

Nearest-neighbor parameter for inosine-cytosine pairs through a combined experimental and computational approach

Shun Sakuraba^{1,2}, Junichi Iwakiri¹, Michiaki Hamada^{3,4}, Tomoshi Kameda⁵, Genichiro Tsuji^{6,7}, Yasuaki Kimura⁶, Hiroshi Abe⁶, and Kiyoshi Asai^{*1,5}

¹Graduate School of Frontier Sciences, The University of Tokyo, Japan

²Quantum Beam Science Research Directorate, National Institutes for Quantum and Radiological Science and Technology, Japan

³Faculty of Science and Engineering, Waseda University, Japan

⁴Computational Bio Big-Data Open Innovation Laboratory (CBBDOIL), National Institute of Advanced Industrial Science and Technology (AIST), Japan

⁵Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), Japan

⁶Department of Chemistry, Graduate School of Science, Nagoya University, Furo, Chikusa, Nagoya 464-8602, Japan

⁷Division of Organic Chemistry, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan

Short Title Nearest-neighbor parameters for inosine-cytosine pairs

Key words Inosine-cytosine pair, nearest-neighbor model, free energy, modified RNAs, molecular dynamics simulation

*E-mail: asai at k dot u-tokyo.ac.jp, Phone: +81-47-136-3986, Fax: +81-47-136-3986

Abstract

In RNA secondary structure prediction, nearest-neighbor parameters are used to determine the stability of a given structure. We derived the nearest-neighbor parameters for RNAs containing inosine-cytosine pairs. For parameter derivation, we developed a method that combines UV adsorption measurement experiments with free-energy calculations using molecular dynamics simulations. The method provides fast drop-in parameters for modified bases. Derived parameters were compared and found to be consistent with existing parameters for canonical RNAs. A duplex with an internal inosine-cytosine pair is 0.9 kcal/mol more unstable than the same duplex with an internal guanine-cytosine pair, and is as stable as the one with an internal adenine-uracil pair (only 0.1 kcal/mol more stable) on average.

1 Introduction

RNA molecules undergo various types of co- and post-transcriptional modifications. According to the MODOMICS database (Boccaletto et al., 2018), over 150 types of modifications are enzymatically introduced to transcripts across various species. Adenosine-to-inosine modification (a.k.a. A-to-I RNA editing) is catalyzed by adenosine deaminases acting on RNA (ADARs) converting adenosine in a RNA duplex (i.e., adenosine in A-U base-pairs) to inosine (Nishikura, 2010) and is one of the most prominent modifications in mammals. Currently, high-throughput sequencing methods, such as RNA-seq (Tan et al., 2017) and ICE-seq (Sakurai et al., 2014), provide the transcriptomic landscape of the A-to-I RNA editome. In the human tran-

scriptome, more than 20,000 protein-coding/non-coding RNAs harboring 2.5 million A-to-I editing sites have been collected in the RADAR database (Ramaswami and Li, 2014).

Over the past decades, the secondary structures of RNA molecules have been experimentally investigated and computationally predicted, clarifying the relationships between their molecular functions and secondary structures (Hofacker, 2014; Seetin and Mathews, 2012). For example, pre-miRNAs and tRNAs are characterized by hairpin (Kozomara and Griffiths-Jones, 2014) and cloverleaf (Chan and Lowe, 2009) structures, respectively, despite the diversity of their primary sequences (Griffiths-Jones et al., 2003). Most of the computational predictions of RNA secondary structures are based on the nearest-neighbor (NN) model in which the stability of RNA duplexes is estimated as the summation of the free-energies of two adjacent base-pairs. These parameters, also known as NN parameters (Mathews et al., 1999; Xia et al., 1998), are derived from the thermodynamic measurement of RNA duplexes consisting of various pairs of the four standard RNA nucleotides (A, C, G, and U).

It is well known (Crick, 1966; Nishikura, 2010) that inosine forms a Watson-Crick type pair with cytosine and a wobble pair with uridine (Fig. 1A). However, computational predictions of secondary structures for inosine-containing RNA molecules are currently difficult because the NN parameters for inosine-cytosine (I-C) pairs have not been investigated, though the parameters for inosine-uridine (I•U) pairs have been reported (Wright et al., 2007). In an alternative method, inosines are treated as guanines in primary sequences (based on their similarity) for the purpose of predicting the

secondary structures of inosine-containing RNAs. However, this method overestimates the stability of I-C pairs, because a G-C pair consists of three hydrogen bonds, whereas an I-C pair consists of only two.

In previous research, we have shown that the combination of experimental free-energy measurements and free-energy calculations based on physicochemical simulation can be used to accurately estimate NN parameters (Nishida et al., 2018; Sakuraba et al., 2015). In the present work, we applied this method to determine the NN parameters for inosine-cytosine pairs (Fig. 1B). We compare the derived NN parameters with the corresponding parameters for other bases.

2 Results

Experimental and computational duplex forming free-energy differences Tables 1 and 2 list the experimentally observed ΔG_{37}° values and computationally derived $\Delta\Delta G_{37}^{\circ}$ values, respectively. Each ΔG_{37}° obtained from experimental measurement had high reproducibility, within a 0.17 kcal/mol error. Errors observed in computationally obtained $\Delta\Delta G_{37}^{\circ}$ values were also within a similar range of the previously reported results.

NN parameters and their validation Combining both the experimentally observed and computationally derived free energies, we derived NN parameters for I•U pairs as $\Delta G_{37,\text{exp+comp}}$ (Table 3).

To validate these parameters, we first predicted the free-energy differences for the duplex formation based on the derived NN model and com-

pared them with the experimental ΔG values and computational $\Delta\Delta G$ values. Figure 2 presents the observed or calculated free energies versus the predicted free energies. As explained in the Methods section, the NN parameters were determined in favor of experimentally determined values. This is reflected in Fig. 2, as the experimental ΔG values have smaller deviations (mean absolute difference (MAD), 0.20 kcal/mol) from the predicted values while the calculated $\Delta\Delta G$ values have larger deviations (MAD, 0.55 kcal/mol). Overall, the NN-predicted free energies are in good agreement with the observed ΔG values and calculated $\Delta\Delta G$ values, demonstrating the consistency between the experimental ΔG values and the calculated ΔG values.

The parameters were further validated using an off-by-one test. Because the parameters were determined with larger errors in the simulated free-energy difference, there is a risk that the results overfit the experimental free-energy difference. To address this possibility, we removed one experimental ΔG_{37}° value from the dataset and derived the NN parameters. Then, we estimated the binding free-energy according to the NN model and compared it with the real experimental ΔG_{37}° value. We repeated this procedure for all experimentally derived ΔG_{37}° values (Fig. 3). Values from the model and the experiment agreed well (MAD = 0.41 kcal/mol and $R^2 = 0.94$), indicating that the estimated free-energies are not sensitive to the elimination of a sample. We also tested whether the NN parameters are stable using an off-by-one analysis. The standard deviation of the parameters were within 0.12 kcal/mol, below the estimated standard deviations for any parameters. From these results, we concluded that the current parameter set is reasonably

robust.

