Nearest neighbor rules for RNA helix folding thermodynamics: improved end effects

Jeffrey Zuber[1,†], Susan J. Schroeder[2,†], Hongying Sun[3,4], Douglas H. Turner[4,5] and 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

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 5′GGUC3′ paired to 3′CUGG5′ separately from other 5′GU3′/3′UG5′ 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 <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 are effective 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, $\Delta G°$, often augmented with information from chemical mapping and/or sequence comparison. Usually, about half the nucleotides in structured non-coding RNA are canonically paired (17,18). GU pairs play important roles in RNA structure and function as sites for binding metal ions (19,20), therapeutics (21), proteins or metabolites (22).

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 (23). 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 (24,25). Individual parameters for nearest neighbors containing at least one GU pair, however, were revised on the basis of new data (25). In that revision, the penalty of 0.45 kcal/mol applied to terminal AU pairs and previously assumed for terminal GU pairs (24), was found unnecessary for GU pairs. Expansion of the database for duplexes, particularly those with terminal GU pairs (26) 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, i.e. sequence-specific terms for the ends of helices that account for the last and penultimate base pairs. 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 $\frac{5'GU}{3'UG}$ and $\frac{5'AG}{3'UU}$ nearest neighbors, where stacks are shown for a top strand in the 5' to 3' direction pairing to a bottom strand in the opposite direction. With these changes, $\frac{5'GGUC}{3'CUGG}$ fits the NN model rather than being an outlier as considered previously (27). 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 (28). 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) (22). In Figure 1A, the terminal GU pair in the foreground is in a $\frac{5'UG}{3'GU}$ nearest neighbor and has a single hydrogen bond while the penultimate UG pair has two hydrogen bonds. The conformation of the terminal GU pair may be influenced by solvent interactions or crystal contacts through stacking interactions with the terminal GU pair of an adjacent molecule. In Figure 1B, the $\frac{5'UG}{3'GU}$ nearest neighbor is flanked by WCF GC pairs on both sides in the middle of the helix, i.e. $\frac{5'\ldots GUGC\ldots}{3'\ldots CGUG\ldots}$ In that context, both the GU pairs 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 Figure 1B and C reflect the different nucleotide configurations and effect of considering terminal effects. The $\Delta G°_{37}$ of $\frac{5'UG}{3'GU}$ and $\frac{5'GU}{3'UG}$ 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 (25). The new parameters also include an increment of –0.74 kcal/mol for
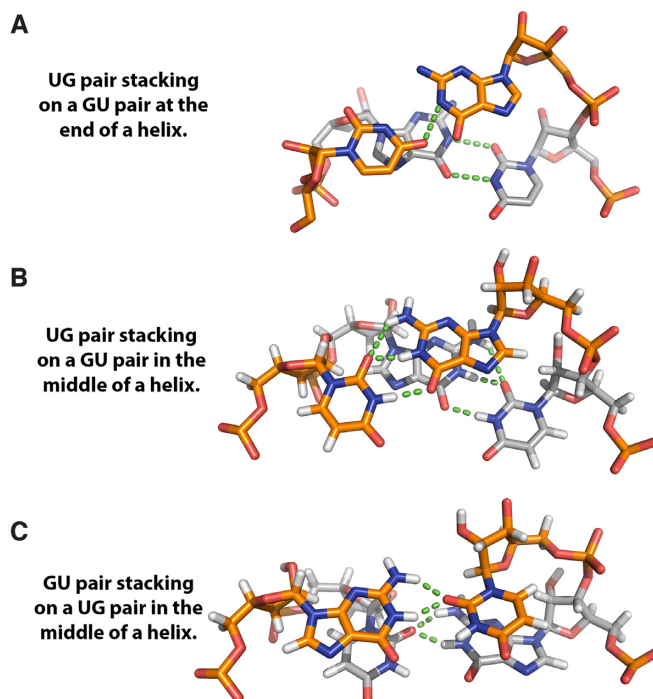

Figure 1. GU pair stacking and hydrogen bonding for three contexts of tandem GU pairs. Three distinct patterns of hydrogen bonding (shown in green) are observed in these three examples of cis Watson–Crick/Watson–Crick pairs by the Leontis and Westhof nomenclature (152). (**A**) X-ray crystal structure (1.4 Å resolution, $R_{\text{free}} = 20.7\%$) with terminal stacking GU pairs in 5'**UG**CUCCUAGUACGUAAGGACCGGAGU**G**, PDB ID# 1MSY (153). 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. Here, the GU pair, although *cis* Watson–Crick/Watson–Crick has a single hydrogen bond. (**B**) NMR structure with internal UG pairs in (5'GAGU**G**CUC)₂, PDB ID# 1EKA (137). 28 unique NOE measurements define the orientation for these nucleotides. (**C**) NMR structure with internal GU pairs in (5'GGC**G**UGCC)₂. PDB ID# 1EKD (137). 26 unique NOE measurements define the orientation for these nucleotides.

