Published online 26 October 2018 *Nucleic Acids Research, 2018, Vol. 46, No. 22* **12099–12108**
doi: 10.1093/nar/gky907

Stability of RNA duplexes containing inosine·cytosine pairs

Daniel J. Wright, Christopher R. Force and Brent M. Znosko*

Department of Chemistry, Saint Louis University, St. Louis, MO 63103, USA

Received June 06, 2018; Revised August 30, 2018; Editorial Decision September 23, 2018; Accepted October 22, 2018

ABSTRACT

Inosine is found naturally in the anticodon loop of tRNA, is a product of adenosine deaminases that act on RNA, and can be used in oligonucleotide probes or to investigate the role of the exocyclic amino group of guanosine. Although the thermodynamics of I·U pairs in RNA have been systematically studied [Wright, D. J., Rice, J. L., Yanker, D. M., and Znosko, B. M. (2007) *Biochemistry 46*, 4625–4634], the thermodynamics of I·C pairs in RNA have not. Here, we have performed optical melting experiments on a series of RNA duplexes containing I·C pairs and compared their thermodynamics to the same duplexes containing A·C and G–C pairs. Nearest neighbor parameters for single I·C pairs adjacent to Watson-Crick pairs were derived. The derived nearest neighbor parameters are compared to those previously predicted blindly through a reweighting of energy-function collection with conformational ensemble sampling in Rosetta [Chou, F.-C., Kladwang, W., Kappel, K., and Das, R. (2016) *Proc. Natl. Acad. Sci. U.S.A. 113*, 8430–8435]. Scientists can use these nearest neighbor parameters to calculate the stability of ADAR products and to calculate the stability of an RNA duplex in which G-to-I substitution was used to determine the role of the exocyclic amino group of G.

INTRODUCTION

Inosine (I) is a purine nucleoside consisting of the nucleobase hypoxanthine. Because the structure of inosine is the same as guanosine but without the exocyclic amino group (Figure 1), it tends to behave as guanosine (1,2). Inosine serves a variety of functions in the cell and in the research laboratory.

Inosine occurs naturally in the anticodon loop of some tRNAs. It is usually found in the wobble position of the anticodon loop and can pair with A, C or U in the codon mRNA (1,3). It has also been found in the middle position of the anticodon loop where it pairs with A in the codon mRNA (2).

Adenosine-to-inosine editing is considered to be the predominant form of RNA editing in mammals (4–6). A-to-I editing has been found in the tRNA of prokaryotes and eukaryotes (7) and in the mRNA of mammalian brains (8). While many of these editing sites are found in the coding regions of gene transcripts, the vast majority of these modifications are in non-coding sequences such as 5′ untranslated regions (UTRs), 3′UTRs, and introns (4). This editing serves to increase proteomic and phenotypic diversity (9). The conversion of A-to-I results in the partial unwinding of dsRNA (10) and may be responsible for creating alternative splice sites. Two enzyme families are mainly responsible for these changes, adenosine deaminases that act on RNA (ADARs) and adenosine deaminases that act on tRNA (ADATs) (11). A-to-I editing has also been shown to occur in some fungi independent of ADAR enzymes (6). The overexpression/downregulation of ADAR enzymes and A-to-I editing has been implicated in human diseases such as cancer, amyotrophic lateral sclerosis (ALS) and Aicardi-Goutières syndrome (12,13).

ADAR enzymes can deaminate A resulting in the conversion of A-to-I within regions of double-stranded RNA and the creation of I·U pairs. Importantly, when an ADAR enzyme encounters an RNA substrate with an A·C pair, it can convert the A to an I more efficiently than an A within an A–U pair (10,11), resulting in an I·C pair. It is hypothesized that the I and C residues shift slightly from the original A·C conformation in order to form two hydrogen bonds, resulting in a more stable conformation than the original A·C pair with one hydrogen bond (Figure 1).

Scientists have found uses for inosine *in vitro* and *in vivo*. Because of its promiscuity, inosine can be used in an oligonucleotide probe when the exact sequence of the nucleic acid target is unknown (1). Probes containing deoxyinosine have been used to screen high complexity genomic DNA and cDNA libraries (3,14). Similarly, degenerate hybridization probes containing inosine have been used to determine mRNA sequence, and degenerate primers containing inosine have been used for the amplification of ambiguous sequences using degenerate PCR (15).

Scientists also intentionally incorporate non-standard nucleotides such as inosine into RNA molecules to investigate the contribution of individual functional groups. Due

*To whom correspondence should be addressed. Tel: +1 314 977 8567; Fax: +1 314 977 2521; Email: brent.znosko@slu.edu

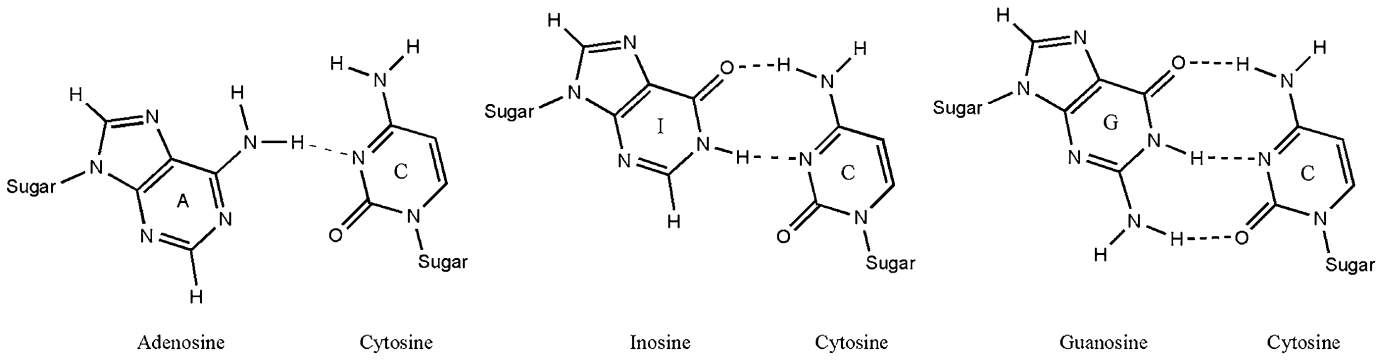

Figure 1. Likely hydrogen bond conformation of an A·C pair (left), I·C pair (center) and Watson–Crick G–C pair (right).

to its structural similarity to guanosine, inosine is a good probe for structure-specific minor groove interactions between G amino groups in RNA duplex regions and proteins (16). For example, an investigation of chaperones in RNA folding determined that some chaperones weaken intermolecular interactions involving G to achieve rapid, proper folding of the RNA. Substituting I in place of G in those RNAs negated the need for a chaperone to achieve rapid folding (17). Also, inosine is reported to have an oxidation potential around 200 mV higher than G (18,19). Therefore, G-to-I replacement has been used as a control experiment to study the dynamics of electron transport in DNA (18–20).