The NN parameters of all possible I-C and Watson-Crick pairs were all negative ΔG (favorable) values. For the case in which an inosine was at the 5'-end, the stabilities of the tandem pairs were ranked as follows: $\left(\frac{IC}{CG}\right) > \left(\frac{IG}{CC}\right) > \left(\frac{IA}{CU}\right) \approx \left(\frac{IU}{CA}\right)$. The difference between the parameters reflects the higher stability of guanine-cytosine (G-C) pairs compared with adenine-uracil (A-U) pairs, where the G-C pairs contain three hydrogen bonds between bases but the A-U pairs contain only two. The relative stability between the pairs with the same number of hydrogen bonds may reflect a more complex effect from the stacking. For the case in which an inosine was at the 3'-end, the stabilities of the tandem pairs were ranked as follows: $\left(\frac{CI}{GC}\right) > \left(\frac{GI}{CC}\right) > \left(\frac{AI}{UC}\right) \approx \left(\frac{UI}{AC}\right)$. This again reflects the number of hydrogen bonds. For tandem inosines, the trend was as follows: $\left(\frac{IC}{CI}\right) > \left(\frac{CC}{II}\right) > \left(\frac{CI}{IC}\right)$.

We also compared the NN parameters with those obtained from experimental results alone (third column of Table 3). The differences between the parameters were marginal, all less than 0.40 kcal/mol. This is as expected because we used larger errors for the computationally obtained $\Delta\Delta G$ values in the calculation of NN parameters.

3 Discussion

Comparison to the related NN parameters We compared the obtained NN parameters with the canonical NN parameters (Table 3). The fourth column (“I-C→G-C”) represents the NN parameters when I-C pairs were replaced by the G-C pairs (the column thus represents the NN param-

eters for $\left(\begin{smallmatrix} \text{GA} \\ \text{CU} \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} \text{GC} \\ \text{CG} \end{smallmatrix}\right)$, and so on). The NN parameters with I-C replaced by G-C (excluding any terminal bonus) had 1.17 kcal/mol lower free-energy differences on average compared with the I-C counterparts. The result is reasonable because each G-C pair has three hydrogen bonds, whereas each I-C pair has only two hydrogen bonds. The difference is thus considered to reflect the increased stability of G-C pairs with their additional hydrogen bond. The tendency is more prominent when parameters for only tandem I-C pairs are considered, where G-C containing pairs are 1.92 kcal/mol more stable than I-C counterparts. This is again reasonable because tandem G-C pairs have two additional hydrogen bonds compared with tandem I-C pairs. When I-C pairs were replaced by A-U pairs (“I-C→A-U”), the free-energy differences were comparable (0.08 kcal/mol in favor of I-C pairs). Again, this is considered to reflect the similar relative stability of A-U pairs (with two hydrogen bonds) and I-C pairs (also with two hydrogen bonds). The difference in the average parameters for tandem pairs was small as well (0.03 kcal/mol).

When the I-C pairs were replaced by I•U pairs, the free-energy difference of the duplex formation increased by 1.93 kcal/mol on average (Table 3, “I-C→I•U”). Even when the non-tandem NN parameters are considered, the NN parameters for I-C pairs are 1.19 kcal/mol more stable than the counterparts with I•U pairs. This result reflects the relative stability of Watson-Crick pairs and wobble pairs. Despite there being two hydrogen bonds between nucleotides in both pairs, wobble pairs are structurally disfavored compared with Watson-Crick pairs. The free-energy difference between tandem I-C pairs and tandem I•U pairs is 3.91 kcal/mol on average.

This large difference indicates that stacking interactions of tandem wobble I•U pairs in RNA duplexes are strongly disfavored compared with tandem Watson-Crick I-C pairs. This tendency is also found in the comparison of NN parameters between tandem A-U (i.e., Watson-Crick) pairs (Xia et al., 1998) and tandem G•U (i.e., wobble) (Mathews et al., 1999) pairs. Tandem A-U pairs were generally more stable than tandem G•U pairs despite both having two hydrogen bonds between pairs (Table 3, “I-C→A-U” and “I-C→G•U”), possibly due to the structural instability of consecutive wobble pairs.

I-C bonus at RNA termini In Table 3, ΔG_{37}° for a terminal I-C pair is -0.28 kcal/mol, indicating that there is a *bonus* when an I-C pair occurs at a terminus. This implies that I-C pairs at termini are more favored than G-C pairs. Indeed, the same tendency was also observed in the experimentally derived parameters (Table 3). Furthermore, terminal I•U wobble pairs have been reported (Wright et al., 2007) to be more stable than terminal G-C pairs (by -1.33 kcal/mol). Although the detailed mechanism is unclear, these results indicate that inosine-containing base pairs are favored on 5'- and 3'-ends compared with C-G pairs. It is possible that single-stranded inosine is unstable because of structural restraints, entropic contributions, and/or solvation effects. Another possibility is that the observed stability of a terminal inosine reflects biases in the sequence design. Experimentally measured sequences containing I-C and I•U pairs are very limited in variation. All sequences with terminal I-C or I•U pairs in Table 1 and Wright et al. (2007) were GC rich (i.e., > 50% of base pairs are G-C). Errors that

had accumulated due to this lack of variation may have caused this rather counterintuitive result. It has recently been shown that the accumulation of errors may sum up to unexpectedly large errors in the case of canonical NN parameters (Zuber et al., 2018). Limited variation in the sequences may have similar consequences in the case of both current I-C and I•U parameters.

Future Research and Conclusions Because the simulations inherently have a systematic bias from the experiments, the current NN parameters may be somewhat inaccurate. Ultimately, the parameters will be superseded by fully experimentally determined parameters. Meanwhile, the methodology of combining experiments with low-cost calculations should provide drop-in replacements for modified RNA parameters.

Recently, multiple RNA force fields (Aytenfisu et al., 2017; Kührová et al., 2018; Steinbrecher et al., 2012; Tan et al., 2018; Yang et al., 2017) have been proposed for the performance of precise RNA simulations. The main differences between the current RNA force field parameters and these new parameters are dihedral angle parameters (Aytenfisu et al., 2017) and van der Waals parameters (Steinbrecher et al., 2012; Tan et al., 2018; Yang et al., 2017). In the present research, because only the duplex-forming energy was considered, the difference in the dihedral angle parameters affects the stability of only single-stranded structures, whose contribution to the total free-energy difference may not be significant. The modification of the van der Waals parameters, on the other hand, may affect the stability of both single-stranded and duplex structures. These parameters may thus enable more accurate free-energy calculation and hence more accurate free-energy

parameters. In parallel to these approaches, an implicit-solvent force field has recently been used to successfully estimate free energy (Chou et al., 2016). Combining these methodologies may contribute to the accurate, low-cost determination of NN parameters for various modified RNAs.