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 (26,29–33). 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 (25,26,29,34–62). Enumeration of all melting experiments included in

this analysis is available in Supplementary Table S1 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, duplexes that were reported by the original authors to have non-2-state unfolding transitions were excluded from this analysis (43). A total of 223 experiments were included in the fits, with 125 for Watson–Crick-Franklin pair parameters and 98 for GU pair parameters.

Most melting experiments for generating NN parameters have used 1 M Na$^+$. This was initially chosen to assure formation of duplexes rather than hairpins, as also seen for deoxy A–T oligonucleotides (63) and to allow measurements of the concentration dependence of duplexes that could only be synthesized with many AU pairs (64,65). The high melting temperatures dependent on 1 M NaCl became more important when comparisons with calorimetry revealed that interpretation of optical melting improved if sloping upper and lower baselines were considered (66).

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 (67–72). 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 (73). For A-form double helical and single strand RNA, respectively, the local 'ion atmosphere' in the absence of multiple charged cations is predicted to have 1.7 M and 0.4 M of M$^+$ ions (74). They respectively neutralize 0.8 and 0.6 of the phosphate charge. In the absence of M$^+$ cations, M^{2+} cations are predicted to neutralize 0.9 and 0.8, respectively, of backbone charge. More detailed computations and experiments agree qualitatively with expectations from Manning theory (29,52,58,67,68,70,71,75–77). GU pairs can be the sites of metal ion binding (20,58,78,79), but optical melting experiments of duplexes with consecutive GU pairs did not find differences in stabilities in 1 M Na$^+$ and in 150 mM K$^+$ with 10 mM Mg^{2+} (26). Together these suggest that, while salt conditions vary between and within cells (80–84), they are unlikely to affect dramatically the relative stabilities of NN canonical pairs.

Feature correlations

Feature correlations were calculated for each model using the R statistical programming language. The resulting correlation matrices were then visualized with the *R corrplot* library, available at https://github.com/taiyun/corrplot.

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} (24,85). 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*. 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 (86,87). To perform covariation analysis, the optical melting data were resampled within experimental error ($\Delta H^\circ_\sigma = 12\% \Delta H^\circ$ and $\Delta S^\circ_\sigma = 13.5\% \Delta S^\circ$ (24)). The resampling was performed with the *mvrnorm* function from the R MASS library (88), which preserves the observed correlation between ΔH° and ΔS° ($\rho = 0.9996$ (24)). 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 (86,87). The standard errors of regression, which neglect the correlations and the effect of nested regressions, for the NN parameters can be found in Supplementary Table S2.

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 to test the revised model

Optical melting experiments were conducted on three additional duplexes, (5′UGUCGAUA)$_2$, (5′AUAGCUGU)$_2$ and (5′AUUCGAGU)$_2$, following standard protocols described in (89). 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$_2$EDTA. 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 280 nm. Data was analyzed with Meltwin software (52).

Stacking term counts

An archive of RNA sequences of known secondary structure (18,90) was analyzed to count the number of occurrences of each NN stacking parameter. A Python script was used to parse each structure into individual helices and then to parse each helix into component NN stacking and helix end parameters.

Table 1A. The $\Delta G°_{37}$ and for $\Delta H°$ nearest neighbor parameters for helices composed of WCF pairs. The base pair stacks are represented with the top strand (5′ to 3′), the slash, and then the bottom strand (3′ to 5′). For example, AC/UG is the stack $\frac{5'AC}{3'UG}$. The end terms are added in addition to the stacks, with example calculations in Figure 4

Feature	New Model[§]			1998 Model [¶]	
	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$(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 $\Delta G°_{37}$ and for $\Delta H°$ nearest neighbor parameters for stacks with GU pairs

Feature	New Model[§]			2012 Model [*]	
	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$(eu)	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$(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 Supplementary Table S2.
[¶]Parameters taken from (24).
[‡]The 1998 model did not have separate values for each AU End variant.
[*]Values taken from (25).
[†]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.

