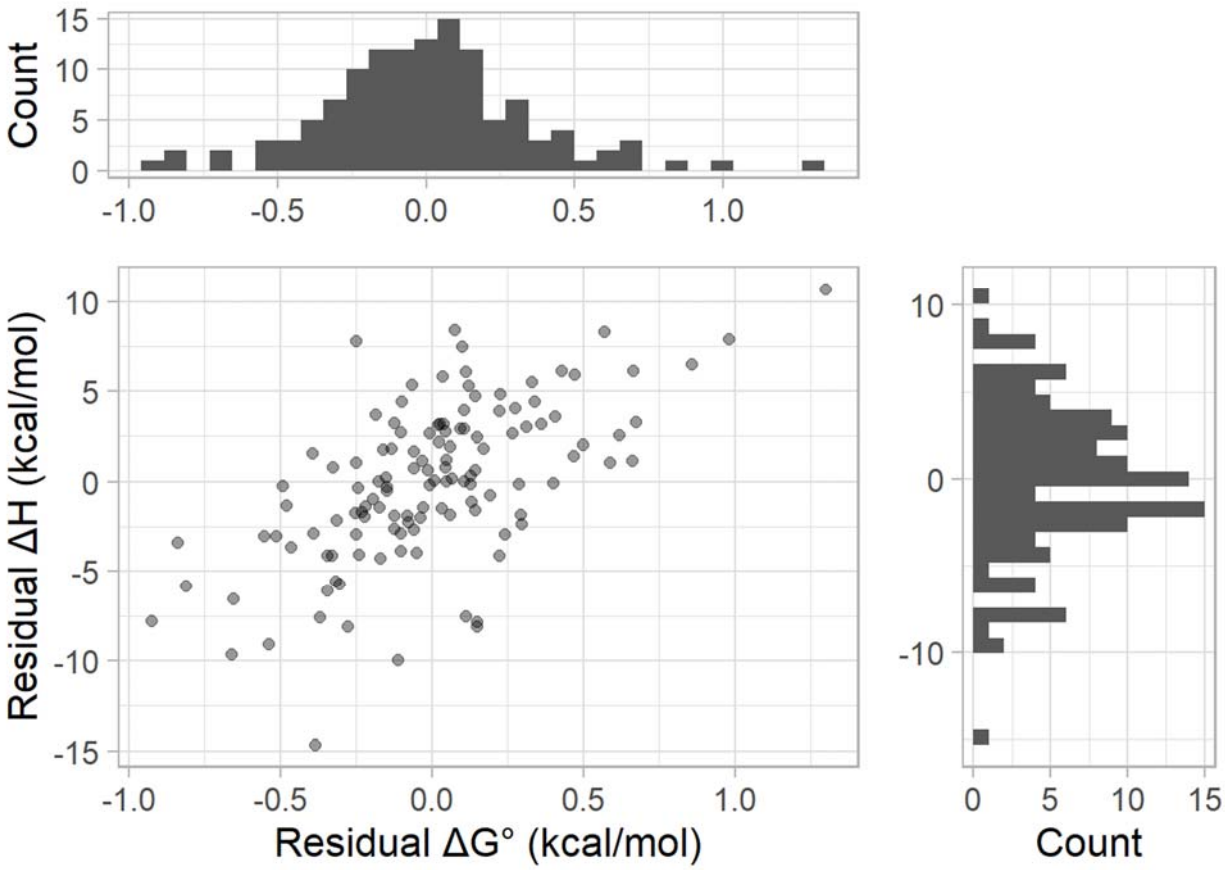
[§]Uncertainty values were calculated from the experiment covariation analysis. [¶]Uncertainty values were derived from standard errors of regression (24)



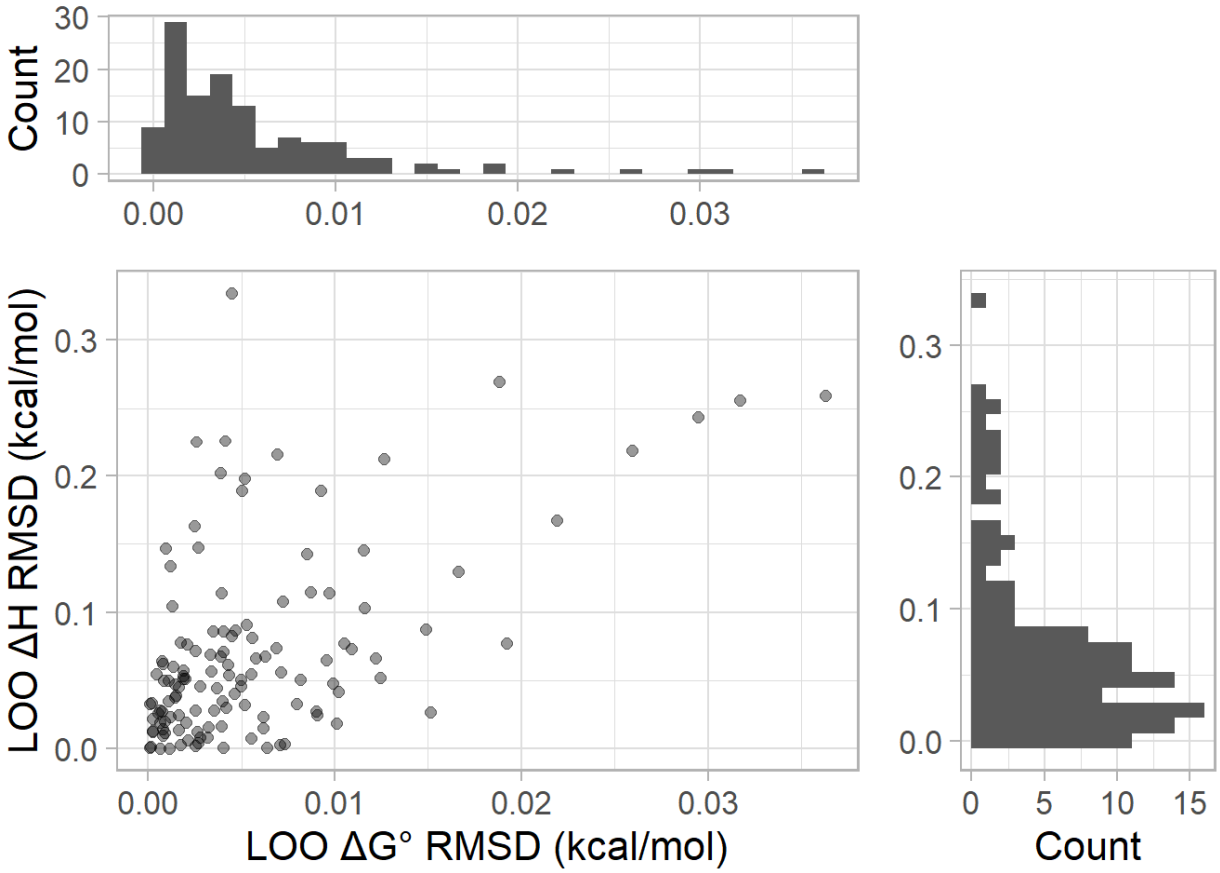
Supplemental Figure 1: ΔH° predictions as a function of experimentally-determine ΔH° for canonical WCF nearest neighbor parameters. ΔH° values predicted from updated nearest neighbor parameters for duplexes composed solely canonical WCF base pairs in Table 1A plotted against values determined from optical melting experiments.