Knowledge of the thermodynamics of inosine pairing with A, C or U in RNA would be beneficial when comparing the stability of ADAR substrates to products, when designing probes containing inosine, and when intentionally replacing guanosine with inosine to study structure-function relationships. The thermodynamics of a set of deoxynucleotide duplexes containing deoxyinosine opposite the four standard DNA bases have been reported (1,3), and a comprehensive set of deoxyinosine nearest neighbor parameters have been derived (15).

Despite the widespread use of inosine and the extensive thermodynamic data for deoxyinosine, there have been relatively few thermodynamic studies with RNA containing inosine. The thermodynamics of I·U pairs have been systematically studied (21,22). Changing an A–U pair in the center of a Watson-Crick helix to an I·U pair destabilizes the duplex by an average of 2.3 kcal/mol. Interestingly, changing an A–U pair at the end of a Watson-Crick helix to an I·U pair has the opposite effect, stabilizing the duplex by an average of 0.8 kcal/mol (21).

Although the thermodynamics of I·U pairs in RNA have been systematically studied, the thermodynamics of I·C pairs in RNA have not. Base pairing between I and C is possible by forming two hydrogen bonds as in A–U pairs instead of the three that occur in G–C base pairs (Figure 1). This was confirmed in a crystal structure of the ribosomal decoding center (23). The structure shows that an I·C pair is the same as a canonical G–C base pair except for the missing N2–O2 hydrogen bond. Substitution of a G–C base pair with I·C has been found to influence the thermal stability of DNA in a highly sequence dependent manner (1,18,24). A complete characterization of the thermodynamics of an I·C pair within an RNA oligonucleotide requires the measurement of a large number of sequences to generate a complete set of nearest neighbor parameters. Here, we have performed optical melting experiments on a series of RNA duplexes containing I·C pairs, compared the resulting thermodynamics to the same duplexes containing A·C and G–C pairs, and derived nearest neighbor parameters for single I·C pairs adjacent to Watson-Crick pairs.

MATERIALS AND METHODS

Design of sequences

Duplexes containing internal I·C pairs were designed to have melting temperatures at approximately 50°C and to have minimal formation of hairpin structures or misaligned duplexes. Terminal G–C pairs were chosen to prevent end fraying during melting experiments. Duplexes were designed to contain all possible combinations of Watson-Crick base pairs adjacent to the I·C pair. The I·C pair was placed directly in the center of the duplex in 15 of the duplexes. Duplexes containing a single I·C pair in the terminal position were also utilized, including four duplexes with a 3′ terminal I·C pair and four duplexes with a 5′ terminal I·C pair. In addition, four duplexes containing tandem I·C pairs were designed and optically melted.

RNA synthesis and purification

All oligonucleotides containing internal inosine nucleotides were synthesized at the University of Rochester (Rochester, NY, USA) on an Applied Biosystems 392 DNA/RNA synthesizer or obtained from the Keck Lab at Yale University (New Haven, CT, USA). All oligonucleotides containing only standard nucleotides were obtained from Integrated DNA Technologies (Coralville, IA, USA). Purification of oligonucleotides followed standard procedures including Waters Sep-Pak C18 chromatography and preparative thin-layer chromatography (21).

Concentration calculations and duplex formation

The total concentration of each single strand was calculated from the absorbance at 260 nm at 80°C and the extinction coefficient using the Beer-Lambert law. Samples were diluted so that the absorbance was between 0.2 and 2.25. The absorbance was measured at

80°C to disrupt any single-strand folding. Extinction coefficients of the single strands were calculated from the molar extinction coefficient of each base at 260 nm and pairwise values for each nearest neighbor combination at 260 nm. Values for standard bases were obtained from Integrated DNA Technologies (http://www.idtdna.com/support/technical/TechnicalBulletinPDF/Oligonucleotide_Yield_Resuspension_and_Storage.pdf), and values for inosine were obtained from Proligo (http://www.proligo.com/pro_primprobes/PP_08-8_UVAbsorbance.html). Individual single strand concentrations were used to mix equal molar amounts of non-self-complementary strands to form a duplex containing an I·C pair.

Optical melting experiments

Newly formed duplexes were lyophilized and redissolved in 1 M NaCl, 20 mM sodium cacodylate, and 0.5 mM Na_2EDTA, pH 7.0. A melt scheme was designed that consisted of a dilution series resulting in a minimum of nine samples to allow for a concentration range typically >50-fold. Using a heating rate of 1°C/min on a Beckman–Coulter DU800 spectrophotometer with a Beckman–Coulter high performance temperature controller, curves of absorbance at 260 nm versus temperature were obtained.

Determination of thermodynamic parameters for duplexes

MeltWin (25) was used to fit melting curves to a two-state model, assuming linear sloping baselines and temperature-independent ΔH° and ΔS° values (26,27). Additionally, T_M values at different concentrations were used to calculate thermodynamic parameters according to Borer *et al.* (28):

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

The Gibb's free energy change at 37°C was calculated as:

$$\Delta G^\circ_{37} = \Delta H^\circ - (310.15\,\mathrm{K})\,\Delta S^\circ \quad (2)$$

MeltWin (25) requires the sequence of the duplex in order to calculate the concentration of the duplex from the high temperature absorbance and extinction coefficients. *MeltWin* does not recognize inosine. Previous work with inosine showed that if inosine is substituted with cytidine within the *MeltWin* software, thermodynamics differed by <1% from manually calculating the inosine extinction coefficient (21); therefore, for simplicity, the molar extinction coefficient for cytidine was used for inosine within the *MeltWin* software.

Determination of the contribution of I·C pairs to duplex thermodynamics

The total free energy change for duplex formation can be approximated by a nearest neighbor model (29) that is the sum of energy increments for helix initiation and nearest neighbor interactions between each of the bases. For example, for terminal I·C pairs,

$$\begin{aligned}\Delta G^\circ_{37}\begin{pmatrix}\mathrm{GCGCAI}\\ \mathrm{CGCGUC}\end{pmatrix} &= \Delta G^\circ_{37,\mathrm{i}} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix}\\ &+ \Delta G^\circ_{37}\begin{pmatrix}\mathrm{CG}\\ \mathrm{GC}\end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix}\\ &+ \Delta G^\circ_{37}\begin{pmatrix}\mathrm{CA}\\ \mathrm{GU}\end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix} + \Delta G^\circ_{37,\,\mathrm{terminal\,IC}}\end{aligned} \quad (3)$$

where $\Delta G^\circ{}_{37}\begin{pmatrix}\mathrm{GCGCAI}\\ \mathrm{CGCGUC}\end{pmatrix}$ was obtained by optical melting, $\Delta G^\circ{}_{37,\mathrm{i}}$ is the free energy change for duplex initiation, 4.09 kcal/mol, $\Delta G^\circ{}_{37,\,\mathrm{terminal\,IC}}$ is a correction factor for a terminal I·C base pair, and the remaining variables are nearest neighbor values for every nearest neighbor combination present (29). Rearranging to isolate the unknowns, a terminal I·C base pair to duplex thermodynamics, $\Delta G^\circ{}_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix}$ in this example, and the terminal I·C base pair correction factor:

$$\begin{aligned}\Delta G^\circ_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix} + \Delta G^\circ_{37,\,\mathrm{terminal\,IC}} &= \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GCGCAI}\\ \mathrm{CGCGUC}\end{pmatrix} - \Delta G^\circ_{37,\mathrm{i}}\\ -\Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix} - \Delta G^\circ_{37}\begin{pmatrix}\mathrm{CG}\\ \mathrm{GC}\end{pmatrix} &- \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix} - \Delta G^\circ_{37}\begin{pmatrix}\mathrm{CA}\\ \mathrm{GU}\end{pmatrix}\end{aligned} \quad (4)$$

$$\begin{aligned}\Delta G^\circ_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix} + \Delta G^\circ_{37,\,\mathrm{terminal\,IC}} &= -10.14 - 4.09 - (-3.42)\\ -(-2.36) - (-3.42) &- (-2.11)\,\mathrm{kcal/mol}\end{aligned} \quad (5)$$

$$\Delta G^\circ_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix} + \Delta G^\circ_{37,\,\mathrm{terminal\,IC}} = -2.92\,\mathrm{kcal/mol} \quad (6)$$

This calculation was repeated for every duplex with a terminal I·C pair.

Similarly, for internal I·C pairs,

$$\begin{aligned}\Delta G^\circ_{37}\begin{pmatrix}\mathrm{GCAIGGC}\\ \mathrm{CGUCCCG}\end{pmatrix} &= \Delta G^\circ_{37,\mathrm{i}} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{CA}\\ \mathrm{GU}\end{pmatrix}\\ +\Delta G^\circ_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UG}\end{pmatrix} &+ \Delta G^\circ_{37}\begin{pmatrix}\mathrm{IG}\\ \mathrm{CC}\end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GG}\\ \mathrm{CC}\end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix}\end{aligned} \quad (7)$$

where $\Delta G^\circ{}_{37}\begin{pmatrix}\mathrm{GCAIGGC}\\ \mathrm{CGUCCCG}\end{pmatrix}$ was obtained by optical melting. The remaining variables are individual nearest neighbor values. Two of these nearest neighbor values, $\Delta G^\circ{}_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix}$ and $\Delta G^\circ{}_{37}\begin{pmatrix}\mathrm{IG}\\ \mathrm{CC}\end{pmatrix}$, were unknown values that had to be determined experimentally. Rearranging this equation to isolate the unknowns gives the following:

$$\begin{aligned}\Delta G^\circ_{37}\begin{pmatrix}\mathrm{AI}\\ \mathrm{UC}\end{pmatrix} + \Delta G^\circ_{37}\begin{pmatrix}\mathrm{IG}\\ \mathrm{CC}\end{pmatrix} &= \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GCAIGGC}\\ \mathrm{CGUCCCG}\end{pmatrix} - \Delta G^\circ_{37,\mathrm{i}}\\ -\Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix} - \Delta G^\circ_{37}\begin{pmatrix}\mathrm{CA}\\ \mathrm{GU}\end{pmatrix} &- \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GG}\\ \mathrm{CC}\end{pmatrix} - \Delta G^\circ_{37}\begin{pmatrix}\mathrm{GC}\\ \mathrm{CG}\end{pmatrix}\end{aligned} \quad (8)$$

Substituting in known values for the standard nearest neighbor parameters:

$$\Delta G^{\circ}_{37}\begin{pmatrix}\text{AI}\\ \text{UC}\end{pmatrix} + \Delta G^{\circ}_{37}\begin{pmatrix}\text{IG}\\ \text{CC}\end{pmatrix} = -11.42 - 4.09 - (-3.42) \\ -(-2.11) - (-3.26) - (-3.42)\text{kcal/mol} \quad (9)$$

$$\Delta G^{\circ}_{37}\begin{pmatrix}\text{AI}\\ \text{UC}\end{pmatrix} + \Delta G^{\circ}_{37}\begin{pmatrix}\text{IG}\\ \text{CC}\end{pmatrix} = -3.30\,\text{kcal/mol} \quad (10)$$

This calculation was performed for all duplexes containing internal I·C pairs.

In order to determine free energy contributions for the nearest neighbor parameters containing I·C pairs, *Excel*'s LINEST function (a complete linear least squares curve fitting routine) was used to determine linearly independent I·C nearest neighbor values. One tandem duplex melted in a non-two-state manner. The other three tandem duplexes did not contain enough instances of each of the tandem nearest neighbors to result in meaningful parameters and were therefore excluded from analysis. Therefore, to derive nearest neighbor parameters for I·C pairs, we included data from 15 duplexes containing single internal I·C pairs, eight duplexes containing single terminal I·C pairs, and two duplexes from the literature each containing a terminal I·C pair on both ends of the duplex (30).

RESULTS

Thermodynamic parameters

Table 1 lists the thermodynamics of 29 duplexes. Of these, 23 duplexes contain a single I·C pair (four contain a 3′ terminal I·C pair, four contain a 5′ terminal I·C pair, and 15 contain an internal I·C pair). An additional four duplexes contain tandem internal I·C base pairs. Data for these 27 duplexes were combined with data for two additional duplexes from the literature, each containing a terminal I·C pair on both ends of the duplex (30). Table 1 shows the ΔH°, ΔS° and ΔG°_{37} values from the $T_M{}^{-1}$ versus log C_T plots and from the average of curve fits for all duplexes.

It was immediately obvious that the duplex $\begin{pmatrix}\text{GGCCIGCC}\\ \text{CCGICCGG}\end{pmatrix}$ melted in a non-two-state manner, with ΔH° values from the average of curve fits and the log C_T plots differing by >15%. All other melts appeared to be two-state.

Thermodynamic comparison to duplexes containing A·C and G–C pairs

ADAR enzymes act on substrates containing A·C pairs and convert them to products containing I·C pairs (11); therefore, we have used nearest neighbor parameters (29,31–35) to calculate the ΔG°_{37} values for each duplex studied here containing an A·C pair(s) instead of an I·C pair(s) (Table 2). This allows us to compare the stability of substrate to product for an ADAR-mediated deamination. On average, duplexes containing an internal I·C pair are more stable than duplexes containing an internal A·C pair by 4.1 kcal/mol. Also, duplexes containing a terminal I·C pair are more stable than duplexes containing a terminal A·C pair by an average of 0.8 kcal/mol. We also note that I·C pairs are isosteric with G–C pairs (Figure 1), and G-to-I substitution is used to investigate the role of the exocyclic amino group of guanosine. Therefore, we have also used nearest neighbor parameters (29) to calculate the ΔG°_{37} value for each duplex studied here containing a G–C pair instead of an I·C pair (Table 2). Duplexes containing an internal G–C pair are more stable than duplexes containing an internal I·C pair by an average of 2.0 kcal/mol. Similarly, duplexes containing a terminal G–C pair are more stable than duplexes containing a terminal I·C pair by an average of 0.8 kcal/mol. On average, duplexes with tandem I·C pairs were found to be 5.7 kcal/mol (or 2.8 kcal/mol per substitution) more stable than similar duplexes with tandem A·C pairs and 4.0 kcal/mol (or 2.0 kcal/mol per substitution) less stable than similar duplexes with tandem G–C pairs. Comparable calculations for entropy and enthalpy are available in Supplementary Tables S1 and S2.