RESULTS

AU end parameters depend on penultimate pair

Prior work demonstrated that multiple GU terminal base pairs impact the stability of helical duplexes (26), 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 $\frac{5'\dots AU}{3'\dots UA}$, $\frac{5'\dots AA}{3'\dots UU}$, $\frac{5'\dots UU}{3'\dots AA}$, or $\frac{5'\dots AU}{3'\dots UA}$ helix ends) and a parameter for an AU termi-

nal pair on a CG penultimate pair (applying to $\frac{5'\dots GU}{3'\dots CA}$, $\frac{5'\dots GA}{3'\dots CU}$, $\frac{5'\dots CU}{3'\dots GA}$, or $\frac{5'\dots CA}{3'\dots GU}$ helix ends). In these terminal stacks, the end of the helix is to the right, a top strand is shown from 5′ to 3′, and pairs are shown to a bottom strand running in the opposite direction. These parameters are applied in addition to the base pair stacking parameter for these end stacks. This model for terminal AU pairs considers the identity but not the orientation of pairs and is consistent with experiments on terminal GU pairs (26). Those experiments revealed that a bonus for multiple terminal GU pairs is largely independent of orientation. Those observations inspired this model.
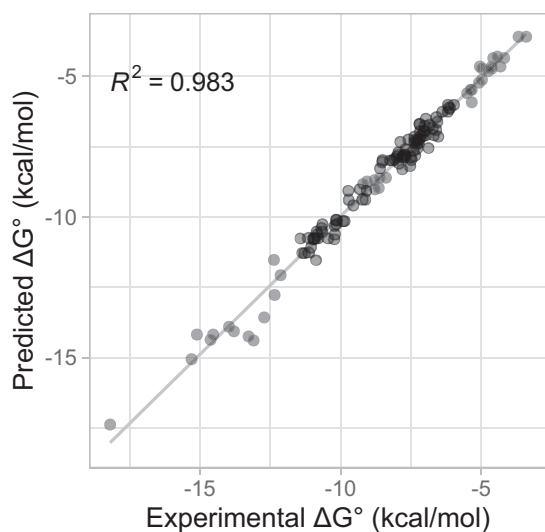
Figure 2. Correlation between predicted and measured $\Delta G°_{37}$ for duplexes with only WCF pairs. $\Delta G°_{37}$ values predicted from updated nearest neighbor parameters for duplexes composed solely of WCF base pairs (Table 1A) are plotted against values determined from optical melting experiments.

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

Comparisons between the updated model and those used in the 1998 and 2004 NN models are in Table 1A. 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 $\Delta G°_{37}$ and $\Delta H°$. 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 $\Delta G°_{37}$ ($R^2 = 0.9830$, Figure 2) and $\Delta H°$ ($R^2 = 0.8877$, Supplemental Figure S1). 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 S2). The predicted folding $\Delta G°_{37}$ were within 0.5 kcal/mol of the measured value for 86.4% of the experiments (Supplemental Figure S3). Predicted $\Delta H°$ are within 5 kcal/mol of the measured value for 76% of the experiments (Supplemental Figure S3).

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 $\Delta G°_{37}$ and $\Delta H°$ parameter values for these leave-one-out (LOO) data sets can be seen in Supplemental Figure S4. No one individual experiment heavily impacted the parameter values. The biggest impacts were RMSDs of 0.0363 kcal/mol in $\Delta G°_{37}$ and 0.3346 kcal/mol in $\Delta H°$, substantially smaller than uncertainty in the parameter values.

GU stacking parameters

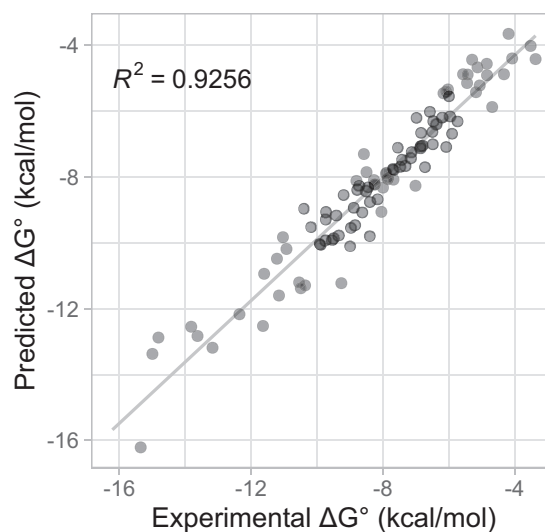A similar model for terminal AU and GU stacks was used when fitting duplexes with GU base pairs. The model re-