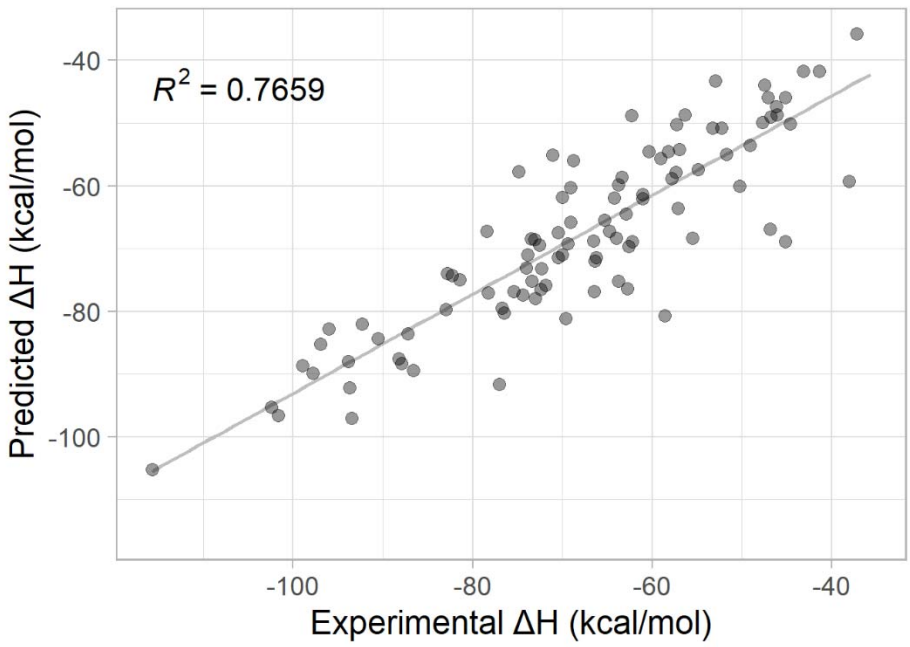
Supplemental Figure 2: Pearson correlations between model features. Pairwise Pearson correlation coefficients were calculated for the observed frequencies of WCF stacking parameters. The correlation coefficients are encoded in the size and color of each circle.



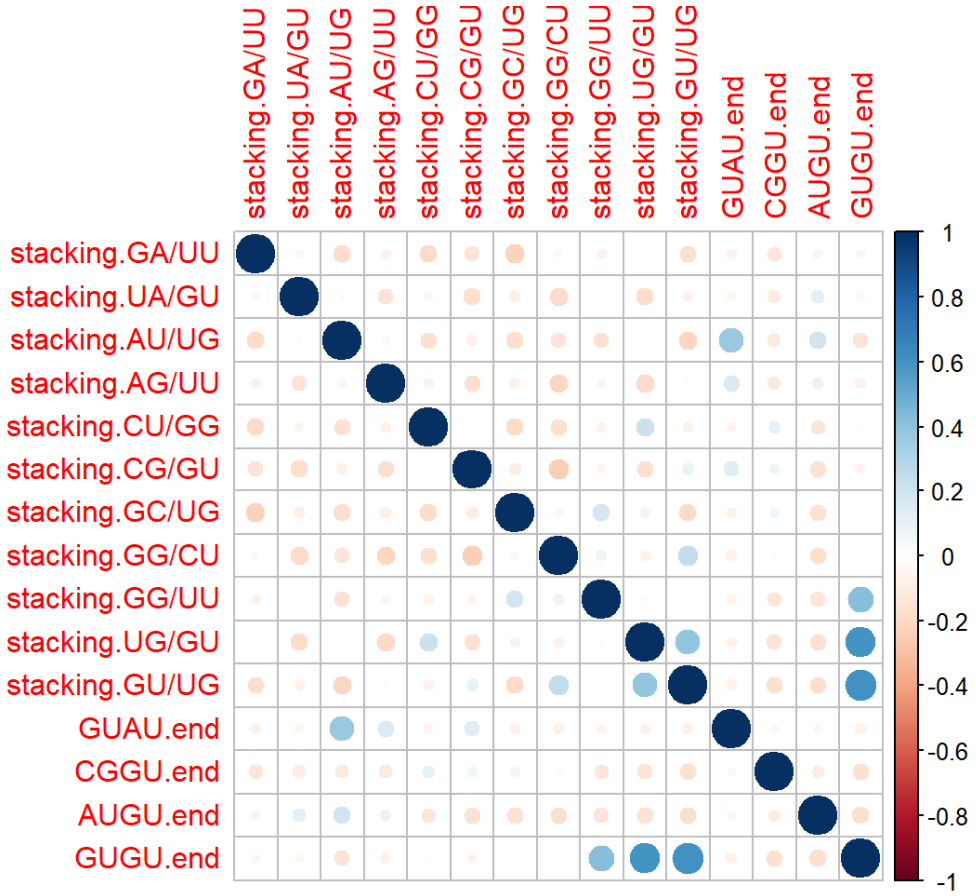
Supplemental Figure 3: Residuals for the Watson-Crick-Franklin stacking parameter predictions. Shown are the scatter plot between the residuals in ΔH° and folding $\Delta G°_{37}$ change predictions (main plot), a histogram of the ΔH° residuals (right plot), and a histogram of the $\Delta G°_{37}$.residuals (top plot).



