

# Blind tests of RNA nearest-neighbor energy prediction

Fang-Chieh Chou<sup>a</sup>, Wipapat Kladwang<sup>a</sup>, Kalli Kappel<sup>b</sup>, and Rhiju Das<sup>a,b,c,1</sup>

<sup>a</sup>Department of Biochemistry, Stanford University, Stanford, CA 94305; <sup>b</sup>Biophysics Program, Stanford University, Stanford, CA 94305; and <sup>c</sup>Department of Physics, Stanford University, Stanford, CA 94305

Edited by David Baker, University of Washington, Seattle, WA, and approved May 31, 2016 (received for review November 25, 2015)

The predictive modeling and design of biologically active RNA molecules requires understanding the energetic balance among their basic components. Rapid developments in computer simulation promise increasingly accurate recovery of RNA's nearest-neighbor (NN) free-energy parameters, but these methods have not been tested in predictive trials or on nonstandard nucleotides. Here, we present, to our knowledge, the first such tests through a RECCES–Rosetta (reweighting of energy-function collection with conformational ensemble sampling in Rosetta) framework that rigorously models conformational entropy, predicts previously unmeasured NN parameters, and estimates these values' systematic uncertainties. RECCES–Rosetta recovers the 10 NN parameters for Watson–Crick stacked base pairs and 32 single-nucleotide dangling-end parameters with unprecedented accuracies: rmsd of 0.28 kcal/mol and 0.41 kcal/mol, respectively. For set-aside test sets, RECCES–Rosetta gives rmsd values of 0.32 kcal/mol on eight stacked pairs involving G–U wobble pairs and 0.99 kcal/mol on seven stacked pairs involving nonstandard isocytidine–isoguanosine pairs. To more rigorously assess RECCES–Rosetta, we carried out four blind predictions for stacked pairs involving 2,6-diaminopurine–U pairs, which achieved 0.64 kcal/mol rmsd accuracy when tested by subsequent experiments. Overall, these results establish that computational methods can now blindly predict energetics of basic RNA motifs, including chemically modified variants, with consistently better than 1 kcal/mol accuracy. Systematic tests indicate that resolving the remaining discrepancies will require energy function improvements beyond simply reweighting component terms, and we propose further blind trials to test such efforts.

RNA helix | ensemble prediction | simulated tempering | thermodynamics | blind prediction

RNA plays central roles in biological processes, including translation, splicing, regulation of genetic expression, and catalysis (1, 2), and in bioengineering efforts to control these processes (3–5). These critical RNA functions are defined at their most fundamental level by the energetics of how RNA folds and interacts with other RNAs and molecular partners, and how these processes change upon naturally occurring or artificially introduced chemical modifications. Experimentally, the folding free energies of RNA motifs can be precisely measured by optical melting experiments, and a compendium of these measurements have established the nearest-neighbor (NN) model for the most basic RNA elements, including double helices with the four canonical ribonucleotides (6). In the NN model, the stability of a base pair is assumed to only be affected by its adjacent base pairs, and the folding free energy of a canonical RNA helix can be estimated based on NN parameters for each stacked pair, an initialization term for the entropic cost of creating the first base pair, and corrections for different terminal base pairs. Although next-NN effects and tertiary contacts are not treated in the NN model (7–9), the current NN model gives accurate predictions for the folding free energies of canonical RNA helices (<0.5 kcal/mol for helices with 6–8 base pairs) (10, 11) and can be extended to single-nucleotide dangling ends, chemically modified nucleotides, and more complex motifs, such as noncanonical base pairs, hairpins, and internal loops (11–14). However, it is currently not feasible to experimentally characterize the energetics of all RNA motifs due to the large number of possible motif sequences and the

requirement of specialized experiments to address complex motif topologies, such as three-way junctions (15–17). These considerations, and the desire to test physical models of RNA folding, have motivated several groups to pursue automated computational methods to calculate the folding free energies of RNA motifs.