In the present research, only RNA duplex formation parameters were derived. Although parameters for the duplex formation are essential for RNA structure prediction, several other parameters are also known to contribute to more accurate RNA structure prediction. In particular, Zuber et al. (2017) recently reported that dangling and loop-initiation parameters strongly affect structural prediction accuracy for canonical RNAs. Combined with a recent report on the computational derivation of $\Delta\Delta G$ of hairpins (Smith et al., 2018), methodologies for deriving non-stem parameters would be a valuable addition to the current parameter derivation scheme. Additionally, in the present research, the enthalpic and entropic terms were considered in only the experiments, not the simulations. Hence, the derived parameters cover ΔG s at only 37°C. Deriving parameters for a wider range of temperatures would enable prediction of structures at different temperatures. Additionally, recent research (Spasic et al., 2018) has suggested that using a multiple-state model improves the fit to UV-melting experiments. Adjusting the current methodology to fit such cases will be beneficial.

Finally, combined with the I•U wobble pair parameters by Wright et al. (2007), it is now possible to predict the secondary structure of inosine-containing sequences. The development of bioinformatic tools that enable structure prediction of inosine-containing sequences is important from a practical perspective.

In conclusion, we derived the NN parameters for RNAs containing I-C pairs by combining experiments and simulation. The derived parameters are consistent with the existing NN parameters for canonical bases. The current work also exemplifies a methodology for combining experiments with calculations in order to derive parameters for various modified RNAs.

4 Materials and Methods

4.1 NN model

In the NN model, the standard free-energy difference ΔG° of forming a duplex is approximated by the sum of the energies for duplex initiation and those for two base pairs. For example, the standard free-energy difference for forming the duplex 5'-CAIA-3' and 5'-UCUG-3' at 37°C is approximated as

$$\Delta G_{37}^\circ \left(\begin{array}{c} \text{CAIA} \\ \text{GUCU} \end{array} \right) \approx \Delta G_{37,\text{init}}^\circ + \Delta G_{37}^\circ \left(\begin{array}{c} \text{CA} \\ \text{GU} \end{array} \right) + \Delta G_{37}^\circ \left(\begin{array}{c} \text{AI} \\ \text{UC} \end{array} \right) + \Delta G_{37}^\circ \left(\begin{array}{c} \text{IA} \\ \text{CU} \end{array} \right) + \Delta G_{37,\text{termAU}}^\circ. \quad (1)$$

where $\Delta G_{37,\text{init}}^\circ$ is the duplex initiation energy and $\Delta G_{37}^\circ \left(\begin{array}{c} \text{XY} \\ \text{ZW} \end{array} \right)$ is the free-energy difference for each of the two neighboring base pairs. $\Delta G_{37,\text{termAU}}^\circ$ represents the penalty for each A-U pair appearing at the 5'- and/or 3'-ends. The free-energy difference for each two neighboring base pairs as well as penalties or bonuses at termini are represented by the NN parameters. The NN parameters are widely adapted as a fundamental parameter to estimate the secondary structure of RNAs (Hofacker, 2014; Seetin and Mathews, 2012). In this research, we only consider Watson-Crick type or wobble base

pairs. Furthermore, unless otherwise stated, we only consider sequences at 37°C (i.e., 310K), and the temperature subscripts $_{37}$ may be omitted from the notation.

As explained in the Introduction, our goal in this research is the derivation of the NN parameters for duplexes containing inosine-cytosine pairs. The NN parameters were typically determined from UV melting experiments; as also explained in the Introduction, this research has combined UV melting experiments with molecular simulations to estimate the parameters. Figure 1 represents the overview of approaches to determine the NN parameters. In addition to experimentally determined duplex-forming energies, we obtain the relative free-energy differences to supplement the parametrization.

4.2 UV adsorption measurements

The free-energy difference upon duplex formation for inosine-containing sequences were measured using UV adsorption experiments.

Sequence design The sequences of inosine-containing RNAs were determined to meet the following criteria: (1) some NN parameters can be derived directly from experimental results, (2) self-pairing or staggered pairing is prevented, and (3) duplexes are stable enough to allow measurements around room temperature. We used `DuplexFold` in `RNAstructure` (Reuter and Mathews, 2010) to estimate the stability of RNA pairs, where we replaced inosine with guanine during the prediction. Table 1 lists the designed sequences.

Preparation of oligo RNAs Solid phase synthesis of oligo RNAs was performed with an automatic DNA/RNA synthesizer (NTS M-2-MX, NIHON TECHNO Co., Ltd., Japan) using the following phosphoramidite monomers: A-TOM-CE Phosphoramidite, C-TOM-CE Phosphoramidite, G-TOM-CE Phosphoramidite, U-TOM-CE Phosphoramidite (10-3004, 10-3014, 10-3024, 10-3034, respectively; Glen Research, Sterling, VA), Inosine CED phosphoramidite (ANP-5680; ChemGenes, Wilmington, MA), and CPGs (Ac-C-RNA-CPG; 20-3315, Ac-G-RNA-CPG; 20-3324, Glen Research). For oligo RNAs with an inosine at the 3' end, Universal UnyLinker Support (2000Å, N-4000-20, ChemGenes) was used as a solid support. After solid phase synthesis, work-up procedures were performed as follows; the CPGs were washed with CH₃CN, dried with flowing air, treated with a mixed solution of triethylamine and CH₃CN (1:1, *v/v*), and kept at room temperature for 10 min. Then, the solution was removed and the CPGs were washed with CH₃CN. Cleavage of the synthesized oligo RNAs from the CPGs and deprotection were performed with a mixed solution of concentrated NH₄OH and 40% aqueous methylamine (1:1, *v/v*) at 35°C for 6 h or at room temperature for 12 h. The resulting RNA solution was filtered and transferred to another tube. This solution was concentrated to dryness using a centrifuge evaporator. The residue was treated with TBAF (1 M in THF), and the mixture was heated at 50°C for 10 min and vortexed. The mixture was then incubated at 35°C for > 6 h, and 1 M Tris-HCl buffer (pH 7.4) was added. Then, the volatiles were removed by a centrifuge evaporator. The resulting solution was desalted with a NAP-25 column (GE Healthcare, Chicago, IL). The collected fractions were concentrated with a

centrifuge evaporator.