Contribution of I·C nearest neighbors to duplex thermodynamics

The contribution of I·C nearest neighbors to duplex stability is defined in Equations (4)–(10). The contribution of I·C nearest neighbors to duplex ΔG°_{37}, ΔH° and ΔS° is shown in Supplementary Table S3. The I·C nearest neighbor combinations in every duplex in this study had a stabilizing effect on duplex thermodynamics. This was true for single internal I·C pairs, terminal I·C pairs, and tandem I·C pairs (Supplementary Table S3).

I·C nearest neighbor parameters

A linear regression analysis was completed on the data resulting in the derivation of thermodynamic parameters for all possible combinations of Watson-Crick nearest neighbors containing I·C pairs (Table 3). All I·C/Watson–Crick nearest neighbor combinations contribute negative ΔG°_{37}, negative ΔH°, and negative ΔS° to duplex thermodynamics. A terminal I·C pair was shown to contribute a slightly negative ΔG°_{37} (–0.08 kcal/mol), positive ΔH° (2.0 kcal/mol), and positive ΔS° (6.5 eu) to duplex thermodynamics. Tandem I·C nearest neighbor parameters were not derived from this dataset due to the limited number of occurrences for each possible tandem parameter. Previously published Watson–Crick nearest neighbor parameters (29) as well as the derived I·C nearest neighbor parameters were used to predict the thermodynamics of each duplex measured in this study (Table 1). The average deviation between measured and predicted values was calculated using the data from the van't Hoff plot (or from the average of curve fit data if the van't Hoff data was unavailable). The average deviations between the predicted values and the measured values for all duplexes were 4.6%, 6.2%, 6.6% and 2.0°C for ΔG°_{37}, ΔH°, ΔS° and T_M, respectively. These values are comparable to previously published values for deoxyinosine (3.5%, 4.8%, 5.0% and 1.2°C, respectively) (15), I·U (5.1%, 4.6%, 5.1% and 2.8°C, respectively) (21), and RNA Watson–Crick (3.2%, 6.0%, 6.8% and 1.3°C, respectively) (29) nearest neighbor parameters.

The average deviation between the predicted free energies and the measured free energies for all duplexes was 0.46

Table 1. Thermodynamic parameters of duplex formation

Duplex[a]	T_M^{-1} versus log C_T plots				Average of curve fits				Predicted[e]			
	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)	T_M[b] (°C)	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)	T_M[b] (°C)	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)	T_M[b] (°C)
Internal												
GCAICGC CGUCGCG	77.5 ± 8.3	215.4 ± 25.5	10.75 ± 0.45	54.8	74.8 ± 7.0	206.7 ± 21.5	10.69 ± 0.40	55.3	72.1	197.9	10.68	56.0
GCAIGGC CGUCCCG	72.9 ± 9.8	198.1 ± 29.6	11.42 ± 0.69	59.3	72.4 ± 7.0	196.7 ± 21.8	11.40 ± 0.41	59.4	78.7	215.2	11.92	59.7
GCAIUGC CGUCACG	66.0 ± 3.0	184.8 ± 9.4	8.71 ± 0.08	47.6	71.7 ± 5.2	202.5 ± 16.0	8.88 ± 0.26	47.5	68.9	191.3	9.56	51.2
GCCIAGC CGGCUCG	76.6 ± 8.1	211.3 ± 25.0	11.10 ± 0.45	56.6	72.8 ± 3.5	199.5 ± 10.7	10.92 ± 0.25	56.9	78.1	215.9	11.13	56.4
GCCICGC CGGCGCG	75.4 ± 10.9	202.1 ± 32.5	12.75 ± 0.91	64.9	75.2 ± 5.0	201.4 ± 15.7	12.72 ± 0.20	64.8	73.6	198.0	12.12	62.7
GCCIGGC CGGCCCG	85.6 ± 9.1	230.6 ± 27.2	14.11 ± 0.73	67.1	85.3 ± 6.6	229.4 ± 19.4	14.12 ± 0.58	67.3	80.2	215.3	13.36	65.9
GCCIUGC CGGCACG	76.4 ± 6.3	210.6 ± 19.4	11.11 ± 0.33	56.7	72.5 ± 3.4	198.5 ± 9.9	10.94 ± 0.30	57.1	70.4	191.4	11.00	58.1
GCGIAGC CGCCUCG	75.4 ± 5.3	207.2 ± 16.4	11.12 ± 0.26	57.1	76.1 ± 6.7	209.4 ± 20.8	11.13 ± 0.34	57.0	79.4	220.8	10.99	55.4
GCGICGC CGCCGCG	68.6 ± 8.7	184.3 ± 26.2	11.40 ± 0.63	60.7	69.6 ± 6.9	187.3 ± 20.4	11.48 ± 0.62	60.7	74.9	202.9	11.98	61.4
GCGIGGC CGCCCCG	83.9 ± 5.9	226.2 ± 17.6	13.75 ± 0.48	66.2	83.5 ± 10.7	224.8 ± 31.5	13.75 ± 0.92	66.4	81.5	220.2	13.22	64.7
GCGIUGC CGCCACG	72.2 ± 4.1	198.3 ± 12.6	10.65 ± 0.23	55.8	72.1 ± 5.5	198.1 ± 16.5	10.66 ± 0.39	55.8	71.7	196.3	10.86	56.9
GCUIAGC CGACUCG	81.7 ± 5.1	234.5 ± 16.1	8.97 ± 0.15	46.5	82.4 ± 10.1	236.4 ± 31.8	9.05 ± 0.28	46.8	74.2	210.2	9.05	47.9
GCUICGC CGACGCG	71.7 ± 2.0	196.4 ± 6.1	10.81 ± 0.09	56.7	69.3 ± 1.3	189.0 ± 3.8	10.70 ± 0.09	56.9	69.7	192.3	10.04	53.5
GCUIGGC CGACCCG	75.5 ± 1.4	206.9 ± 4.2	11.33 ± 0.08	58.1	72.3 ± 6.6	197.0 ± 19.9	11.20 ± 0.43	58.4	76.3	209.7	11.28	57.5
GCUIUGC CGACACG	56.6 ± 13.2	156.4 ± 42.0	8.11 ± 0.91	45.9	57.2 ± 7.6	158.0 ± 24.1	8.17 ± 0.34	46.2	66.5	185.8	8.92	48.5
Terminal[f]												
GCGCAI CGCGUC	62.6 ± 3.4	169.2 ± 10.6	10.14 ± 0.17	56.0	60.3 ± 4.7	162.0 ± 14.2	10.05 ± 0.29	56.2	59.5	163.1	8.87	49.7
GCGCCI CGCGGC	62.2 ± 3.7	167.2 ± 11.3	10.34 ± 0.20	57.2	61.1 ± 4.1	163.7 ± 12.1	10.31 ± 0.32	57.5	61.0	163.2	10.31	57.7
GCGCGI CGCGCC	69.9 ± 6.4	192.3 ± 19.8	10.28 ± 0.31	54.6	68.3 ± 2.5	187.1 ± 7.6	10.24 ± 0.22	54.8	62.3	168.1	10.17	56.3
GCGCUI CGCGAC	58.3 ± 6.3	161.4 ± 19.8	8.28 ± 0.27	46.6	58.2 ± 5.3	161.1 ± 16.6	8.24 ± 0.29	46.4	57.1	157.6	8.23	46.6
IAGCGC CUCGCG	58.7 ± 8.7	162.2 ± 27.1	8.42 ± 0.49	47.3	56.9 ± 8.4	156.4 ± 26.5	8.40 ± 0.31	47.6	60.7	168.3	8.45	47.2
ICGCGC CGCGCG	48.2 ± 8.9	126.2 ± 26.8	9.10 ± 0.89	54.4	46.6 ± 8.4	120.7 ± 25.4	9.14 ± 0.76	55.4	56.1	150.4	9.44	54.2
IGGCGC CCCGCG	61.5 ± 12.3	166.5 ± 37.3	9.82 ± 0.92	54.5	61.5 ± 4.8	166.7 ± 15.4	9.82 ± 0.32	54.5	62.7	167.8	10.68	59.0
IUGCGC CACGCG	59.2 ± 2.9	158.5 ± 8.9	10.07 ± 0.16	56.7	59.7 ± 3.3	159.9 ± 10.0	10.10 ± 0.23	56.7	53.0	143.9	8.32	47.9
ICCGGC[c] CGGCCI	N.A.	N.A.	N.A.	N.A.	53.6	146.9	8.03	51.3	51.2	138.4	8.30	47.9
CGGCCI[c] ICCGGC	N.A.	N.A.	N.A.	N.A.	53.0	143.2	8.55	54.8	59.6	162.2	9.30	52.1