Current computational approaches are beginning to recover NN parameters for the simplest RNA motifs with accuracies within a few-fold of the errors of experimental approaches. For example, the Rosetta package has been developed and extensively tested for structure prediction and design of macromolecules, including RNA. Recent successes at near-atomic resolution have leveraged an all-atom “score function” that includes physics-based terms (for hydrogen bonding, van der Waals packing, and orientation-dependent implicit solvation) and knowledge-based terms (for, e.g., RNA torsional preferences) (18). When interpreting the total score as an effective energy for a conformation, simple Rosetta calculations recover the NN parameters for all canonical stacked pairs with an rmsd of less than 0.5 kcal/mol upon fitting two phenomenological parameters, the Rosetta energy scale and a constant offset parameterizing the conformational entropy loss upon folding each base pair (ref. 18 and see below). In parallel, molecular dynamics studies have demonstrated calculation of folding free energies of short RNA hairpins using umbrella sampling, molecular mechanics–Poisson Boltzmann surface area (MM–PB/SA), free energy perturbation, and other methods (19–22). Although these calculations have not yet accurately recovered folding free energies (errors > 10 kcal/mol) (21, 22), relative differences of NN parameters between different sequences and other aspects of RNA motif energetics have been recovered with accuracies between 0.6–1.8 kcal/mol (22–24). These error ranges are similar or lower than uncertainties of

## Significance

Understanding RNA machines and how their behavior can be modulated by chemical modification is increasingly recognized as an important biological and bioengineering problem, with continuing discoveries of riboswitches, mRNA regulons, CRISPR-guided editing complexes, and RNA enzymes. Computational strategies for understanding RNA energetics are being proposed, but have not yet faced rigorous tests. We describe a modeling strategy called RECCES–Rosetta (reweighting of energy-function collection with conformational ensemble sampling in Rosetta) that models the full ensemble of motions of RNA in single-stranded form and in helices, including nonstandard nucleotides, such as 2,6-diaminopurine, a variant of adenosine. When compared with experiments, including blind tests, the energetic accuracies of RECCES–Rosetta calculations are at levels close to experimental error, suggesting that computation can now be used to predict and design basic RNA energetics.

Author contributions: F.-C.C., W.K., K.K., and R.D. designed research; F.-C.C., W.K., and K.K. performed research; F.-C.C. contributed new reagents/analytic tools; F.-C.C. and K.K. analyzed data; and F.-C.C. and R.D. wrote the paper.

The authors declare no conflict of interest.

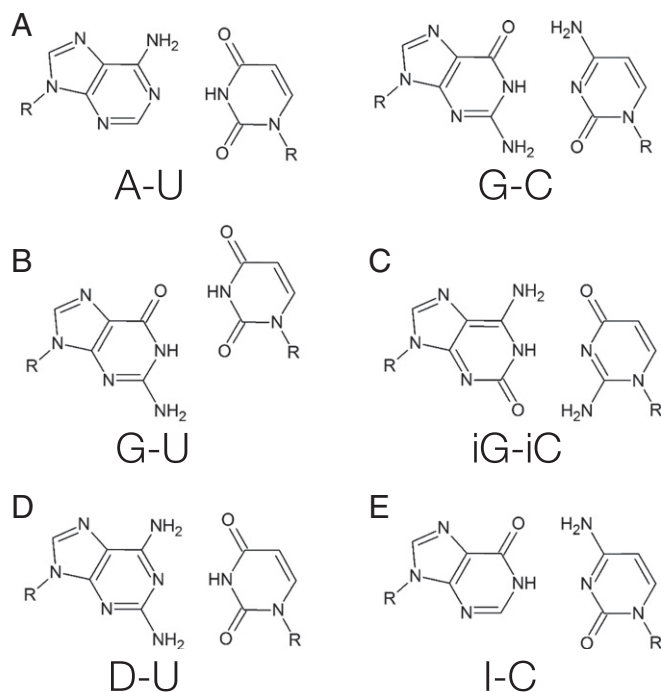
This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. Email: rhiju@stanford.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1523335113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1523335113/-DCSupplemental).

empirically defined NN parameters for most motifs, which are on the scale of 1 kcal/mol. For example, original NN energy estimates for G–U stacked pairs, single-nucleotide bulges, and tetraloop free energies have been corrected by >1 kcal/mol when revisited in detailed studies (11, 25–27). Overall, computational approaches may be ready for calculations of new energetic parameters, including parameters for these uncertain motifs as well as for motifs involving nonstandard nucleotides that are being found throughout natural coding and noncoding RNAs (28, 29) or used to engineer new RNA systems (30, 31). However, the predictive power of these methods has not been evaluated through tests on previously unmeasured NN parameters. Predictive tests are particularly important because models are increasing in complexity and risk overtraining on previously available data.