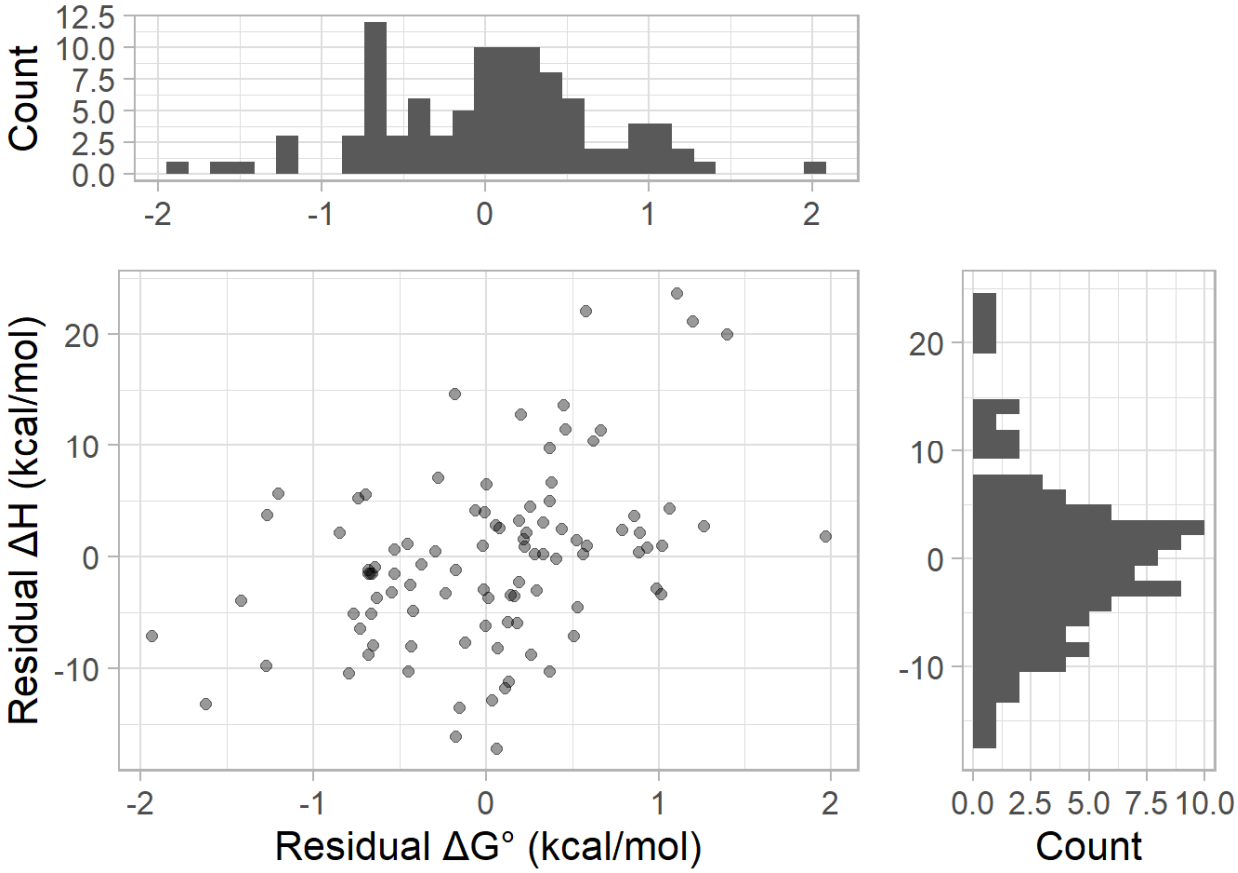
Supplemental Figure 4: Leave-One-Out (LOO) analysis for WCF helix parameters. The plotted value is the RMSD of the parameter values for leaving out a single experiment as compared to when all experiments are used.



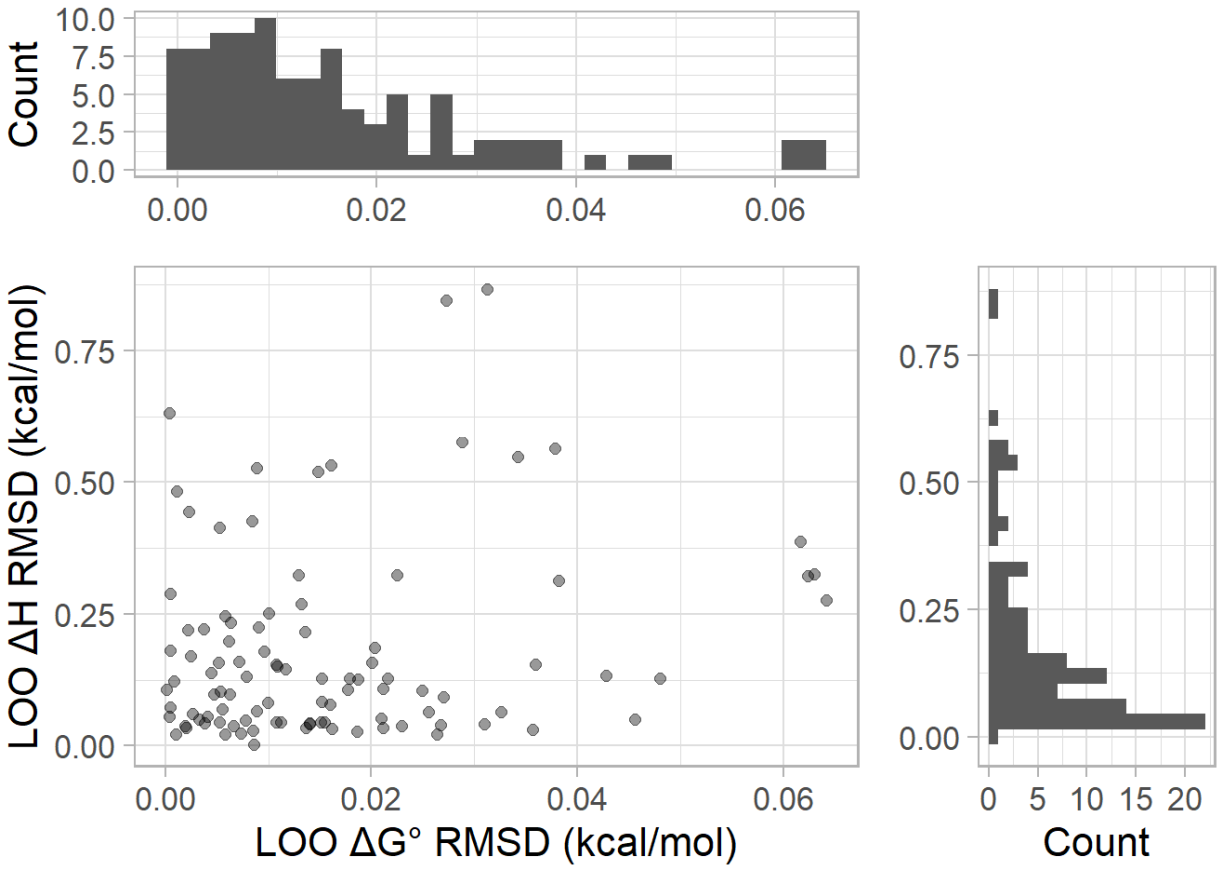
Supplemental Figure 5: Correlation between predicted and observed ΔH° for duplexes with WCF and GU pairs. ΔH° values predicted from parameters in Table 1 plotted against values determined from optical melting experiments.