Table 1. Continued

Duplex[a]	T_M^{-1} versus log C_T plots				Average of curve fits				Predicted[e]			
	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)	T_M[b] (°C)	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)	T_M[b] (°C)	$-\Delta H^\circ$ (kcal/mol)	$-\Delta S^\circ$ (eu)	$-\Delta G^\circ_{37}$ (kcal/mol)	T_M[b] (°C)
Tandem[g]												
GUCCIGAC CAGICCUG	77.5 ± 1.9	220.8 ± 6.0	9.02 ± 0.07	51.0	72.6 ± 4.3	205.3 ± 13.5	8.89 ± 0.17	51.4	N.P.	N.P.	N.P.	N.P.
GGCICGCC CCGCICGG	84.9 ± 3.6	232.0 ± 10.6	12.91 ± 0.29	65.9	85.8 ± 4.2	234.8 ± 12.5	13.00 ± 0.35	65.9	N.P.	N.P.	N.P.	N.P.
GGCCIGCC[d] CCGICCGG	94.7 ± 3.2	260.5 ± 9.6	13.96 ± 0.26	66.7	72.7 ± 4.2	194.9 ± 12.8	12.21 ± 0.36	67.7	N.P.	N.P.	N.P.	N.P.
GGCIIGGC CCGCCCCG	90.7 ± 7.2	244.6 ± 21.3	14.82 ± 0.58	68.2	85.5 ± 4.5	229.0 ± 13.4	14.44 ± 0.42	68.6	N.P.	N.P.	N.P.	N.P.

[a]Solutions are 1 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na_2EDTA, pH 7.0. The top strand of each duplex is written 5′ to 3′, and the bottom strand is written 3′ to 5′.
[b]Calculated for 10^{-4} M oligonucleotide concentration.
[c]Ref (30).
[d]The duplex appears to melt in a non-two-state manner.
[e]Predicted thermodynamic parameters based on a linear regression model using experimental data.
[f]N.A. indicates the data is not available.
[g]N.P. indicates no available parameters.

kcal/mol, illustrating the goodness-of-fit for the derived model. In order to test the predictive power of the model, a leave-one-out cross-validation was performed. Here, one data point was selected as the test set, and the thermodynamic model was built with LINEST using all of the remaining data points. The resulting model was used to predict the free energy of the data point that was held out. This procedure was repeated for each data point, resulting in an average deviation of 0.75 kcal/mol, illustrating how the model might work in a predictive context.

DISCUSSION

Thermodynamic comparison of duplexes containing I·C pairs to duplexes containing A·C and G–C pairs

Comparison of the thermodynamics of duplexes containing A·C pairs to duplexes containing I·C pairs allows for the comparison of the stability of ADAR substrates to products. Similarly, the comparison of the thermodynamics of duplexes containing G–C pairs to duplexes containing I·C pairs allows for the comparison of duplex stability before and after intentional incorporation of the non-standard nucleotide inosine as a probe for the role of the exocyclic amino group of G. It can be assumed that an A·C, I·C, and G–C pair in these short oligonucleotides adopt a conformation with one (N6 amino of A to N3 of C), two (O6 carbonyl of I to N4 amino of C and N1 imino of I to N3 of C), and three hydrogen bonds (O6 carbonyl of G to N4 amino of C, N1 imino of G to N3 of C, and N2 amino of G to O2 carbonyl of C), respectively (Figure 1).

Note that G–C and I·C pairs are isosteric and both contain O6 carbonyl to N4 amino of C and N1 imino to N3 of C hydrogen bonds. Therefore, the most direct comparison between duplexes can be made between those containing an I·C pair and those containing a G–C pair. The major difference between an I·C and G–C pair is the additional N2 amino of G to O2 carbonyl of C hydrogen bond in a G–C pair (Figure 1). Based on the data reported here, a duplex containing an internal I·C pair is 2.0 ± 0.6 kcal/mol less stable than the same duplex containing an internal G–C pair, suggesting that the N2 amino of G to O2 carbonyl of C hydrogen bond is worth ∼-2.0 kcal/mol (although other factors such as electrostatic effects on the strength of hydrogen bonds, stacking, pKa, local structural variations, and hydration could also affect stability). However, this appears to be sequence dependent, with experimental values ranging from –0.8 to –3.2 kcal/mol (Table 2). Similar values for the contribution of a hydrogen bond have been reported in the literature (36).

The comparison between duplexes containing an I·C pair and those containing an A·C pair are less direct. An A·C pair and an I·C pair are not isosteric. If an ADAR enzyme converted an A to an I, the resulting I would need to shift position in order to form two hydrogen bonds with C. This shift likely affects the stacking interactions with the bases adjacent to the I·C pair. Based on the discussion above and focusing solely on the number of hydrogen bonds, one could hypothesize that an internal I·C pair, with two hydrogen bonds, would be ∼2 kcal/mol more stable than an internal A·C pair with one hydrogen bond. However, the data reported here shows that duplexes containing an internal I·C pair are, on average, 4.1 ± 0.7 kcal/mol more stable than the same duplexes containing internal A·C pairs, more than twice as expected based on the I·C versus G–C results. This also appears to be sequence dependent, with experimental values ranging from 3.1 to 5.4 kcal/mol. Perhaps computational or structural studies could shed some light on the surprising stability of I·C pairs in comparison to A·C pairs.