quires 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 $\frac{5'GGUC}{3'CUGG}$ as a special, non-nearest neighbor case. When results from a fitting model including a parameter for the non-nearest neighbor quadruplet $\frac{5'GGUC}{3'CUGG}$, 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 $\frac{5'GGUC}{3'CUGG}$, predicted $\Delta G°_{37}$ and $\Delta H°$ values from each model were also close to each other and to predicted $\Delta G°_{37}$ and $\Delta H°$ values from the 2012 model (25) (Supplementary Table S3). Additionally, the inclusion of the special $\frac{5'GGUC}{3'CUGG}$ parameter results in almost identical R^2 value for the fit (0.9256 versus 0.9267) (Figure 3 and Supplemental Figure S5). 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 $\frac{5'GU}{3'UG}$ stack, where the $\Delta G°_{37}$ changed from +0.72 kcal/mol to –0.19 kcal/mol between the previous (25) 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 $\frac{5'\ldots AG}{3'\ldots UU}$ stack, which went from having a $\Delta G°_{37}$ contribution of –0.35 kcal/mol

(25) 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 $\Delta G°_{37}$ ($R^2 = 0.9256$, Figure 3) and $\Delta H°$ ($R^2 = 0.7659$, Supplemental Figure S6). 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 S7). For folding $\Delta G°_{37}$, 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 S8). For $\Delta H°$, 57.1% of experiments had predictions within 5 kcal/mol of the measured value (79.6% were within 10 kcal/mol) (Supplemental Figure S8).

RMSDs in $\Delta G°_{37}$ and $\Delta H°$ parameter values for the LOO data sets can be seen in Supplemental Figure S9. 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 $\Delta G°_{37}$ and 0.8675 kcal/mol in $\Delta H°$, 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. We previously found that this approach is important for estimating the uncertainties of folding free energies and enthalpies because the correlated nature of the NN parameters and the use of sequential regressions break the assumptions used in the calculations of standard errors of regression (87). Covariances between parameters are presented in Supplemental Figures S10 and S11. 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 (24) models use values of –0.71 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 –0.02 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 mea-

sured thermodynamic values for these three duplexes. The new model predicts that on average the duplexes are 1.60 kcal/mol more stable at 37°C than predicted by the prior model. This equates to an average 13-fold more favorable equilibrium constant for duplex formation. 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 (14,91–98). These RNA structure prediction programs enable design of mRNA vaccine sequences (13,99,100), analysis of metaproperties of transcriptomic changes in response to stress (101–103), determination of effects of nucleotide modifications on folding stability (104–106), discovery of accessible regions to target with antisense DNA or siRNA (107–110), and rational design of small molecules targeting RNA (111–114). 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. Supplemental Figure S12 shows the improvement in the residuals for the new model compared to the previous.

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 $\frac{5'GGUC}{3'CUGG}$ 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.

This work presents the next advance in development of a robust NN model for predicting RNA duplex stabilities. The NN model for $\Delta G°$ of RNA helices composed of canonical pairs uses stacks of adjacent base pairs (115–117). This assumes that the total $\Delta G°$ and temperature dependence, $\Delta H°$, for helix formation can be approximated by summing $\Delta G°$s and $\Delta H°$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 (64,65). 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 developed for polynucleotides was tested and found to fit the database (118). This suggested that the NN model would allow predictions for unmeasured sequences (116). In collaboration with related efforts in the Crothers lab, this led to original rules for predicting the thermodynamics of RNA folding (115).

Table 2. Optical melting experimental tests of nearest neighbor parameters. Melting buffer was 1 M NaCl, 20 mM sodium cacodylate pH 7, and 0.5 mM Na$_2$EDTA. *a*) Predicted $\Delta G°_{37}$ are calculated using Table 1 values for the new model parameters. *b*) Predicted $\Delta G°_{37}$ are calculated from the previous models described in (25) and (24). The duplex values in italics show borderline apparent 2-state behavior with a difference in $\Delta H°$ between the two analyses of 18.2%, rather than the typical <15% criterion

Duplex	New model[a] $-\Delta G°_{37}$ (kcal/mol)	Prior model[b] $-\Delta G°_{37}$ (kcal/mol)	Van't Hoff plot analysis				Curve fit analysis			
			$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_m (°C)†	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_m (°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*

$^\dagger T_m$ values were calculated for an RNA strand concentration of 1×10^{-4} M.