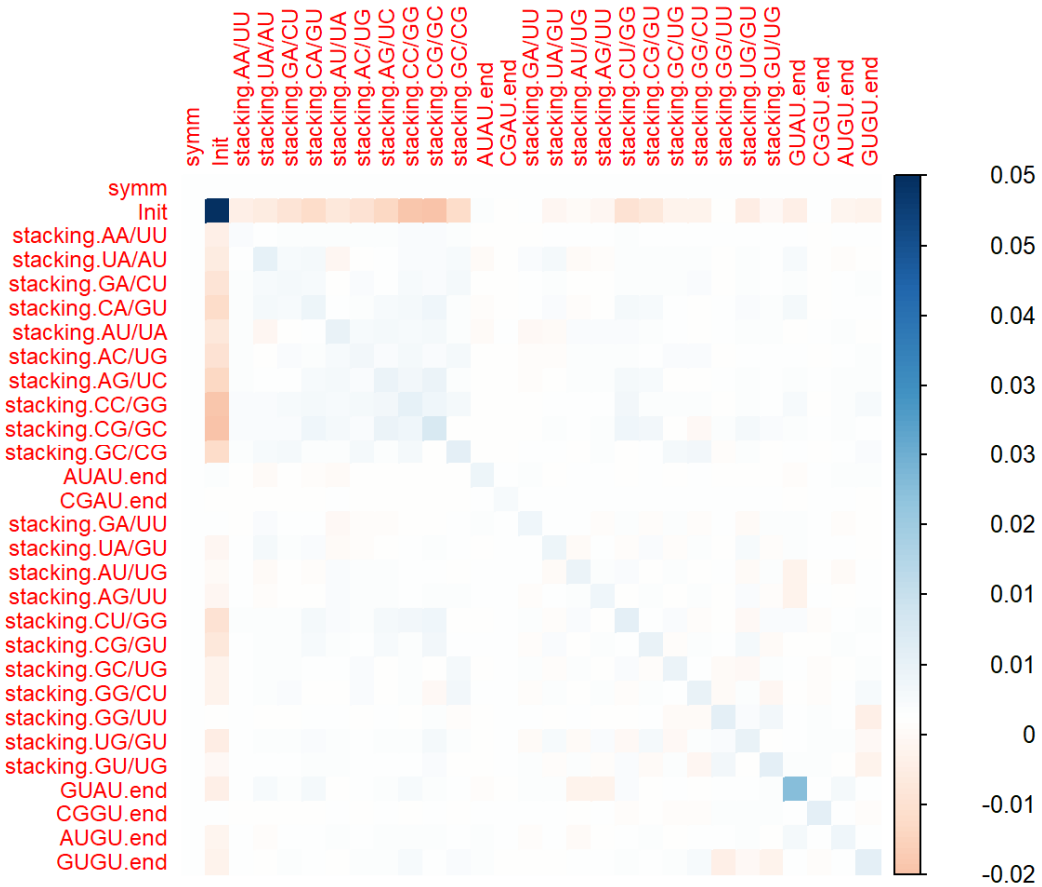
Supplemental Figure 6: Pearson correlation between features of GU stacking model. Pairwise Pearson correlation coefficients were calculated for the observed frequencies of GU stacking parameters. The correlation coefficients are encoded in the size and color of each circle.



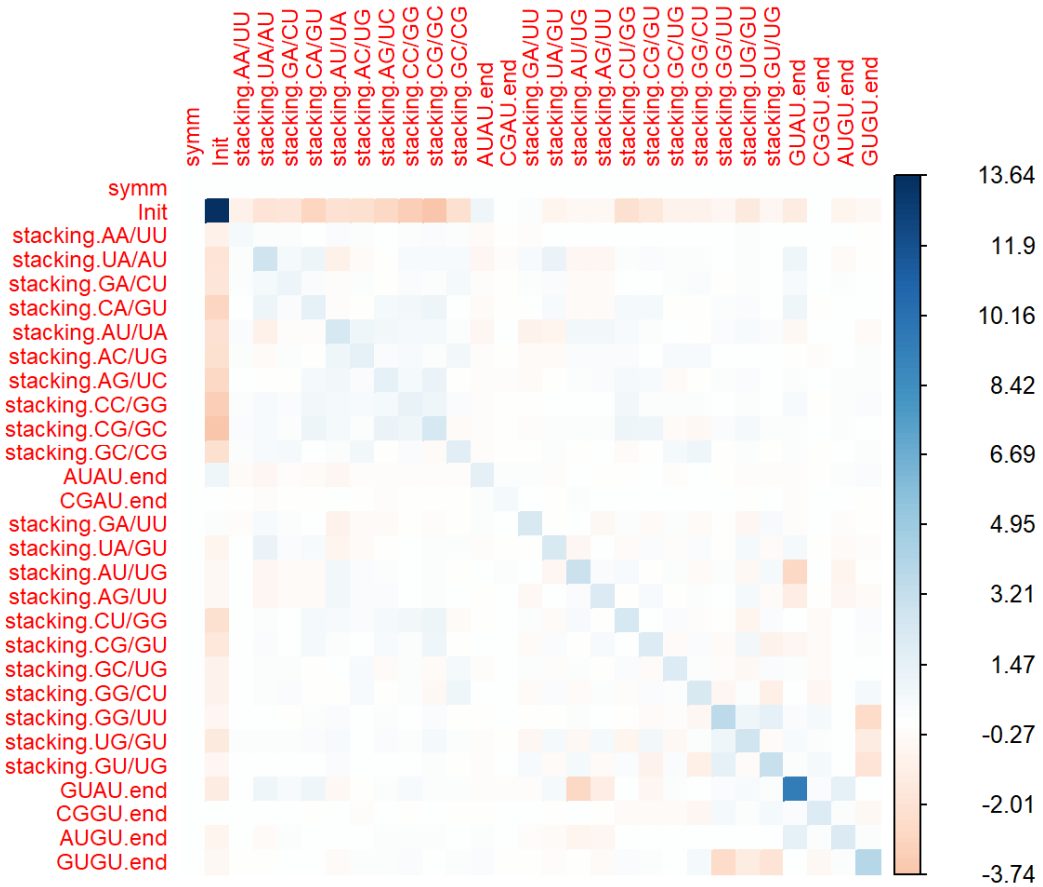
Supplemental Figure 7: Residuals for the GU stacking parameter predictions. Shown are the scatter plot between the residuals in ΔH° and folding ΔG°_{37} predictions (bottom left plot), a histogram of the folding ΔG°_{37} residuals (top plot), and a histogram of the ΔH° residuals right plot).