Similar comparisons can be made between duplexes containing terminal I·C pairs and those containing terminal A·C and G–C pairs. Here, the results are less surprising. On average, duplexes with a terminal G–C pair are 0.8 kcal/mol more stable than duplexes with a terminal I·C pair (with a range of 0.8 kcal/mol less stable to 1.8 kcal/mol more stable), suggesting that the extra terminal hydrogen bond is worth ∼0.8 kcal/mol. Similarly, on average, duplexes with a terminal I·C pair are 0.8 kcal/mol more stable (with a range of 0.1 to 2.3 kcal/mol) than duplexes with a terminal A·C

Table 2. Free energy comparison between duplexes containing I·C, A·C and G–C pairs

Duplex[a]	$-\Delta G^{o}_{37}$ (kcal/mol)[b]	–NN A–C (kcal/mol)[c]	ΔA–C (kcal/mol)[d]	–NN G–C (kcal/mol)[e]	ΔG–C (kcal/mol)[f]	Duplex[a]	$-\Delta G^{o}_{37}$ (kcal/mol)[b]	–NN A–C (kcal/mol)[c]	ΔA–C (kcal/mol)[d]	–NN G–C (kcal/mol)[e]	ΔG–C (kcal/mol)[f]
Internal						*Terminal*					
GCAICGC CGUCGCG	10.75	7.02	−3.73	12.72	1.97	GCGCAI CGCGUC	10.14	8.02	−2.12	9.30	−0.84
GCAIGGC CGUCCCG	11.42	6.92	−4.50	13.46	2.04	GCGCCI CGCGGC	10.34	9.87	−0.47	10.73	0.39
GCAIUGC CGUCACG	8.71	5.57	−3.14	11.29	2.58	GCGCGI CGCGCC	10.28	8.67	−1.61	10.73	0.45
GCCIAGC CGGCUCG	11.10	6.89	−4.21	12.80	1.70	GCGCUI CGCGAC	8.28	7.99	−0.29	9.30	1.02
GCCICGC CGGCGCG	12.75	8.77	−3.98	14.15	1.40	IAGCGC CUCGCG	8.42	7.89	−0.53	9.54	1.12
GCCIGGC CGGCCCG	14.11	9.27	−4.84	14.89	0.78	ICGCGC CGCGCG	9.10	8.57	−0.53	10.46	1.36
GCCIUGC CGGCACG	11.11	7.32	−3.79	12.72	1.61	IGGCGC CCCGCG	9.82	9.37	−0.45	11.63	1.81
GCGIAGC CGCCUCG	11.12	5.99	−5.13	12.80	1.68	IUGCGC CACGCG	10.07	7.82	−2.25	9.46	−0.61
GCGICGC CGCCGCG	11.40	8.47	−2.93	14.15	2.75	ICCGGC[g] CGGCCI	8.03	6.56	−1.47 (−0.73)	11.20	3.17 (1.59)
GCGIGGC CGCCCCG	13.75	8.37	−5.38	14.89	1.14	CGGCCI[g] ICCGGC	8.55	8.42	−0.13 (−0.07)	10.14	1.59 (0.79)
GCGIUGC CGCCACG	10.65	7.02	−3.63	12.72	2.07	Average[h]			−0.82		0.79
GCUIAGC CGACUCG	8.97	4.51	−4.46	11.37	2.40						
						Tandem					
GCUICGC CGACGCG	10.81	6.39	−4.42	12.72	1.91	GUCCIGAC CAGICCUG	9.02	3.66	−5.36 (−2.68)	13.54	4.52 (2.26)
GCUIGGC CGACCCG	11.33	6.89	−4.44	13.46	2.13	GGCICGCC CCGCICGG	12.91	7.84	−5.07 (−2.54)	16.98	4.07 (2.04)
GCUIUGC CGACACG	8.11	4.94	−3.17	11.29	3.18	GGCIIGGC CCGCCCCG	14.82	8.27	−6.55 (−3.28)	18.15	3.33 (1.67)
Average			−4.12		1.96	**Average**[h]			−2.83		1.99

[a]For each duplex, the top strand is written 5′ to 3′, and the bottom strand is written 3′ to 5′.
[b]The measured $-\Delta G^{\circ}_{37}$ of the listed duplex, using values from the log C_T plots.
[c]$-\Delta G^{\circ}_{37}$ calculated using the nearest neighbor model (29,31–35) for the duplex if the I·C pair was an A·C pair.
[d]The difference in $-\Delta G^{\circ}_{37}$ between the –NN A·C value and the measured $-\Delta G^{\circ}_{37}$. Values shown in parenthesis represent the free energy difference per I·C pair since duplex contains two I·C pairs.
[e]$-\Delta G^{\circ}_{37}$ calculated using the nearest neighbor model, reference (29), for the duplex if the I·C pair was a G–C pair.
[f]The difference in $-\Delta G^{\circ}_{37}$ between the –NN G–C value and the measured $-\Delta G^{\circ}_{37}$. Values shown in parenthesis represent the free energy difference per I·C pair since duplex contains two I·C pairs.
[g]Ref(30).
[h]Average free energy difference per I·C pair.

pair, which is consistent with a terminal hydrogen bond being worth ∼0.8 kcal/mol. These results are consistent with the number of hydrogen bonds between the A·C (one hydrogen bond), I·C (two hydrogen bonds) and G–C (three hydrogen bonds) pairs. The reported value of -0.8 kcal/mol for a hydrogen bond within a terminal RNA base pair is consistent with literature values derived from other RNA systems (36).

Unlike A–U pairs, there is no free energy penalty assigned to a terminal I-C pair (Table 3). The 0.45 kcal/mol penalty assigned to terminal A–U pairs accounts for the one fewer hydrogen bond in comparison to a terminal G–C pair (29). Both terminal G–U pairs (37) and terminal I-C pairs, which

Table 3. Nearest neighbor parameters for I·C pairs

Nearest neighbors[a]	Number of occurrences[b]	ΔH° (kcal/mol)	ΔS° (eu)	ΔG°_{37} (kcal/mol)	RECCES-Rosetta prediction[c] ΔG°_{37} (kcal/mol)	Difference in ΔG°_{37}[d] (kcal/mol)	Percent Difference in ΔG°_{37}[e]
IG CC	5	− 14.5 ± 3.1	− 39.6 ± 9.2	− 2.23 ± 0.40	− 2.24 ± 0.16	0.01	0.45%
IC CG	7	− 10.6 ± 2.4	− 28.2 ± 7.0	− 1.89 ± 0.31	− 1.96 ± 0.13	0.07	3.64%
IA CU	4	− 15.3 ± 3.4	− 45.7 ± 10.0	− 1.18 ± 0.44	− 1.06 ± 0.11	− 0.12	10.71%
IU CA	5	− 7.7 ± 3.1	− 21.5 ± 9.2	− 1.02 ± 0.40	− 0.95 ± 0.21	− 0.07	7.11%
GI CC	5	− 16.8 ± 3.1	− 45.9 ± 9.2	− 2.62 ± 0.40	− 2.07 ± 0.26	− 0.55	23.45%
CI GC	7	− 12.7 ± 2.4	− 35.0 ± 7.0	− 1.86 ± 0.31	− 1.98 ± 0.20	0.12	6.25%
AI UC	4	− 14.2 ± 3.4	− 40.7 ± 10.0	− 1.57 ± 0.44	− 1.09 ± 0.14	− 0.48	36.09%
UI AC	5	− 11.8 ± 3.1	− 35.0 ± 9.2	− 0.96 ± 0.40	− 0.98 ± 0.25	0.02	2.06%
Terminal I·C	12	2.0 ± 1.9	6.5 ± 5.5	− 0.08 ± 0.24	0.88 ± 0.21	− 0.96	240.00%