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 (43,85,119). Application of T4 RNA ligase and development of chemical synthesis on polymer supports allowed expansion of sequences available (42,43,120,121). 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 (122,123) led to discovery that duplexes with the same nearest neighbors can have different thermodynamics depending on the terminal base pair (24).

Applications of the method were expanded to larger RNAs by modifying dynamic programming algorithms to predict folding that optimized $\Delta G°$ (91,124,125) rather than base pairing (92). The thermodynamic approach lends itself to modeling ensembles of structures, including calculations of base pairing probabilities and stochastic samples (98,126,127). Additional applications include predicting structures for multiple interacting strands (128), designing sequences to fold to specific structures (129,130), and integrating mapping or conservation data into structure prediction (131–134). Recently, applications to even larger RNAs have become possible due to linearization of the algorithms (135,136).

NN parameters for canonical pairs have undergone substantial revisions over time, including treatment of end effects (24,25,27,43,116). GU terminal base pairs were initially assumed to be equivalent to AU terminal base pairs and were given the same penalty term (27). Expansion of the database and refitting of the model indicated that GU terminal base pairs do not require an end penalty (25). Measurements on helices with consecutive terminal GU pairs, however, revealed they are surprisingly more stable than predicted (26). 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 (22). 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 $\frac{5'GU}{3'UG}$

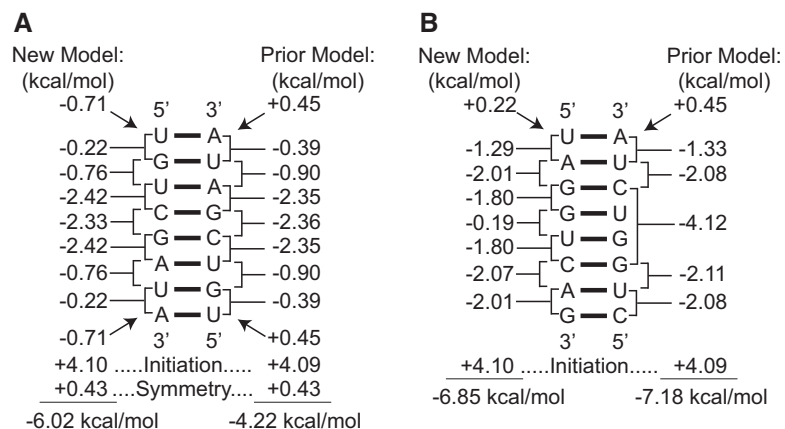**Figure 4.** Example calculations of helical ΔG°_{37}. Panel (**A**) shows the stability calculation for (5′UGUCGAUA)$_2$, which is shown in Table 2 to have an experimentally determined ΔG°_{37} of –6.10 kcal/mol. This sequence is self-complementary and therefore the symmetry penalty is added. Panel (**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 (24,25). The total stability is the sum of the stability increments.

had one NN parameter value with an exception for the motif $\frac{5'\text{GGUC}}{3'\text{CUGG}}$, 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 (52,53,137,138). In addition, the crystal structure of consecutive terminal GU pairs in (5′GGUGGCUGUU3′)$_2$ 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 (30). NMR studies of duplexes with consecutive terminal GU pairs usually showed broad resonances and few or weak NOEs in the final two GU pairs (26,30). Fluorescence and NMR studies have quantified different base pair dynamics in the middle and ends of helices for various types of base pairs (139–142). 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 of +0.22 and +0.44 kcal/mol at 37°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 (24). 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 (91,92). In prior models, the special cases for GU pairs required additional considerations in dynamic programming algorithm computations. In current application of the nearest neighbor parameters (143), 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 in accepted RNA secondary structures at a rate of approximately 13 per 1000 bases in the sequences (Supplemental Figure S13). 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 (21) and the motif $\frac{5'\text{UUUU}}{3'\text{GAGG}}$ 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 (144) and computationally (145). One end of SL3 terminates in a $\frac{5'\text{UG}}{3'\text{GU}}$ 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 RNA 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 struc-

ture 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.