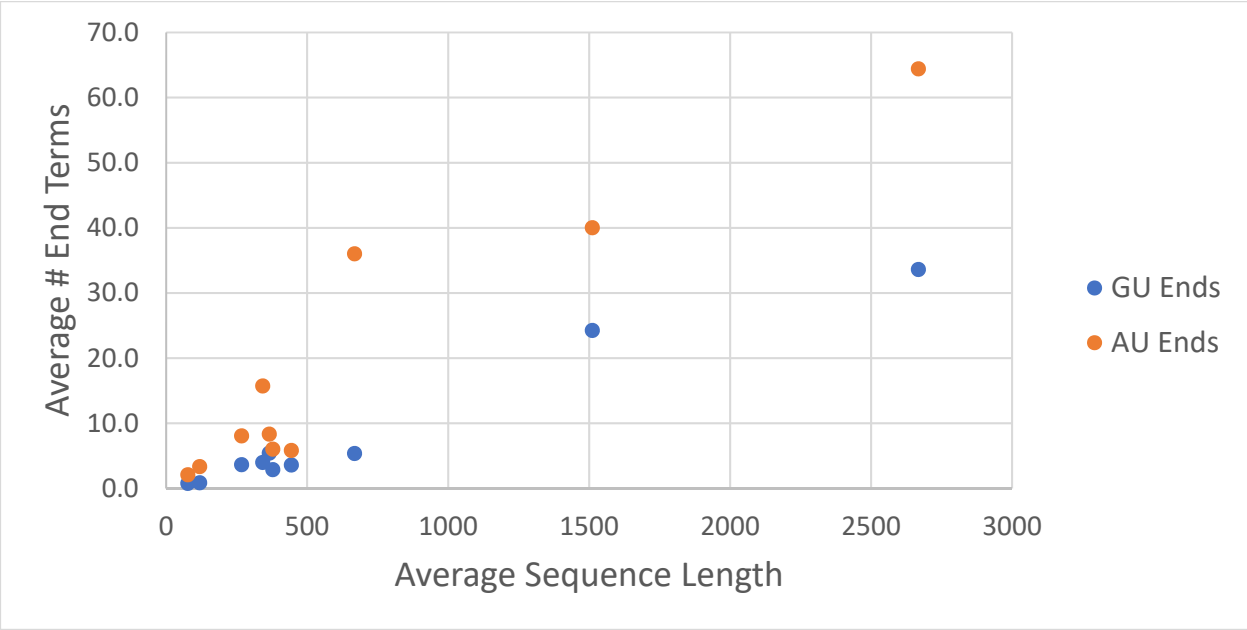
Supplemental Figure 8: Leave-One-Out (LOO) analysis for GU stacking parameters. The plotted value is the RMSD of the parameter values for leaving out a single experiment as compared to when all experiments are used.



Supplemental Figure 9. The covariation between ΔG°_{37} parameter values. Pairwise covariation values for ΔG°_{37} values fit to 1,000 randomly perturbed sets of optical melting experiments.



Supplemental Figure 10. The covariation between ΔH° parameter values. Pairwise covariation values for ΔH° values fit to 1,000 randomly perturbed sets of optical melting experiments.



Supplemental Figure 11: AU and GU helix end counts across RNA families. The accepted secondary structures of RNA sequences in different RNA families were parsed to count the number of AU and GU helix ends that existed in the structures. The RNA families included 5S rRNA, 16S rRNA, 23S rRNA, Group 1 and Group 2 introns, RNAP, SRP, telomerase, tmRNA, and tRNA. AU and GU ends were counted if they closed exterior loops, interior loops, hairpin loops, multibranch loops, or bulge loops larger than a single nucleotide.

References

1. Encode Project Consortium. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.
2. Encode Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A. *et al.* (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, **583**, 699-710.
3. Doudna, J.A. and Cech, T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222-228.
4. Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr Opin Struct Biol*, **15**, 342-348.
5. Serganov, A. and Nudler, E. (2013) A decade of riboswitches. *Cell*, **152**, 17-24.
6. Wu, L. and Belasco, J.G. (2008) Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol Cell*, **29**, 1-7.
7. Blackburn, E.H. and Collins, K. (2011) Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol*, **3**.
8. Dasgupta, A., Baron, M.H. and Baltimore, D. (1979) Poliovirus replicase: a soluble enzyme able to initiate copying of poliovirus RNA. *Proc Natl Acad Sci U S A*, **76**, 2679-2683.
9. Ohman, M. and Bass, B. (2001) In Soll, D., Nishimura, S. and Moore, P. B. (eds.), *RNA*. Pergamon/Elsevier Science, pp. 125-136.
10. Simpson, L. (2006) In R.F. Gesteland, T. R. C., and J.F. Atkins (ed.), *The RNA World*. Cold Spring Harbor Laboratory Press, pp. 401-417.
11. Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A. *et al.* (2017) RNA targeting with CRISPR-Cas13. *Nature*, **550**, 280-284.
12. Lambowitz, A.M. and Zimmerly, S. (2011) Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*, **3**, a003616.
13. Corbett, K.S., Edwards, D.K., Leist, S.R., Abiona, O.M., Boyoglu-Barnum, S., Gillespie, R.A., Himansu, S., Schafer, A., Ziwawo, C.T., DiPiazza, A.T. *et al.* (2020) SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature*, **586**, 567-571.
14. Watkins, A.M., Rangan, R. and Das, R. (2020) FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds. *Structure*, **28**, 963-976 e966.
15. Gruber, A.R., Findeiss, S., Washietl, S., Hofacker, I.L. and Stadler, P.F. (2010) RNAz 2.0: improved noncoding RNA detection. *Pac Symp Biocomput*, 69-79.
16. Fu, Y., Xu, Z.Z., Lu, Z.J., Zhao, S. and Mathews, D.H. (2015) Discovery of Novel ncRNA Sequences in Multiple Genome Alignments on the Basis of Conserved and Stable Secondary Structures. *PLoS One*, **10**, e0130200.
17. Petrov, A.I., Kay, S.J.E., Gibson, R., Kulesha, E., Staines, D., Bruford, E.A., Wright, M.W., Burge, S., Finn, R.D., Kersey, P.J. *et al.* (2015) RNAcentral: an international database of ncRNA sequences. *Nucleic Acids Res*, **43**, D123-129.
18. 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.
19. Keel, A.Y., Rambo, R.P., Batey, R.T. and Kieft, J.S. (2007) A general strategy to solve the phase problem in RNA crystallography. *Structure*, **15**, 761-772.
20. Ding, P., Kharytonchyk, S., Waller, A., Mbaekwe, U., Basappa, S., Kuo, N., Frank, H.M., Quasney, C., Kidane, A., Swanson, C. *et al.* (2020) Identification of the initial nucleocapsid recognition element in the HIV-1 RNA packaging signal. *Proc Natl Acad Sci U S A*, **117**, 17737-17746.