HPLC analysis and purification of oligo RNAs HPLC purification and analysis were performed with a LaChrom Elite HPLC System (L-2455, L-2130, L-2485, HITACHI Co., Japan) equipped with a YMC HPLC column (Hydrosphere C18, 4.6×250 mm for analysis, 10×250 mm for purification, YMC Co., Ltd, Kyoto, Japan) with flow rates of 1 mL/min for analysis and 3 mL/min for purification. HPLC was conducted at room temperature and with a monitoring wavelength of 250 nm. The mobile phase consisted of solutions A (50 mM TEAA (pH 7.0) containing 5% CH_3CN) and B (CH_3CN), with a linear gradient of A/B = 100 : 0 to 70 : 30 used over 20 min. The collected fractions were lyophilized, and the resulting RNAs were re-dissolved in Milli-Q water (MilliporeSigma, Burlington, MA) and lyophilized by freeze drying once again.

MALDI-TOF MS analysis of oligo RNAs The obtained RNA samples were characterized by MALDI-TOF MS (Ultraflex III, Bruker Daltonics, Billerica, MA) with a 3-hydroxypicolinic acid matrix. Calculated and observed MS values are summarized in Table S1.

T_m measurements The concentration of RNA stock solution (Milli-Q) was estimated by UV absorbance at 260 nm at room temperature using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA USA). The molar extinction coefficient of oligo RNAs was calculated using Oligo Analyzer 3.1 (Integrated DNA Technologies, Coralville, IA), an online software tool. Equal moles of complementary single strands of RNA were mixed to form

the duplex. This solution was lyophilized, and then the RNAs were redissolved in 20 mM sodium cacodylate, 1 M NaCl, and 0.5 mM Na₂EDTA (pH 7.0). The T_m values of the samples were recorded on a spectrophotometer (V-650; JASCO, Tokyo, Japan) equipped with an 8-microcell changer and Peltier temperature controller system using a heating rate of 2°C/min from 10°C to 90°C to obtain curves of absorbance at 260 nm versus temperature. Spectrum manager software (JASCO) was used to fit melting curves to a two-state model to calculate T_m values. T_m values at different concentrations (2, 4, 6, 8, 10, 50, 100, 120, and 150 μ M) were used to calculate thermodynamic parameters.

Data analysis The duplex-forming free-energy difference was estimated from the regression between the concentration of the RNA (C_t) and the melting temperature (T_m). We followed the procedure detailed by Xia et al. (1998) to derive ΔG_{37}° , ΔH° , and ΔS° upon the duplex formation. We calculated the parameters from three independent measurements. Averages and the standard deviations are listed in Table 1.

4.3 Free-energy calculation

Although the direct calculation of free-energy differences upon duplex formation is possible (Deng and Roux, 2009; Tan et al., 2018; Wang et al., 2006), such calculation often requires considerable computational resources and/or dedicated hardware to achieve an accuracy comparable to that based on experiments. This is because there are large structural changes between the two end states (i.e., the duplex form and the unbound form),

and the various conformational changes from two states should be sampled to achieve better convergence. Instead, we considered the calculation of the *relative* free-energy difference. Here, we considered the relative free-energy difference with respect to the mutations to nucleotide bases, that is, the *difference* in duplex-forming energies between two RNA duplexes $\Delta\Delta G_{AB} = \Delta G_B - \Delta G_A$. This kind of free-energy difference causes smaller conformational changes, and thus the resulting relative free-energy difference converges faster (Krepl et al., 2013; Sakuraba et al., 2015). Similar approaches have also been demonstrated to be computationally efficient in analyses of protein-drug complexes (Wang et al., 2015). Furthermore, we have shown that such relative free-energy differences can also be used to estimate NN parameters in combination with experiments (Nishida et al., 2018). The relative binding free-energy changes were calculated for various inosine-containing RNA duplexes using the free-energy perturbation method.

Sequence design We chose pairs of inosine-containing duplexes so as to meet the following goals: (a) the pairs cover a variety of the NN parameters containing inosine, (b) the mutations between duplexes do not contain unnecessary canonical NN changes, (c) all duplexes are stable enough to hold a duplex form at 37°C, (d) the sequence does not contain excessive repetition, (e) there are only two or fewer base pair mutations, (f) termini are paired as G-C or I-C (for stability), and (g) sequence lengths are 6 or 7 bases. We optimized the pairs using in-house software to meet the criteria above. During the design process, we used the Vienna RNA package (Lorenz et al., 2011)

to predict the stability of various duplexes, with inosine bases replaced by guanine bases during the prediction.

System setup The relative free-energy differences were calculated with the free-energy perturbation method. We used the same protocol as described by Sakuraba et al. (2015) to calculate the free-energy difference; for the detailed procedure, please consult this previous study. Parameters associated with atoms (van der Waals radii and charges) were gradually changed between two end states (pairs A and B) to compute the free-energy difference. The 1-1-48 soft-core potential function (Pham and Shirts, 2011) was used to remove numerical instability during the calculation. Both charges and the van der Waals radii were treated with the soft-core potential.

For the 30 duplex pairs listed in Table 2, all-atom, explicit-water models were constructed. The AMBER14SB (AMBER99 + bsc0 + χ OL3 (Cheatham III et al., 1999; Perez et al., 2007; Zgarbová et al., 2011) for nucleic acids) force field and a monovalent cation force field developed by Joung and Cheatham (Joung and Cheatham, 2008) were used to model the system. Water molecules were modeled using TIP3P, and the system was solvated with 1 M NaCl solution. The atomic charges of an inosine molecule were determined by the RESP method (Bayly et al., 1993; Cieplak et al., 1995). The dihedral angle parameter with respect to the χ angle was set equal to that from the adenine molecule in a χ OL3 force field (Zgarbová et al., 2011). The total number of atoms in the system was between 13,741 and 16,481. For each pair of duplexes, the free-energy difference with respect to the mutation was calculated using the weighted histogram analysis method (multistate

Bennett acceptance ratio method) (Kumar et al., 1993; Shirts and Chodera, 2008; Souaille and Roux, 2001). The Hamiltonian replica exchange method (Fukunishi et al., 2002) was used to sample various conformations between two end states (28 replicas in total). Free-energy differences were calculated for the duplex state and two single-stranded states, and the total relative free energy difference was calculated according to the thermodynamic cycle in Fig. 4. For each state, the production run was performed for 20 ns (20 ns \times 28 replicas \times 3 states \approx 1.7 μ s for each pair of duplexes). For the duplex states of pairs 17, 19, 20, 22, and 25, we restrained the termini with I-C (and corresponding termini at the other state) to keep Watson-Crick type pairs because inosine bases at termini are unstable at 37°C and often result in a dissociation between the two strands. The restraint was realized by applying a potential that includes a penalty if the distance between N1 of inosine and N3 of cytosine exceeds 4 Å, with a force constant of 10^4 kJ/mol/nm².

4.4 NN parameter estimation

We derived the NN parameters using both the experimentally derived free-energy differences and the computationally derived relative free-energy differences. We used a method described by Nishida et al. (2018) to combine experimental and computational results. Here, we briefly describe the method to derive the NN parameters.