While the parameters in Tables 1A and 1B provide excellent predictions of the measured duplex stabilities at 37°C, there is slightly less agreement for duplexes with WCF and GU pairs (Figures 2 and 3). This is not surprising because the database with GU pairs is smaller than that with only WCF pairs. Additionally, GU pairs are more likely to have different structures (Figure 1) (22). One example of this is (GGC$\underline{GU}$GCC)$_2$, where measured and predicted values for $\Delta G°_{37}$ are, respectively, –9.72 and –11.24 kcal/mol. On the basis of NMR and nucleobase substitution results (137), and mesoscopic modeling based on melting temperatures (146), the GU pairs in (GGC$\underline{GU}$GCC)$_2$ were determined to have only one H-bond each (Figure 1C) rather than the usual two. That may explain the overprediction of thermodynamic stability.

There are known limitations to the current parameterization of the nearest neighbor rules. These parameters depend on the two-state fits of melting data and assume that the enthalpy and entropy changes are temperature independent. It has been shown that parameters can be determined without assuming two-state melting by fitting directly to the optical melting data (absorbance as a function of temperature) (147), although much of these data are not currently available for the duplexes studied here. It is also known that enthalpy change and entropy change both depend on temperature (148–151). The changes have antagonizing effects with respect to free energy change, however, so folding free energy estimates at 37°C are probably little affected by the temperature dependencies because most strands are designed to have melting temperatures close to 37°C (149). On the other hand, extrapolation of folding free energies to other temperatures and estimates of melting temperatures are likely to be affected by these temperature dependencies. Future work could develop new nearest neighbor parameters that do not rely on these assumptions.

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 $\Delta G°$ and $\Delta H°$ 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.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

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**, a003558.
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: Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds). *The RNA World (3rd edn.)*. 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.
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–D129.
18. Bellaousov,S. and Mathews,D.H. (2010) ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA*, **16**, 1870–1880.
19. 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.
20. 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.
21. 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.

22. Westhof,E., Yusupov,M. and Yusupova,G. (2019) The multiple flavors of GoU pairs in RNA. *J. Mol. Recognit.*, **32**, e2782.

23. 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.

24. 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.

25. 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.

26. Nguyen,M.T. and Schroeder,S.J. (2010) Consecutive terminal GU pairs stabilize RNA helices. *Biochemistry*, **49**, 10574–10581.

27. 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.

28. 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.

29. Schroeder,S.J. and Turner,D.H. (2000) Factors affecting the thermodynamic stability of small asymmetric internal loops in RNA. *Biochemistry*, **39**, 9257–9274.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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 helices. *Biochemistry*, **24**, 4533–4539.

38. 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.

39. 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.

40. 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.

41. Sugimoto,N., Kierzek,R., Freier,S.M. and Turner,D.H. (1986) Energetics of internal GU mismatches in riboologonucleotide helixes. *Biochemistry*, **25**, 5755–5759.

42. 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.

43. 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.

44. 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.

45. Sugimoto,N., Kierzek,R. and Turner,D.H. (1987) Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, **26**, 4559–4562.

46. Longfellow,C.E., Kierzek,R. and Turner,D.H. (1990) Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry*, **29**, 278–285.

47. 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.

48. 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.

49. 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.

50. 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.

51. Wu,M., McDowell,J.A. and Turner,D.H. (1995) A periodic table of symmetric tandem mismatches in RNA. *Biochemistry*, **34**, 3204–3211.

52. 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.

53. 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.

54. 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.

55. Kierzek,R., Burkard,M.E. and Turner,D.H. (1999) Thermodynamics of single mismatches in RNA duplexes. *Biochemistry*, **38**, 14214–14223.

56. 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.

57. Schroeder,S.J. and Turner,D.H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA. *Biochemistry*, **40**, 11509–11517.

58. 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.

59. 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.

60. 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.

61. 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.

62. 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.

63. Scheffler,I.E., Elson,E.L. and Baldwin,R.L. (1970) Helix formation by d(TA) oligomers. II. Analysis of the helix-coli transitions of linear and circular oligomers. *J. Mol. Biol.*, **48**, 145–171.

64. 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.

65. 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.

66. Breslauer,K.J., Sturtevant,J.M. and Tinoco,I. Jr (1975) Calorimetric and spectroscopic investigation of the helix-to-coil transition of a ribo-oligonucleotide: rA7U7. *J. Mol. Biol.*, **99**, 549–565.