21. Westhof, E., Yusupov, M. and Yusupova, G. (2019) The multiple flavors of GoU pairs in RNA. *J Mol Recognit*, **32**, e2782.
22. Andronescu, M., Condon, A., Turner, D.H. and Mathews, D.H. (2014) The determination of RNA folding nearest neighbor parameters. *Methods Mol Biol*, **1097**, 45-70.
23. 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.
24. Chen, J.L., Dishler, A.L., Kennedy, S.D., Yildirim, I., Liu, B., Turner, D.H. and Serra, M.J. (2012) Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*, **51**, 3508-3522.
25. Nguyen, M.T. and Schroeder, S.J. (2010) Consecutive terminal GU pairs stabilize RNA helices. *Biochemistry*, **49**, 10574-10581.
26. 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.
27. Sponer, J., Leszczynski, J. and Hobza, P. (2001) Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers*, **61**, 3-31.
28. Schroeder, S.J. and Turner, D.H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry*, **39**, 9257-9274.
29. Gu, X., Mooers, B.H., Thomas, L.M., Malone, J., Harris, S. and Schroeder, S.J. (2015) Structures and Energetics of Four Adjacent G.U Pairs That Stabilize an RNA Helix. *J Phys Chem B*, **119**, 13252-13261.
30. Rinnenthal, J., Klinkert, B., Narberhaus, F. and Schwalbe, H. (2010) Direct observation of the temperature-induced melting process of the Salmonella fourU RNA thermometer at base-pair resolution. *Nucleic Acids Res*, **38**, 3834-3847.
31. Bothe, J.R., Nikolova, E.N., Eichhorn, C.D., Chugh, J., Hansen, A.L. and Al-Hashimi, H.M. (2011) Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods*, **8**, 919-931.
32. Huang, Y., Weng, X. and Russu, I.M. (2011) Enhanced base-pair opening in the adenine tract of a RNA double helix. *Biochemistry*, **50**, 1857-1863.
33. Nelson, J.W., Martin, F.H. and Tinoco, I., Jr. (1981) DNA and RNA oligomer thermodynamics: the effect of mismatched bases on double-helix stability. *Biopolymers*, **20**, 2509-2531.
34. 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.
35. Freier, S.M., Burger, B.J., Alkema, D., Neilson, T. and Turner, D.H. (1983) Effects of 3' Dangling End Stacking on the Stability of Ggcc and Ccgg Double Helices. *Biochemistry*, **22**, 6198-6206.
36. Freier, S.M., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985) Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry*, **24**, 4533-4539.
37. 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.
38. 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.
39. Freier, S.M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M.H. and Turner, D.H. (1986) Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry*, **25**, 3214-3219.

40. Sugimoto, N., Kierzek, R., Freier, S.M. and Turner, D.H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helixes. *Biochemistry*, **25**, 5755-5759.
41. 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.
42. 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.
43. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) Sequence dependence for the energetics of dangling ends and terminal base pairs in ribonucleic acid. *Biochemistry*, **26**, 4554-4558.
44. Sugimoto, N., Kierzek, R. and Turner, D.H. (1987) Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, **26**, 4559-4562.
45. Longfellow, C.E., Kierzek, R. and Turner, D.H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry*, **29**, 278-285.
46. 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.
47. Hall, K.B. and McLaughlin, L.W. (1991) Thermodynamic and structural properties of pentamer DNA.DNA, RNA.RNA, and DNA.RNA duplexes of identical sequence. *Biochemistry*, **30**, 10606-10613.
48. He, L., Kierzek, R., SantaLucia, J., Jr., Walter, A.E. and Turner, D.H. (1991) Nearest-neighbor parameters for G.U mismatches: 5'GU3'/3'UG5' is destabilizing in the contexts CGUG/GUGC, UGUA/AUGU, and AGUU/UUGA but stabilizing in GGUC/CUGG. *Biochemistry*, **30**, 11124-11132.
49. Walter, A.E., Wu, M. and Turner, D.H. (1994) The stability and structure of tandem GA mismatches in RNA depend on closing base pairs. *Biochemistry*, **33**, 11349-11354.
50. Wu, M., McDowell, J.A. and Turner, D.H. (1995) A periodic table of symmetric tandem mismatches in RNA. *Biochemistry*, **34**, 3204-3211.
51. 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.
52. McDowell, J.A., He, L., Chen, X. and Turner, D.H. (1997) Investigation of the structural basis for thermodynamic stabilities of tandem GU wobble pairs: NMR structures of (rGGAGUUCC)2 and (rGGAUGUCC)2. *Biochemistry*, **36**, 8030-8038.
53. Xia, T., 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.
54. Kierzek, R., Burkard, M.E. and Turner, D.H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, **38**, 14214-14223.
55. Burkard, M.E., Xia, T. and Turner, D.H. (2001) Thermodynamics of RNA internal loops with a guanosine-guanosine pair adjacent to another noncanonical pair. *Biochemistry*, **40**, 2478-2483.
56. Schroeder, S.J. and Turner, D.H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA. *Biochemistry*, **40**, 11509-11517.
57. Serra, M.J., Baird, J.D., Dale, T., Fey, B.L., Retatagos, K. and Westhof, E. (2002) Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA*, **8**, 307-323.
58. 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.
59. Serra, M.J., Smolter, P.E. and Westhof, E. (2004) Pronounced instability of tandem IU base pairs in RNA. *Nucleic Acids Res*, **32**, 1824-1828.