Linear least square model With the NN model, the RNA duplex-forming energy can be approximated as the matrix-vector product

$$\begin{pmatrix} \Delta \mathbf{G}_{\text{exp}}^{\circ} \\ \Delta \Delta \mathbf{G}_{\text{calc}}^{\circ} \end{pmatrix} \approx \begin{pmatrix} S_{\text{exp}} \\ S_{\text{calc,B}} - S_{\text{calc,A}} \end{pmatrix} \Delta \mathbf{G}_{\text{NN}}, \quad (2)$$

where $\Delta \mathbf{G}_{\text{exp}}^{\circ}$ is a vector representing experimentally measured ΔG° values, $\Delta \Delta \mathbf{G}_{\text{calc}}^{\circ}$ is a vector representing calculated $\Delta \Delta G^{\circ}$ values, and $\Delta \mathbf{G}_{\text{NN}}$ is a vector representing the NN parameters we wish to obtain. S_{exp} and S_{calc} are matrices constructed as follows: each row of S_{exp} represents the number of occurrences of NNs in the duplex, whereas each row of $S_{\text{calc,A}}$ and $S_{\text{calc,B}}$ represents the number of occurrences in sequences before and after the mutation, respectively. In this research, we only consider the NN parameters containing I/C pairs. This can be achieved by subtracting offset energies based on the occurrences of all the canonical NN pairs from $\Delta \mathbf{G}_{\text{exp}}^{\circ}$ and $\Delta \Delta \mathbf{G}_{\text{calc}}^{\circ}$.

From Eq. (2), the NN parameters are determined by solving the least square fitting problem. The parameters are readily obtained as

$$\Delta \mathbf{G}_{\text{NN}} = \left[\Sigma^{-1} \begin{pmatrix} S_{\text{exp}} \\ S_{\text{calc,B}} - S_{\text{calc,A}} \end{pmatrix} \right]^{\dagger} \Sigma^{-1} \begin{pmatrix} \Delta \mathbf{G}_{\text{exp}}^{\circ} \\ \Delta \Delta \mathbf{G}_{\text{calc}}^{\circ} \end{pmatrix}, \quad (3)$$

where M^{\dagger} is the Moore-Penrose pseudoinverse of the matrix M , and Σ is a diagonal matrix whose diagonal elements are standard deviations of all experiments and calculations combined. In this work, we assumed that all experiments and all calculations are independent and that their errors are not

correlated. Expected standard deviations can be derived similarly (Nishida et al., 2018) and are presented alongside the parameters in Table 3.

Derivation of parameters The simulation inherently has a *systematic* bias compared with the experimental counterpart. Although both the simulation and experiments have biases compared with the “true” values of Gibbs free energy (or differences in Gibbs free energy), the bias from the simulation is considered to be much larger than that from the UV melting experiment. Consistent with the case of the canonical NN parameter, in this research, we consider that the values obtained from the UV melting experiments represent the “true” values of the Gibbs free energy. We thus consider that the values from UV experiments have only statistical errors, while the values from simulations have both statistical errors and systematic biases. In our previous work (Sakuraba et al., 2015), we observed a mean absolute deviation of 0.55 kcal/mol in the simulation of canonical bases. Based on this result, to compensate for the systematic biases, we offset the estimated statistical errors from the free-energy simulation by 0.50 kcal/mol. This protocol adjusts the NN parameters to favor experimentally measured free energies, reflecting the uncertainty arising from the systematic bias in the computationally derived free energies. For a comparison, we also derived the NN parameters from the simulation-derived free energies only. Derived parameters are presented in Table SII. Derived parameters indicated a slight deviation from the experimental values as presented in Fig. SI, therefore, we concluded that the inclusion of experimental parameters is necessary for the accurate parameter derivation.

Acknowledgement

The computations were performed using ACCMS, Kyoto University, Japan, and the Research Center for Computational Science, Okazaki, Japan. This work was supported by JSPS KAKENHI Grant Numbers 16H02484 to SS, MH, TK, HA, and KA, 16K17778 to SS, and 16K16143 to JI, and by a MEXT Grant-in-Aid for Scientific Research on Innovative Areas (16H06279) to KA.

Supporting Information

Characterization data for the synthesized inosine oligo RNAs, NN parameters from only simulations, and parameter files we used for inosine simulation are available in the supporting information.

References

- Aytenfisu AH, Spasic A, Grossfield A, Stern HA, Mathews DH. 2017. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J Chem Theory Comput*, **13**: 900–915.
- Bayly CI, Cieplak P, Cornell W, Kollman PA. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*, **97**: 10269–10280.
- Boccaletto P, Machnicka MA, Purta E, Piatkowski P, Baginski B, Wirecki TK, de Crecy-Lagard V, Ross R, Limbach PA, Kotter A et al.

2018. MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res*, **46**: D303–D307.
- Chan PP, Lowe TM. 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res*, **37**: D93–97.
- Cheatham III TE, Cieplak P, Kollman PA. 1999. A modified version of the cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn*, **16**: 845–862.
- Chou FC, Kladwang W, Kappel K, Das R. 2016. Blind tests of RNA nearest-neighbor energy prediction. *Proc Natl Acad Sci U S A*, **113**: 8430–8435.
- Cieplak P, Cornell WD, Bayly C, Kollman PA. 1995. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J Comput Chem*, **16**: 1357–1377.
- Crick F. 1966. Codon-anticodon pairing: The wobble hypothesis. *J Mol Biol*, **19**: 548–555.
- Deng Y, Roux B. 2009. Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B*, **113**: 2234–2246.
- Fukunishi H, Watanabe O, Takada S. 2002. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J Chem Phys*, **116**: 9058–9067.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res*, **31**: 439–441.
- Hofacker IL. 2014. Energy-directed RNA structure prediction. *Methods Mol Biol*, **1097**: 71–84.

- Joung IS, Cheatham TE. 2008. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B*, **112**: 9020–9041.
- Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, **42**: 68–73.
- Krepl M, Otyepka M, Banáš P, Sponer J. 2013. Effect of guanine to inosine substitution on stability of canonical DNA and RNA duplexes: molecular dynamics thermodynamics integration study. *J Phys Chem B*, **117**: 1872–1879.
- Kührová P, Mlynsky V, Zgarbova M, Krepl M, Bussi G, Best RB, Otyepka M, Sponer J, Banas P. 2018. Improving the performance of the RNA AMBER force field by tuning the hydrogen-bonding interactions. *bioRxiv*, doi: 10.1101/410993.
- Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. 1993. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem*, **13**: 1011–1021.
- Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**: 26.
- Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, **288**: 911–940.
- Nishida S, Sakuraba S, Asai K, Hamada M. 2018. Estimating energy parameters for RNA secondary structure predictions using both experimental and computational data. *IEEE/ACM Trans Comput Biol Bioinform*

matics, in press. Doi: 10.1109/TCBB.2018.2813388.