67. Draper,D.E. (2004) A guide to ions and RNA structure. *RNA*, **10**, 335–343.

68. Draper,D.E. (2008) RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.*, **95**, 5489–5495.

69. Pabit,S.A., Qiu,X., Lamb,J.S., Li,L., Meisburger,S.P. and Pollack,L. (2009) Both helix topology and counterion distribution contribute to the more effective charge screening in dsRNA compared with dsDNA. *Nucleic Acids Res.*, **37**, 3887–3896.

70. 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.

71. Lipfert,J., Doniach,S., Das,R. and Herschlag,D. (2014) Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.*, **83**, 813–841.

72. Gebala,M. and Herschlag,D. (2019) Quantitative studies of an RNA duplex electrostatics by ion counting. *Biophys. J.*, **117**, 1116–1124.

73. Manning,G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.*, **11**, 179–246.

74. Turner,D.H. (2013) Fundamental interactions in RNA: questions answered and remaining. *Biopolymers*, **99**, 1097–1104.

75. 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.

76. 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.

77. 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.

78. Gonzalez,R.L. Jr and Tinoco,I. Jr (1999) Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *J. Mol. Biol.*, **289**, 1267–1282.

79. Wang,W., Zhao,J., Han,Q., Wang,G., Yang,G., Shallop,A.J., Liu,J., Gaffney,B.L. and Jones,R.A. (2009) Modulation of RNA metal binding by flanking bases: ^{15}N NMR evaluation of GC, tandem GU, and tandem GA sites. *Nucleosides. Nucleotides Nucleic Acids*, **28**, 424–434.

80. Leamy,K.A., Assmann,S.M., Mathews,D.H. and Bevilacqua,P.C. (2016) Bridging the gap between in vitro and in vivo RNA folding. *Q. Rev. Biophys.*, **49**, e10.

81. Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K. and Watson,J.D. (1994) In: *Molecular Biology of the Cell (3rd edn.)*. Garland Publishing, NY.

82. London,R.E. (1991) Methods for measurement of intracellular magnesium: NMR and fluorescence. *Annu. Rev. Physiol.*, **53**, 241–258.

83. Romani,A.M. (2007) Magnesium homeostasis in mammalian cells. *Front. Biosci.*, **12**, 308–331.

84. Lusk,J.E., Williams,R.J. and Kennedy,E.P. (1968) Magnesium and the growth of *Escherichia coli*. *J. Biol. Chem.*, **243**, 2618–2624.

85. Cantor,C.R. and Schimmel,P. (1980) In: *Biophysical Chemistry. Part III, The Behavior of Biological Macromolecules*. W.H. Freeman and Company, San Francisco.

86. 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.

87. Zuber,J. and Mathews,D.H. (2019) Estimating uncertainty in predicted folding free energy changes of RNA secondary structures. *RNA*, **25**, 747–754.

88. Venables,W.R. and Ripley,B.D. (2002) In: *Modern Applied Statistics with S*. 4th edn. Springer, NY.

89. Schroeder,S.J. and Turner,D.H. (2009) Optical melting measurements of nucleic acid thermodynamics. *Methods Enzymol.*, **468**, 371–387.

90. Sloma,M.F. and Mathews,D.H. (2016) Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures. *RNA*, **22**, 1808–1818.

91. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.

92. 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.

93. Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48–52.

94. 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.

95. Sato,K., Akiyama,M. and Sakakibara,Y. (2021) RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat. Commun.*, **12**, 941.

96. 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.

97. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinf.*, **11**, 129.

98. Ding,Y. and Lawrence,C.E. (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, **31**, 7280–7301.

99. 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.

100. 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.

101. 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.

102. 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.

103. 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.

104. 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.

105. Kierzek,E., Zhang,X., Watson,R.M., Kennedy,S.D., Szabat,M., Kierzek,R. and Mathews,D.H. (2021) Secondary structure prediction for RNA sequences including N^6-methyladenosine. *Nat. Commun.*, **13**, 1271.

106. 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.

107. Lu,Z.J. and Mathews,D.H. (2008) Efficient siRNA selection using hybridization thermodynamics. *Nucleic Acids Res.*, **36**, 640–647.

108. Lu,Z.J. and Mathews,D.H. (2008) Fundamental differences in the equilibrium considerations for siRNA and antisense oligodeoxynucleotide design. *Nucleic Acids Res.*, **36**, 3738–3745.

109. 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.

110. 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. *Nat. Biotechnol.*, **26**, 578–583.

111. 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**, 925.

112. 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.