Here we report, to our knowledge, the first blind tests of a method to computationally predict NN energetic parameters. The newly measured parameters involve RNA stacked pairs with the nonnatural nucleotide 2,6-diaminopurine (D) paired to uracil (Fig. 1). To ensure a rigorous comparison, calculations were carried out by one author (F.-C.C.) and subsequently tested in independent experiments by another author (W.K.). In preparation for this blind test, we developed a reweighting of energy-function collection with conformational ensemble sampling in Rosetta (RECCES–Rosetta) framework to calculate free energies based on density-of-states estimation and expected errors from statistical precision, inaccuracies in the NN assumption, and uncertainties in the weights of the underlying energy function. Furthermore, to address previous ad hoc assumptions used to fit conformational entropy from data, RECCES calculates the conformational entropy of helix and single-stranded states without fitting of additional parameters. These systematic improvements—and calibration based on previously measured NN parameters—ensured that our blind tests carried sufficient power to rigorously establish the accuracy and limitations of NN energy calculations that seek to make nontrivial predictions.



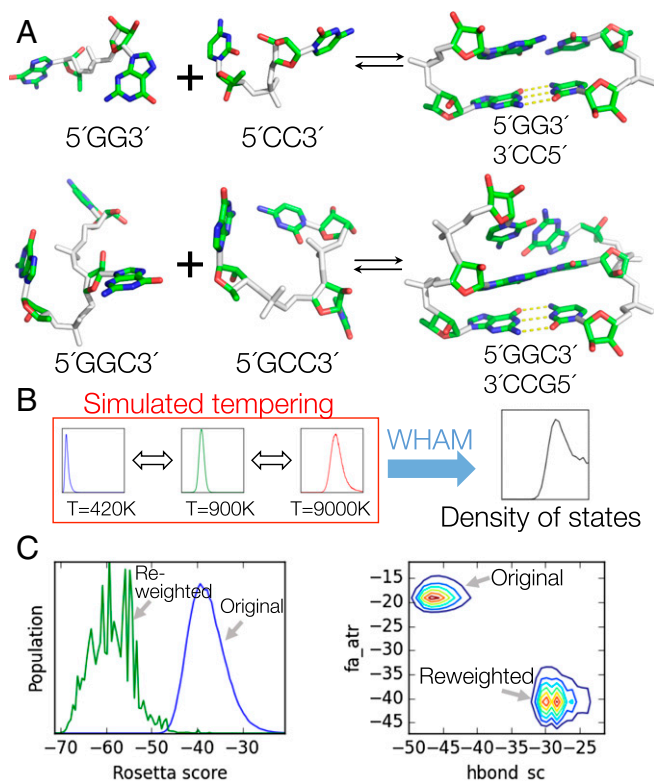
**Fig. 1.** Base pairs involved in NN parameters considered in this study. (A) Canonical pairs adenosine–uracil and guanosine–cytosine, (B) guanosine–uracil wobble pair, (C) nonnatural isoguanosine–isocytidine, (D) nonnatural 2,6-diaminopurine–uracil, and (E) inosine–cytosine.

## Results

**Recovery of Canonical Helix and Dangling-End Parameters.** Blind tests of a prediction method are not worthwhile if the expected prediction errors significantly exceed the range of possible experimental values—on the order of several kilocalories per mole for NN parameters. We therefore first sought to determine whether folding free-energy calculations with the Rosetta all-atom energy function, previously developed for RNA structure prediction and design, could recover NN energetics for canonical Watson–Crick stacked pairs and whether these calculations’ uncertainties were acceptable for making blind predictions. The Rosetta energy function involves separate component terms for hydrogen bonding, electrostatics, van der Waals interactions, nucleobase stacking, torsional potentials, and an orientation-dependent solvation model. Prior structure prediction and design studies did not strongly constrain the weights of these components (18). Thus, we anticipated that NN parameter prediction would require optimization of the weights and care in uncertainty estimation. To assess whether the errors due to weight uncertainties would allow nontrivial predictions, we sought not just a single weight set but instead a large collection of weight sets consistent with available data.

To discover these weight sets, we developed the RECCES framework for sampling conformational ensembles of the single-stranded and helix conformations relevant to NN energy estimation (Fig. 2 and *SI Appendix*, Table S1). Through the use of a density-of-states formalism, simulated tempering, and weighted histogram analysis method (WHAM) integration, RECCES allowed the estimation of free energies with bootstrapped errors of less than 0.003 kcal/mol, significantly less than systematic errors of 0.3 kcal/mol (estimated below; *SI Appendix*, Tables S2–S4), using two central processing unit (CPU) hours of computation per molecule. These methods are similar to replica exchange methods in common use in molecular dynamics studies, but are simpler in that they do not require running multiple parallel processes (*SI Appendix*, *Supporting Methods*). Importantly, the overall RECCES framework did not require separate fitting of conformational entropy factors, reducing the likelihood of overfitting. Furthermore, starting from these initial simulations, RECCES enabled evaluation of alternative weight sets with negligible additional computation (<0.1 s) through a rapid reweighting of cached energies. Though noisy at low energies (compare green to blue curves in Fig. 2C), we confirmed that this reweighting procedure nevertheless led to an acceptable mean calculation error of 0.28 kcal/mol (*SI Appendix*, Table S4), significantly smaller than the several kilocalories per mole range of experimental NN parameters (*SI Appendix*, Table S1). Further tests of the NN assumption, based on simulations with different helix contexts for each stacked pair, also gave systematic errors of 0.2–0.3 kcal/mol (*SI Appendix*, Table S2). Hereafter, we conservatively describe the systematic errors of the RECCES–Rosetta NN parameter estimates to be the higher value in this range, 0.3 kcal/mol.