- Nishikura K. 2010. Functions and regulation of rna editing by adar deaminases. *Annu Rev Biochem*, **79**: 321–349.
- Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, Orozco M. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, **92**: 3817–3829.
- Pham TT, Shirts MR. 2011. Identifying low variance pathways for free energy calculations of molecular transformations in solution phase. *J Chem Phys*, **135**: 034114.
- Ramaswami G, Li JB. 2014. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res*, **42**: D109–113.
- Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**: 129.
- Sakuraba S, Asai K, Kameda T. 2015. Predicting RNA Duplex Dimerization Free-Energy Changes upon Mutations Using Molecular Dynamics Simulations. *J Phys Chem Lett*, **6**: 4348–4351.
- Sakurai M, Ueda H, Yano T, Okada S, Terajima H, Mitsuyama T, Toyoda A, Fujiyama A, Kawabata H, Suzuki T 2014. A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res*, **24**: 522–534.
- Seetin MG, Mathews DH. 2012. RNA structure prediction: an overview of methods. *Methods Mol Biol*, **905**: 99–122.
- Shirts MR, Chodera JD. 2008. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*, **129**: 124105.

- Smith LG, Tan Z, Spasic A, Dutta D, Salas-Estrada LA, Grossfield A, Mathews D. 2018. Chemically accurate relative folding stability of rna hairpins from molecular simulations. *bioRxiv*, doi: 10.1101/354332.
- Souaille M, Roux B. 2001. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput Phys Commun*, **135**: 40–57.
- Spasic A, Berger KD, Chen JL, Seetin MG, Turner DH, Mathews DH. 2018. Improving rna nearest neighbor parameters for helices by going beyond the two-state model. *Nucleic Acids Res*, **46**: 4883–4892.
- Steinbrecher T, Latzer J, Case DA. 2012. Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput*, **8**: 4405–4412.
- Tan D, Piana S, Dirks RM, Shaw DE. 2018. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci U S A*, **115**: E1346–E1355.
- Tan MH, Li Q, Shanmugam R, Piskol R, Kohler J, Young AN, Liu KI, Zhang R, Ramaswami G, Ariyoshi K et al. 2017. Dynamic landscape and regulation of RNA editing in mammals. *Nature*, **550**: 249–254.
- Wang J, Deng Y, Roux B. 2006. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J*, **91**: 2798–2814.
- Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J et al. 2015. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc*, **137**: 2695–2703.

- Wright DJ, Rice JL, Yanker DM, Znosko BM. 2007. Nearest neighbor parameters for inosine–uridine pairs in RNA duplexes. *Biochemistry*, **46**: 4625–4634.
- Xia T, SantaLucia J, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, **37**: 14719–14735.
- Yang C, Lim M, Kim E, Pak Y. 2017. Predicting rna structures via a simple van der waals correction to an all-atom force field. *J Chem Theory Comput*, **13**: 395–399.
- Zgarbová M, Otyepka M, Šponer J, Mladek A, Banas P, Cheatham TE, Jurecka P. 2011. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput*, **7**: 2886–2902.
- Zuber J, Cabral BJ, McFadyen I, Mauger DM, Mathews DH. 2018. Analysis of RNA nearest neighbor parameters reveals interdependencies and quantifies the uncertainty in RNA secondary structure prediction. *RNA*, *in press*. Doi: 10.1261/rna.065102.117.
- Zuber J, Sun H, Zhang X, McFadyen I, Mathews DH. 2017. A sensitivity analysis of RNA folding nearest neighbor parameters identifies a subset of free energy parameters with the greatest impact on RNA secondary structure prediction. *Nucleic Acids Res*, **45**: 6168–6176.

List of Figures

1	(A) Structure of the Watson-Crick pairs (G-C and A-U) and an I-C pair. (B) Schematic overview of the nearest-neighbor (NN) parameter derivation in this research. We combine the results from UV melting experiments with those from molecular dynamics simulations to obtain the NN parameters. . . .	28
2	Comparison of nearest-neighbor (NN) model energies, and experimentally derived ΔG values or computationally derived $\Delta\Delta G$ values. The sign of $\Delta\Delta G$ values can be arbitrary inverted by swapping sequences A and B; we thus normalized the results to ensure $\Delta\Delta G < 0$	29
3	Comparison between the experimentally measured ΔG value and the prediction from a nearest-neighbor (NN) model that is determined by excluding a corresponding single experiment.	30
4	The thermodynamic cycle considered in this research. From UV melting experiment ΔG values were directly measured, while from the molecular dynamics simulations, ΔG values corresponding to the mutation of $A \rightarrow B$ were calculated as $\Delta\Delta G_{AB}$	31

Figure 1: (A) Structure of the Watson-Crick pairs (G-C and A-U) and an I-C pair. (B) Schematic overview of the nearest-neighbor (NN) parameter derivation in this research. We combine the results from UV melting experiments with those from molecular dynamics simulations to obtain the NN parameters.

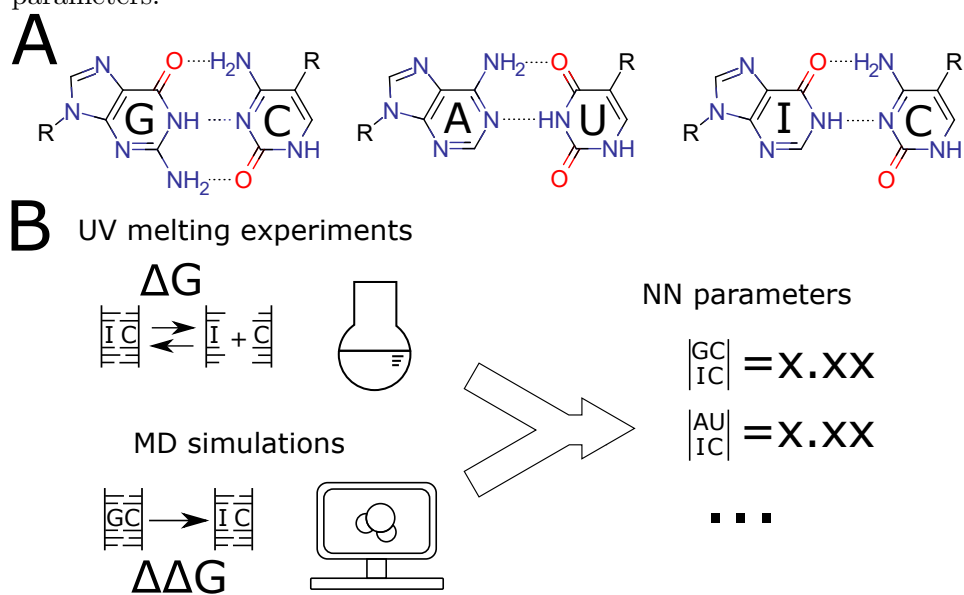


Figure 2: Comparison of nearest-neighbor (NN) model energies, and experimentally derived ΔG values or computationally derived $\Delta\Delta G$ values. The sign of $\Delta\Delta G$ values can be arbitrary inverted by swapping sequences A and B; we thus normalized the results to ensure $\Delta\Delta G < 0$.