113. 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.

114. 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.

115. Tinoco,I. Jr, Borer,P.N., Dengler,B., Levin,M.D., Uhlenbeck,O.C., Crothers,D.M. and Gralla,J. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nat. New Biol.*, **246**, 40–41.

116. 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.

117. Gralla,J. and Crothers,D.M. (1973) Free energy of imperfect nucleic acid helices. II. Small hairpin loops. *J. Mol. Biol.*, **73**, 497–511.

118. Gray,D.M. and Tinoco Jr.,I. (1970) A new approach to the study of sequence-dependent properties of polynucleotides. *Biopolymers*, **9**, 223–244.

119. Breslauer,K.J., Frank,R., Blöcker,H. and Marky,L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 3746.

120. England,T.E. and Uhlenbeck,O.C. (1978) Enzymic oligoribonucleotide synthesis with T4 RNA ligase. *Biochemistry*, **17**, 2069–2076.

121. 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.

122. 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.

123. 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.

124. 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.

125. 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. *J. Biomol. Struct. Dyn.*, **2**, 543–571.

126. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.

127. Mathews,D.H. (2006) Revolutions in RNA secondary structure prediction. *J. Mol. Biol.*, **359**, 526–532.

128. Dirks,R.M., Bois,J.S., Schaeffer,J.M., Winfree,E. and Pierce,N.A. (2007) Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.*, **49**, 65–88.

129. Zadeh,J.N., Wolfe,B.R. and Pierce,N.A. (2011) Nucleic acid sequence design via efficient ensemble defect optimization. *J. Comput. Chem.*, **32**, 439–452.

130. Garcia-Martin,J.A., Clote,P. and Dotu,I. (2013) RNAiFold: a web server for RNA inverse folding and molecular design. *Nucleic Acids Res.*, **41**, W465–W470.

131. Sloma,M.F. and Mathews,D.H. (2015) Improving RNA secondary structure prediction with structure mapping data. *Methods Enzymol.*, **553**, 91–114.

132. Seetin,M.G. and Mathews,D.H. (2012) RNA structure prediction: an overview of methods. *Methods Mol. Biol.*, **905**, 99–122.

133. Havgaard,J.H. and Gorodkin,J. (2014) RNA structural alignments, part I: Sankoff-based approaches for structural alignments. *Methods Mol. Biol.*, **1097**, 275–290.

134. Asai,K. and Hamada,M. (2014) RNA structural alignments, part II: non-Sankoff approaches for structural alignments. *Methods Mol. Biol.*, **1097**, 291–301.

135. 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.

136. 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.

137. 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.

138. Jang,S.B., Hung,L-W., Jeong,M.S., Holbrook,E.L., Chen,X., 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 base pairs. *Biophys. J.*, **90**, 4530–4537.

139. 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.

140. Nikolova,E.N. and Al-Hashimi,H.M. (2010) Thermodynamics of RNA melting, one base pair at a time. *RNA*, **16**, 1687–1691.

141. 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.

142. 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.

143. Turner,D.H. and Mathews,D.H. (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res.*, **38**, D280–D282.

144. 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.

145. 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. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2116269118.

146. Amarante,T.D. and Weber,G. (2016) Evaluating hydrogen bonds and base stacking of single, tandem and terminal GU mismatches in RNA with a mesoscopic model. *J. Chem. Inf. Model.*, **56**, 101–109.

147. Spasic,A., Berger,K.D., Chen,J.L., Seetin,M.G., Turner,D.H. and Mathews,D.H. (2018) Improving RNA nearest neighbor parameters for helices by going beyond the two-state model. *Nucleic Acids Res.*, **46**, 4883–4892.

148. Diamond,J.M., Turner,D.H. and Mathews,D.H. (2001) Thermodynamics of three-way multibranch loops in RNA. *Biochemistry*, **40**, 6971–6981.

149. Lu,Z.J., Turner,D.H. and Mathews,D.H. (2006) A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic. Acids. Res.*, **34**, 4912–4924.

150. Vaitiekunas,P., Crane-Robinson,C. and Privalov,P.L. (2015) The energetic basis of the DNA double helix: a combined microcalorimetric approach. *Nucleic Acids Res.*, **43**, 8577–8589.

151. Privalov,P.L. and Crane-Robinson,C. (2020) Forces maintaining the DNA double helix. *Eur. Biophys. J.*, **49**, 315–321.

152. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.

153. 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.