[a]For each nearest neighbor pair, the top sequence is written 5′ to 3′, and the bottom sequence is written 3′ to 5′.
[b]The number of times that nearest neighbor pair appears in the sequences studied.
[c]Predictions from computer model using the RECCES-Rosetta framework (38).
[d]Difference between the experimentally derived ΔG°_{37} and RECCES-Rosetta model prediction.
[e]Percent difference between the experimentally derived ΔG°_{37} and RECCES-Rosetta model prediction.

can form only two hydrogen bonds, are not assigned a free energy penalty. It is likely, as proposed previously for G–U pairs (37), that terminal G–U and I-C pairs are flexible, allowing for optimization of hydrogen bonding and stacking interactions without distorting the backbone, perhaps compensating for the one fewer hydrogen bond.

I·C nearest neighbor parameters

The stability of an RNA duplex containing an I·C pair can be calculated using a combination of the existing nearest neighbor model (29) with the newly derived I·C nearest neighbor parameters (Table 3). For an example with a terminal I·C pair, the stability of $\binom{\text{GCGCAI}}{\text{CGCGUC}}$ can be calculated as follows:

$$\Delta G^\circ_{37}\binom{\text{GCGCAI}}{\text{CGCGUC}} = \Delta G^\circ_{37,\text{i}} + \Delta G^\circ_{37}\binom{\text{GC}}{\text{CG}} + \Delta G^\circ_{37}\binom{\text{CG}}{\text{GC}} + \Delta G^\circ_{37}\binom{\text{GC}}{\text{CG}} + \Delta G^\circ_{37}\binom{\text{CA}}{\text{GU}} + \Delta G^\circ_{37}\binom{\text{AI}}{\text{UC}} + \Delta G^\circ_{37,\text{ terminal IC}} \quad (11)$$

$$\Delta G^\circ_{37}\binom{\text{GCGCAI}}{\text{CGCGUC}} = 4.09 + (-3.42) + (-2.36) + (-3.42) + (-2.11) + (-1.57) + (-0.08) \quad (12)$$

$$\Delta G^\circ_{37}\binom{\text{GCGCAI}}{\text{CGCGUC}} = -8.87\,\text{kcal/mol} \quad (13)$$

For an example with an internal I·C pair, the stability of $\binom{\text{GCAIGGC}}{\text{CGUCCCG}}$ can be calculated as follows:

$$\Delta G^\circ_{37}\binom{\text{GCAIGGC}}{\text{CGUCCCG}} = \Delta G^\circ_{37,\text{i}} + \Delta G^\circ_{37}\binom{\text{GC}}{\text{CG}} + \Delta G^\circ_{37}\binom{\text{CA}}{\text{GU}} + \Delta G^\circ_{37}\binom{\text{AI}}{\text{UC}} + \Delta G^\circ_{37}\binom{\text{IG}}{\text{CC}} + \Delta G^\circ_{37}\binom{\text{GG}}{\text{CC}} + \Delta G^\circ_{37}\binom{\text{GC}}{\text{CG}} \quad (14)$$

$$\Delta G^\circ_{37}\binom{\text{GCAIGGC}}{\text{CGUCCCG}} = 4.09 + (-3.42) + (-2.11) + (-1.57) + (-2.23) + (-3.26) + (-3.42) \quad (15)$$

$$\Delta G^\circ_{37}\binom{\text{GCAIGGC}}{\text{CGUCCCG}} = -11.92\,\text{kcal/mol} \quad (16)$$

Scientists can use these new nearest neighbor parameters to calculate the stability of ADAR products and to calculate the stability of an RNA duplex in which G-to-I substitution was used to determine the role of the exocyclic amino group of G. To completely evaluate the value of using ribo-inosine as an oligonucleotide probe when the exact sequence of the

RNA target is unknown, an extensive set of data for RNA duplexes containing I·A and I·G pairs is required.

Experimental model compared to energy predictions

The Das lab recently reported blind tests of a method to computationally predict nearest neighbor energetic parameters (38). They developed a reweighting of energy-function collection with conformational ensemble sampling in a Rosetta (RECCES-Rosetta) framework. This purely predictive model factors in separate component terms for hydrogen bonding, electrostatics, van der Waals interactions, base stacking, torsional potentials and an orientation-dependent solvation model (38) and does not require any experimental data. Those authors used this model to blindly predict the nearest neighbor parameters for all eight I·C nearest neighbors as well as a value for a terminal I·C. These predictive values are compared to our experimentally-derived nearest neighbor parameters (Table 3). For most parameters, the RECCES-Rosetta predictions are quite accurate. Six of the predicted parameters are very close to the experimental parameters, with an average percent difference of only 5.0% (or an average difference of only 0.07 kcal/mol). A seventh predicted parameter, $\binom{\text{AI}}{\text{UC}}$, is within experimental error of the experimental value but has a percent difference of 36.1% (or a difference of 0.48 kcal/mol). An eighth predicted parameter, $\binom{\text{GI}}{\text{CC}}$, is within experimental error of the experimental value but has a percent difference of 23.5% (or a difference of 0.55 kcal/mol). Only one parameter is not within experimental error of the experimental value. The terminal I·C parameter is predicted to be 0.96 kcal/mol less stable than measured experimentally (percent difference of 240.0%).

Chou *et al.* (2016) used a different metric to evaluate the accuracy of their RECCES-Rosetta predictions. They calculated an rmsd value between their RECCES-Rosetta predictions and the experimental data. Since I·C data is now available, we calculated the rmsd between the RECCES-Rosetta predictions and our derived I-C nearest neighbor parameters. This rmsd value (0.41) is within the range calculated by Chou *et al.* for other RNA motif categories. Canonical pairs had the lowest rmsd value (0.28) while isoG–isoC pairs had the highest rmsd value (0.99), with the average of all motifs being an rmsd value of 0.50 (38). We are hopeful that these new experimental data can be used to refine the RECCES-Rosetta model to make it even more valuable to RNA researchers.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Saint Louis University Department of Chemistry, Research Corporation [CC6547 to B.M.Z.]; National Institutes of Health (NIH) [2R15GM085699-03 to B.M.Z.]. Funding for open access charge: NIH.
Conflict of interest statement. None declared.