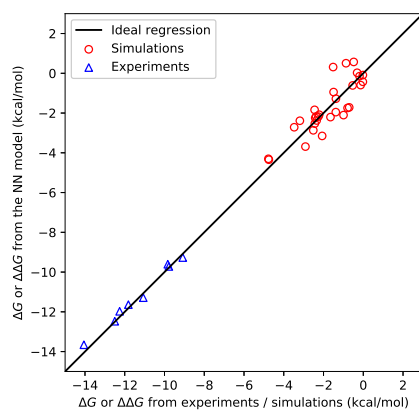


Figure 3: Comparison between the experimentally measured ΔG value and the prediction from a nearest-neighbor (NN) model that is determined by excluding a corresponding single experiment.

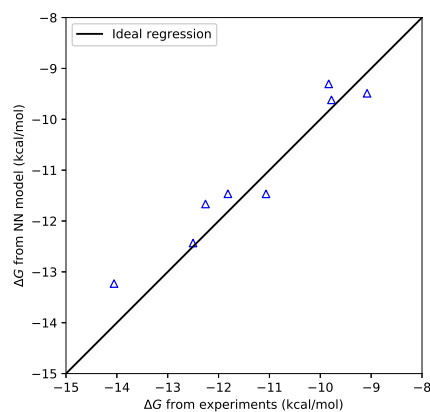
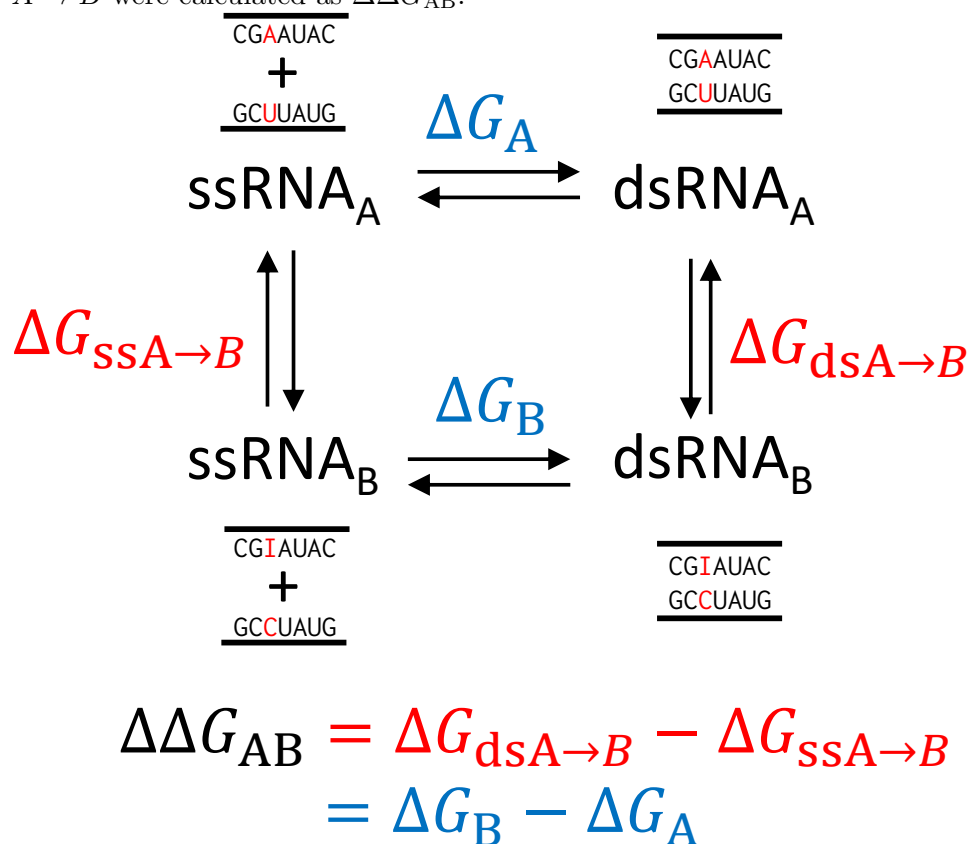


Figure 4: The thermodynamic cycle considered in this research. From UV melting experiment ΔG values were directly measured, while from the molecular dynamics simulations, ΔG values corresponding to the mutation of $A \rightarrow B$ were calculated as $\Delta\Delta G_{AB}$.



List of Tables

1	Experimentally measured RNA duplex-forming free energies.	33
2	The relative difference in duplex-forming energy $\Delta\Delta G_{37}^{\circ} = \Delta G_{37,B}^{\circ} - \Delta G_{37,A}^{\circ}$ obtained from free-energy calculations. . . .	34
3	Derived NN parameters for pairs containing inosine and corresponding NN parameters when inosine-cytosine pairs are replaced with other Watson-Crick or wobble pairs. Units are kcal mol ⁻¹	35

Table 1: Experimentally measured RNA duplex-forming free energies.

Sequences	$-\Delta G_{37}^{\circ}$ (kcal/mol)	$-\Delta H^{\circ}$ (kcal/mol)	$-\Delta S^{\circ}$ (cal/mol/K)
(GCAIAGC CGUCUGG)	9.08 ± 0.04	63.79 ± 0.68	176.41 ± 2.08
(GCCICGC CGGCGCG)	14.06 ± 0.06	88.41 ± 0.82	239.72 ± 2.43
(GCAICGC CGUCGCG)	11.07 ± 0.17	73.97 ± 2.53	202.81 ± 7.61
(GCCIAGC CGGCUUG)	11.82 ± 0.02	81.84 ± 0.10	225.79 ± 0.27
(CAGGCCI GUGCGGC)	12.26 ± 0.15	73.62 ± 1.89	197.83 ± 5.60
(CAGCGAI GUGCGUC)	9.83 ± 0.14	66.25 ± 3.00	181.91 ± 9.21
(ICCACGC CGGUGCG)	12.50 ± 0.04	76.30 ± 0.49	205.70 ± 1.44
(IACACGC CUGUGCG)	9.78 ± 0.00	68.39 ± 0.12	188.96 ± 0.38

Table 2: The relative difference in duplex-forming energy $\Delta\Delta G_{37}^{\circ} = \Delta G_{37,B}^{\circ} - \Delta G_{37,A}^{\circ}$ obtained from free-energy calculations.