60. Chen, G., Znosko, B.M., Jiao, X. and Turner, D.H. (2004) Factors affecting thermodynamic stabilities of RNA 3 x 3 internal loops. *Biochemistry*, **43**, 12865-12876.
61. Clanton-Arrowood, K., McGurk, J. and Schroeder, S.J. (2008) 3' terminal nucleotides determine thermodynamic stabilities of mismatches at the ends of RNA helices. *Biochemistry*, **47**, 13418-13427.
62. Uhlenbeck, O.C., Martin, F.H. and Doty, P. (1971) Self-complementary oligoribonucleotides: effects of helix defects and guanylic acid-cytidylic acid base pairs. *J Mol Biol*, **57**, 217-229.
63. Martin, F.H., Uhlenbeck, O.C. and Doty, P. (1971) Self-complementary oligoribonucleotides: adenylic acid-uridylic acid block copolymers. *J Mol Biol*, **57**, 201-215.
64. Manning, G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys*, **11**, 179-246.
65. Turner, D.H. (2013) Fundamental interactions in RNA: Questions answered and remaining. *Biopolymers*, **99**, 1097-1104.
66. Draper, D.E. (2004) A guide to ions and RNA structure. *RNA*, **10**, 335-343.
67. Bai, Y., Greenfeld, M., Travers, K.J., Chu, V.B., Lipfert, J., Doniach, S. and Herschlag, D. (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J Am Chem Soc*, **129**, 14981-14988.
68. Lipfert, J., Doniach, S., Das, R. and Herschlag, D. (2014) Understanding nucleic acid-ion interactions. *Annu Rev Biochem*, **83**, 813-841.
69. Kirmizialtin, S., Pabit, S.A., Meisburger, S.P., Pollack, L. and Elber, R. (2012) RNA and its ionic cloud: solution scattering experiments and atomically detailed simulations. *Biophys J*, **102**, 819-828.
70. Jiang, T., Kennedy, S.D., Moss, W.N., Kierzek, E. and Turner, D.H. (2014) Secondary structure of a conserved domain in an intron of influenza A M1 mRNA. *Biochemistry*, **53**, 5236-5248.
71. Banerjee, A.R., Jaeger, J.A. and Turner, D.H. (1993) Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure. *Biochemistry*, **32**, 153-163.
72. R Core Team. (2020). R Foundation for Statistical Computing, Vienna, Austria. .
73. Wei, T.S., Viliam. (2017), *R package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from https://github.com/taiyun/corrplot*.
74. Cantor, C.R. and Schimmel, P. (1980) *Biophysical chemistry. Part III, The behavior of biological macromolecules*. W.H. Freeman and Company, San Francisco.
75. Zuber, J., Cabral, B., McFadyen, I., Mauger, D.M. and Mathews, D.H. (2018) Analysis of RNA Nearest Neighbor Parameters Reveals Interdependencies and Quantifies the Uncertainty in RNA Secondary Structure Prediction. *RNA*, **24**, 1568-1582.
76. Zuber, J. and Mathews, D.H. (2019) Estimating uncertainty in predicted folding free energy changes of RNA secondary structures. *RNA*, **25**, 747-754.
77. Venables, W.R., BD. (2002) *Modern Applied Statistics with S*. Fourth ed. Springer, New York, NY.
78. Schroeder, S.J. and Turner, D.H. (2009) Optical melting measurements of nucleic acid thermodynamics. *Methods Enzymol*, **468**, 371-387.
79. Bellaousov, S. and Mathews, D.H. (2010) ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. *RNA*, **16**, 1870-1880.
80. Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*, **9**, 133-148.
81. Nussinov, R. and Jacobson, A.B. (1980) Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci U S A*, **77**, 6309-6313.
82. Zuker, M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48-52.

83. Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
84. Sato, K., Akiyama, M. and Sakakibara, Y. (2021) RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat Commun*, **12**, 941.
85. Shi, H., Rangadurai, A., Abou Assi, H., Roy, R., Case, D.A., Herschlag, D., Yesselman, J.D. and Al-Hashimi, H.M. (2020) Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. *Nat Commun*, **11**, 5531.
86. Reuter, J.S. and Mathews, D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *Bmc Bioinformatics*, **11**, 129.
87. Ding, Y. and Lawrence, C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res*, **31**, 7280-7301.
88. Mauger, D.M., Cabral, B.J., Presnyak, V., Su, S.V., Reid, D.W., Goodman, B., Link, K., Khatwani, N., Reynders, J., Moore, M.J. *et al.* (2019) mRNA structure regulates protein expression through changes in functional half-life. *Proc Natl Acad Sci U S A*, **116**, 24075-24083.
89. Wayment-Steele, H.K., Kim, D.S., Choe, C.A., Nicol, J.J., Wellington-Oguri, R., Watkins, A.M., Parra Sperberg, R.A., Huang, P.S., Participants, E. and Das, R. (2021) Theoretical basis for stabilizing messenger RNA through secondary structure design. *Nucleic Acids Res*, **49**, 10604-10617.
90. Ritchey, L.E., Tack, D.C., Yakhnin, H., Jolley, E.A., Assmann, S.M., Bevilacqua, P.C. and Babitzke, P. (2020) Structure-seq2 probing of RNA structure upon amino acid starvation reveals both known and novel RNA switches in Bacillus subtilis. *RNA*, **26**, 1431-1447.
91. Su, Z., Tang, Y., Ritchey, L.E., Tack, D.C., Zhu, M., Bevilacqua, P.C. and Assmann, S.M. (2018) Genome-wide RNA structurome reprogramming by acute heat shock globally regulates mRNA abundance. *Proc Natl Acad Sci U S A*, **115**, 12170-12175.
92. Tack, D.C., Su, Z., Yu, Y., Bevilacqua, P.C. and Assmann, S.M. (2020) Tissue-specific changes in the RNA structurome mediate salinity response in Arabidopsis. *RNA*, **26**, 492-511.
93. Hoshika, S., Leal, N.A., Kim, M.J., Kim, M.S., Karalkar, N.B., Kim, H.J., Bates, A.M., Watkins, N.E., Jr., SantaLucia, H.A., Meyer, A.J. *et al.* (2019) Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science*, **363**, 884-887.
94. Kierzek, E., Zhang, X., Watson, R.M., Kierzek, R. and Mathews, D.H. (2021) Secondary Structure Prediction for RNA Sequences Including N^6-methyladenosine. *bioRxiv*, 2021.2004.2026.441443.
95. Wright, D.J., Force, C.R. and Znosko, B.M. (2018) Stability of RNA duplexes containing inosine.cytosine pairs. *Nucleic Acids Res*, **46**, 12099-12108.
96. Lu, Z.J. and Mathews, D.H. (2008) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Research*, **36**, 640-647.
97. Lu, Z.J. and Mathews, D.H. (2008) Fundamental differences in the equilibrium considerations for siRNA and antisense oligodeoxynucleotide design. *Nucleic Acids Research*, **36**, 3738-3745.
98. Shao, Y., Chan, C.Y., Maliyekkel, A., Lawrence, C.E., Roninson, I.B. and Ding, Y. (2007) Effect of target secondary structure on RNAi efficiency. *RNA*, **13**, 1631-1640.
99. Tafer, H., Ameres, S.L., Obernosterer, G., Gebeshuber, C.A., Schroeder, R., Martinez, J. and Hofacker, I.L. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nature Biotechnology*, **26**, 578-583.
100. Szabat, M., Lorent, D., Czapik, T., Tomaszewska, M., Kierzek, E. and Kierzek, R. (2020) RNA Secondary Structure as a First Step for Rational Design of the Oligonucleotides towards Inhibition of Influenza A Virus Replication. *Pathogens*, **9**.
101. Angelbello, A.J., Rzuczek, S.G., McKee, K.K., Chen, J.L., Olafson, H., Cameron, M.D., Moss, W.N., Wang, E.T. and Disney, M.D. (2019) Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc Natl Acad Sci U S A*, **116**, 7799-7804.