REFERENCES

1. Martin,F.H. and Castro,M.M. (1985) Base pairing involving deoxyinosine: Implications for probe design. *Nucleic Acids Res.*, **13**, 8927–8938.
2. Corfield,P.W., Hunter,W.N., Brown,T., Robinson,P. and Kennard,O. (1987) Inosine.adenine base pairs in a B-DNA duplex. *Nucleic Acids Res.*, **15**, 7935–7949.
3. Kawase,Y., Iwai,S., Inoue,H., Miura,K. and Ohtsuka,E. (1986) Studies on nucleic acid interactions. I. Stabilities of mini-duplexes ($dG_2A_4XA_4G_2$-$dC_2T_4YT_4C_2$) and self-complementary d(GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. *Nucleic Acids Res.*, **14**, 7727–7736.
4. Nishikura,K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, **79**, 321–349.
5. Nigita,G., Veneziano,D. and Ferro,A. (2015) A-to-I RNA editing: current knowledge sources and computational approaches with special emphasis on non-coding RNA molecules. *Front. Bioeng. Biotechnol.*, **3**, 1–7.
6. Liu,H., Wang,Q., He,Y., Chen,L., Hao,C., Jiang,C., Li,Y., Dai,Y., Kang,Z. and Xu,J.-R. (2016) Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. *Genome Res.*, **26**, 499–509.
7. Su,A.A.H. and Randau,L. (2011) A-to-I and C-to-U editing within transfer RNAs. *Biochemistry-Moscow*, **76**, 932–937.
8. Paul,M.S. and Bass,B.L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.*, **17**, 1120–1127.
9. Mass,S., Kawahara,Y., Tamburro,K.M. and Nishikura,K. (2006) A-to-I RNA editing and human disease. *RNA Biol.*,**3**, 1–9.
10. Wong,S.K., Sato,S. and Lazinski,D.W. (2001) Substrate recognition by ADAR1 and ADAR2. *RNA*, **7**, 846–858.
11. Bass,B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.*, **71**, 817–846.
12. Slotkin,W. and Nishikura,K. (2013) Adenosine-to-inosine RNA editing and human disease. *Genome Med.*, **5**, 13.
13. Rice,G.I., Kasher,P.R., Forte,G.M.A., Mannion,N.M., Greenwood,S.M., Szynkiewicz,M., Dickerson,J.E., Bhaskar,S.S., Zampini,M., Briggs,T.A. *et al.* (2012) Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat. Genet.*, **44**, 1243–1248.
14. Kumar,V.D., Harrison,R.W., Andrews,L.C. and Weber,I.T. (1992) Crystal structure at 1.5-angstroms resolution of d(CGCICICG), an octanucleotide containing inosine and its comparison with d(CGCG) and d(CGCGCG) structures. *Biochemistry*, **31**, 1541–1550.
15. Watkins,N.E. and SantaLucia,J. (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.*, **33**, 6258–6267.
16. Zimmermann,R.A., Gait,M.J. and Moore,M.J. (1998) Incorporation of modified nucleotides into RNA for studies on RNA structure, function and intermolecular interactions. In: Grosjean,H and Benne,R (eds). *Modification and Editing of RNA*. ASM Press, Washington, D.C., pp. 59–79.
17. Grohman,J.K., Gorelick,R.J., Lickwar,C.R., Lieb,J.D., Bower,B.D., Znosko,B.M. and Weeks,K.M. (2013) A guanosine-centric mechanism for RNA chaperone function. *Science*, **340**, 190–195.
18. Keane,P.M., Hall,J.P., Poynton,F.E., Poulsen,B.C., Gurung,S.P., Clark,I.P., Sazanovich,I.V., Towrie,M., Gunnlaugsson,T., Quinn,S.J. *et al.* (2017) Inosine can increase DNA′s susceptibility to photo-oxidation by a Ru^{II} complex due to structural change in the minor groove. *Chem. Eur. J.*, **23**, 10344–10351.
19. Kelley,S.O. and Barton,J.K. (1999) Electron transfer between bases in double helical DNA. *Science*, **283**, 375–381.
20. Wan,C., Fiebig,T., Schiemann,O., Barton,J.K. and Zewail,A.H. (2000) Femtosecond direct observation of charge transfer between bases in DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 14052–14055.
21. Wright,D.J., Rice,J.L., Yanker,D.M. and Znosko,B.M. (2007) Nearest neighbor parameters for inosine·uridine pairs in RNA duplexes. *Biochemistry*, **46**, 4625–4634.
22. Jolley,E.A., Lewis,M. and Znosko,B.M. (2015) A computational model for predicting experimental RNA nearest-neighbor free energy rankings: Inosine·uridine pairs. *Chem. Phys. Lett.*, **639**, 157–160.
23. Murphy,F.V. and Ramakrishnan,V. (2004) Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat. Struct. Mol. Biol.*, **11**, 1251–1252.

24. Reich,N.O. and Sweetnam,K.R. (1994) Sequence-dependent effects on DNA stability resulting from guanosine replacements by inosine. *Nucleic Acids Res.*, **22**, 2089–2093.
25. 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.
26. Petersheim,M. and Turner,D.H. (1983) Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with CCGG, CCGGp, mCCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry*, **22**, 256–263.
27. 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.
28. Borer,P.N., Dengler,B., Tinoco,I. and Uhlenbeck,O. (1974) Stability of ribonucleic-acid double-stranded helices. *J. Mol. Biol.*, **86**, 843–853.
29. Xia,T.B., SantaLucia,J. Jr, Burkard,M. E., Kierzek,R., Schroeder,S.J., Jiao,X.Q., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**, 14719–14735.
30. Turner,D.H., Sugimoto,N., Kierzek,R. and Dreiker,S.D. (1987) Free energy increments for hydrogen bonds in nucleic acid base pairs. *J. Am. Chem. Soc.*, **109**, 3783–3785.
31. Davis,A.R. and Znosko,B.M. (2007) Thermodynamic characterization of single mismatches found in naturally occurring RNA. *Biochemistry*, **46**, 13425–13436.
32. Sheehy,J.P., Davis,A.R. and Znosko,B.M. (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA*, **16**, 417–429.
33. 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.
34. Christiansen,M.E. and Znosko,B.M. (2008) Thermodynamic characterization of the complete set of sequence symmetric tandem mismatches in RNA and an improved model for predicting the free energy contribution of sequence asymmetric tandem mismatches. *Biochemistry*, **47**, 4329–4336.
35. Christiansen,M.E. and Znosko,B.M. (2009) Thermodynamic characterization of tandem mismatches found in naturally occurring RNA. *Nucleic Acids Res.*, **37**, 4696–4706.
36. Jolley,E.A. and Znosko,B.M. (2017) The loss of a hydrogen bond: Thermodynamic contributions of a non-standard nucleotide. *Nucleic Acids Res.*, **45**, 1479–1487.
37. 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.
38. Chou,F.-C., Kladwang,W., Kappel,K. and Das,R. (2016) Blind tests of RNA nearest-neighbor energy prediction. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 8430–8435.