No.	Sequence A	Sequence B	$\Delta\Delta G_{37}^{\circ}$ (kcal/mol)
1	(CGAAUAC)	(CGIAUAC)	0.020 ± 0.241
2	(GCUUAUG)	(GCCUAUG)	1.650 ± 0.443
3	(CGGCIGC)	(CGAUIGC)	-0.868 ± 0.284
4	(GCCICCG)	(GCUACCG)	2.346 ± 0.230
5	(CICGAUG)	(CICAAUG)	-2.449 ± 0.555
6	(GCI CUAC)	(GCGUUAU)	-2.217 ± 0.099
7	(CIGIAAC)	(CIIIAAC)	3.475 ± 0.322
8	(GCCCUUG)	(GCCCUUG)	-0.481 ± 0.169
9	(CUCACAG)	(CUCGUAG)	4.773 ± 0.248
10	(GAGUIUC)	(GAGCAUC)	4.752 ± 0.234
11	(CUIIUC)	(CUGIUC)	0.133 ± 0.656
12	(GACCAGG)	(GACCAGG)	-2.259 ± 0.206
13	(GAUCACC)	(GAUCIUC)	-2.370 ± 0.532
14	(CUAIUGG)	(CUAICAG)	2.514 ± 0.116
15	(GAUCCGC)	(GAUCCAC)	0.994 ± 0.170
16	(CUAGICG)	(CUAGGUG)	0.310 ± 0.194
17	(GCAGCCG)	(GCAIUCG)	0.708 ± 0.278
18	(CGUCGGC)	(CGUCAGC)	0.021 ± 0.394
19	(GCCACAG)	(GUCACAG)	1.495 ± 0.106
20	(CGGUGUC)	(CAIUGUC)	1.516 ± 0.129
21	(GCCUCCG)	(GCCUUUG)	1.382 ± 0.235
22	(CGGAIIC)	(CGGAAAC)	-0.790 ± 0.254
23	(GCIIAC)	(GCIGUAC)	-3.191 ± 0.397
24	(CGCCTUG)	(CGCCAUG)	-2.069 ± 0.231
25	(GCIUUGC)	(GCACUGC)	0.535 ± 0.184
26	(CGCAACG)	(CGUGACG)	2.438 ± 0.092
27	(GGCUUAC)	(GACUUAC)	-0.190 ± 0.384
28	(CCIAAUG)	(CUIAAUG)	1.377 ± 0.478
29	(GICUAAC)	(GICUAAC)	-2.418 ± 0.062
30	(CCGAUUG)	(CCIAUUG)	-2.910 ± 0.518
31	(GIGUUAG)	(GAGUUAG)	
32	(CCCAAUC)	(CUCAAUC)	
33	(ICACCUG)	(GUACCUG)	
34	(CGUGGAC)	(CAUGGAC)	
35	(CGAAIC)	(CGAIAC)	
36	(GCUUCG)	(GCUCUG)	
37	(CGAGGG)	(CGAGGI)	
38	(GCUCC)	(GCUCC)	
39	(CGGGAG)	(CGGGAG)	
40	(GCCUC)	(ICCCUC)	
41	(CUCUCG)	(CCCUCG)	
42	(GAGAGC)	(GIIAGC)	
43	(CUGGUG)	(CUGGCI)	
44	(GACCAC)	(GACCCG)	
45	(CUIAGC)	(CUAGGC)	
46	(GACUCG)	(GAUCCG)	
47	(GACIGC)	(GACIGC)	
48	(CUICCG)	(CUGCCG)	
49	(GAUGCC)	(GAUGCC)	
50	(CUACGG)	(CUACGI)	
51	(GCGGUG)	(GCGIUG)	
52	(CGCCAC)	(CGCCAC)	
53	(GGUIAG)	(GGUIIG)	
54	(CCACUC)	(CCACCC)	
55	(GICUIG)	(GIUCIG)	
56	(CCIACC)	(CCAICC)	
57	(GIIAUG)	(GGIAUG)	
58	(CCCUAC)	(CCCUAC)	
59	(GUCAGG)	(GUCGGG)	
60	(CAIUCC)	(CAGCCC)	

Table 3: Derived NN parameters for pairs containing inosine and corresponding NN parameters when inosine-cytosine pairs are replaced with other Watson-Crick or wobble pairs. Units are kcal mol⁻¹.

Pairs	$\Delta G_{37,\text{exp+comp}}^{\circ}$	$\Delta G_{37,\text{exp}}^{\circ}$	I-C→G-C ^a	I-C→A-U ^a	I-C→I•U ^b	I-C→G•U ^c
(IA CU)	-1.16 ± 0.19	-1.05 ± 0.21	-2.35 ± 0.06	-0.93 ± 0.03	0.43 ± 0.44	-1.27 ± 0.28
(IC CG)	-2.89 ± 0.19	-2.84 ± 0.21	-3.42 ± 0.08	-2.24 ± 0.06	-1.03 ± 0.30	-2.51 ± 0.25
(IG CC)	-2.37 ± 0.36	-	-3.26 ± 0.07	-2.08 ± 0.06	-1.22 ± 0.37	-2.11 ± 0.25
(IU CA)	-0.94 ± 0.42	-	-2.24 ± 0.06	-1.10 ± 0.08	-0.50 ± 0.44	-1.36 ± 0.24
(AI UC)	-1.17 ± 0.20	-1.13 ± 0.21	-2.08 ± 0.06	-0.93 ± 0.03	-0.41 ± 0.47	-0.55 ± 0.32
(CI GC)	-2.39 ± 0.19	-2.70 ± 0.21	-2.36 ± 0.09	-2.11 ± 0.07	-0.77 ± 0.39	-1.41 ± 0.24
(GI CC)	-2.03 ± 0.41	-	-3.26 ± 0.07	-2.35 ± 0.06	-1.34 ± 0.33	-1.53 ± 0.27
(UI AC)	-1.05 ± 0.41	-	-2.11 ± 0.07	-1.33 ± 0.09	0.37 ± 0.39	-1.00 ± 0.30
(IC CI)	-1.70 ± 0.55	-	-3.42 ± 0.08	-1.10 ± 0.08	3.58 ± 1.09	1.29 ± 0.56
(CI IC)	-0.55 ± 0.70	-	-2.36 ± 0.09	-1.33 ± 0.09	2.23 ± 0.91	0.30 ± 0.48
(CC II)	-1.01 ± 0.31	-	-3.26 ± 0.07	-0.93 ± 0.03	2.66 ± 0.88	0.47 ± 0.96
Terminal I-C	-0.28 ± 0.16	-0.41 ± 0.17	0.00 ± 0.00	0.45 ± 0.04	-1.33 ± 0.25	0.45 ± 0.04 ^d
Avg. non-tandem	-1.75	-	-2.64	-1.63	-0.56	-1.47
Avg. tandem	-1.09	-	-3.01	-1.12	2.82	0.69
Avg. all NN ^e	-1.57	-	-2.74	-1.49	0.36	-0.88

^a Xia et al. (1998). ^b Wright et al. (2007). ^c Mathews et al. (1999). ^d Following Mathews et al. (1999), the free energy parameter for each terminal G•U is assumed to be the same as the terminal A-U parameter. ^e Does not include a terminus bonus or penalty. NN, nearest neighbor.