102. Costales, M.G., Aikawa, H., Li, Y., Childs-Disney, J.L., Abegg, D., Hoch, D.G., Pradeep Velagapudi, S., Nakai, Y., Khan, T., Wang, K.W. *et al.* (2020) Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer. *Proc Natl Acad Sci U S A*, **117**, 2406-2411.
103. Suresh, B.M., Li, W., Zhang, P., Wang, K.W., Yildirim, I., Parker, C.G. and Disney, M.D. (2020) A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc Natl Acad Sci U S A*, **117**, 33197-33203.
104. Tinoco, I., Jr., Borer, P.N., Dengler, B., Levin, M.D., Uhlenbeck, O.C., Crothers, D.M. and Bralla, J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nat New Biol*, **246**, 40-41.
105. Borer, P.N., Dengler, B., Tinoco, I., Jr. and Uhlenbeck, O.C. (1974) Stability of ribonucleic acid double-stranded helices. *J Mol Biol*, **86**, 843-853.
106. Gralla, J. and Crothers, D.M. (1973) Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *J Mol Biol*, **73**, 497-511.
107. Breslauer, K.J., Frank, R., Blöcker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences*, **83**, 3746.
108. England, T.E. and Uhlenbeck, O.C. (1978) Enzymic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry*, **17**, 2069-2076.
109. Scaringe, S.A., Wincott, F.E. and Caruthers, M.H. (1998) Novel RNA Synthesis Method Using 5‘-O-Silyl-2‘-O-orthoester Protecting Groups. *J Am Chem Soc*, **120**, 11820-11821.
110. Gray, D.M. (1997) Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest neighbors. *Biopolymers*, **42**, 783-793.
111. Gray, D.M. (1997) Derivation of nearest-neighbor properties from data on nucleic acid oligomers. II. Thermodynamic parameters of DNA.RNA hybrids and DNA duplexes. *Biopolymers*, **42**, 795-810.
112. Jaeger, J.A., Turner, D.H. and Zuker, M. (1989) Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci U S A*, **86**, 7706-7710.
113. Sieger, G., Hofmann, H., Förtsch, J., Gross, H.J., Randies, J.W., Sänger, H.L. and Riesner, D. (1984) Conformational Transitions in Viroids and Virusoids: Comparison of Results from Energy Minimization Algorithm and from Experimental Data. *Journal of Biomolecular Structure and Dynamics*, **2**, 543-571.
114. Huang, L., Zhang, H., Deng, D., Zhao, K., Liu, K., Hendrix, D.A. and Mathews, D.H. (2019) LinearFold: linear-time approximate RNA folding by 5'-to-3' dynamic programming and beam search. *Bioinformatics*, **35**, i295-i304.
115. Zhang, H., Zhang, L., Mathews, D.H. and Huang, L. (2020) LinearPartition: linear-time approximation of RNA folding partition function and base-pairing probabilities. *Bioinformatics*, **36**, i258-i267.
116. Chen, X., McDowell, J.A., Kierzek, R., Krugh, T.R. and Turner, D.H. (2000) Nuclear magnetic resonance spectroscopy and molecular modeling reveal that different hydrogen bonding patterns are possible for G.U pairs: one hydrogen bond for each G.U pair in r(GGCGUGCC)(2) and two for each G.U pair in r(GAGUGCUC)(2). *Biochemistry*, **39**, 8970-8982.
117. Tolbert, B.S., Kennedy, S.D., Schroeder, S.J., Krugh, T.R. and Turner, D.H. (2007) NMR structures of (rGCUGAGGCU)2 and (rGCGGAUGCU)2: probing the structural features that shape the thermodynamic stability of GA pairs. *Biochemistry*, **46**, 1511-1522.
118. Liu, J.D., Zhao, L. and Xia, T. (2008) The dynamic structural basis of differential enhancement of conformational stability by 5'- and 3'-dangling ends in RNA. *Biochemistry*, **47**, 5962-5975.

119. Nikolova, E.N. and Al-Hashimi, H.M. (2010) Thermodynamics of RNA melting, one base pair at a time. *RNA*, **16**, 1687-1691.
120. Dethoff, E.A., Petzold, K., Chugh, J., Casiano-Negroni, A. and Al-Hashimi, H.M. (2012) Visualizing transient low-populated structures of RNA. *Nature*, **491**, 724-728.
121. Rangadurai, A., Szymanski, E.S., Kimsey, I., Shi, H. and Al-Hashimi, H.M. (2020) Probing conformational transitions towards mutagenic Watson-Crick-like G.T mismatches using off-resonance sugar carbon R1rho relaxation dispersion. *J Biomol NMR*, **74**, 457-471.
122. Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F. and Miska, E.A. (2020) The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol Cell*, **80**, 1067-1077 e1065.
123. Li, S., Zhang, H., Zhang, L., Liu, K., Liu, B., Mathews, D.H. and Huang, L. (2021) LinearTurboFold: Linear-Time Global Prediction of Conserved Structures for RNA Homologs with Applications to SARS-CoV-2. *bioRxiv*, 2020.2011.2023.393488.
124. Correll, C.C. and Swinger, K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA*, **9**, 355-363.