To obtain a collection of weight sets, we used RECCES to optimize the weights of all terms in the Rosetta score function over numerous runs with different initial values. These optimization runs minimized the mean square error with respect to the NN parameters of 10 canonical stacked pairs (four base pairs next to four base pairs, removing symmetric cases), 32 single-nucleotide dangling ends (four nucleotides at either the 5′ or 3′ end of four base pairs), and the terminal penalty for A–U vs. G–C. The resulting 9,544 minimized weight sets were highly diverse, even after discarding the weight sets with 5% worst rmsd agreement to training data (*SI Appendix*, Table S5, describes score terms and summarizes mean and SDs of weights; *SI Appendix*, Table S6, gives five example weight sets). Most score terms were recovered with mean weights greater than zero by more than one SD, confirming their importance for explaining RNA structure and energetics. These terms included *stack\_elec*, which models the



**Fig. 2.** RECCES thermodynamic framework and reweighting. (A) Example systems simulated for this study. Degrees of freedom sampled are colored in white. The relative orientation of first base pair in each helix was fixed (Right, yellow dashes). (Upper and Lower) Folding reactions of two-base-pair and three-base-pair systems, respectively. (B) Density of state estimation by simulated tempering and WHAM. (C) Reweighting demonstration. (Left) State population at room temperature before (blue) and after (green) reweighting. (Right) Two-dimensional population histograms of  $fa\_atr$  (Lennard-Jones attraction) vs.  $hbond\_sc$  (hydrogen bonds) energy components, before and after reweighting.

electrostatic interaction between stacked nucleobases, an effect previously posited by several groups to be important for understanding fine-scale RNA energetics (14, 32). Terms with wider variance across weight sets could be explained through their covariance with other terms. For example, some pairs of score terms, such as the nucleobase stacking term  $fa\_stack$  and the van der Waals term  $fa\_atr$ , model similar physical effects, but other pairs model opposing effects in helix association, such as hydrogen bonding  $hbond\_sc$  and the solvation term for burying polar moieties  $geom\_sol\_fast$  (SI Appendix, Table S7). The weights of these pairs varied significantly across optimized weight sets, but linear

combinations of these weight pairs were nearly invariant across the weight set collection (SI Appendix, Fig. S1).

Despite the variations and covariations observed across this large collection of weight sets, each weight set gave an rmsd accuracy of better than 0.58 kcal/mol for canonical base pairs and dangling ends, with a mean accuracy of 0.40 kcal/mol across all training data. These accuracies were significantly better than rmsds of 1.51 kcal/mol and 1.23 kcal/mol, respectively, obtained with the original structure prediction weights, supporting the need for reweighting (SI Appendix, Table S6). The rmsd over just the canonical stacked base pairs was 0.28 kcal/mol (Fig. 3A), comparable in accuracy to the initial experimental estimates of these values (10, 12) and consistent with the estimated systematic errors of our calculation strategies (0.3 kcal/mol) (SI Appendix, Tables S2 and S4). For the dangling-end data, RECCES-Rosetta also gave an excellent rmsd of 0.41 kcal/mol (Fig. 3B). For these data, the largest deviations from experiment were tagged as having the highest expected error from weight uncertainties by RECCES, supporting this method of error computation (see, e.g.,  $5^{CG}$  dangling end in SI Appendix, Table S1). For both sets of NN parameters, the rmsd errors were significantly smaller than the range of experimental values (2.5 kcal/mol and 1.5 kcal/mol for canonical stacked pairs and dangling ends, respectively), leading to the visually clear correlations in Fig. 3A and B. The terminal penalty for A-U relative to G-C was also recovered with a similar error (0.3 kcal/mol) (SI Appendix includes further discussion and computation of other terminal base pair contributions).

Because we directly trained the RECCES score function against the experimental dataset, the accuracies of these results were expected. Nevertheless, we gained further confidence in the use of Rosetta-derived energy functions and RECCES framework by comparing its performance to the results of two simpler models trained on the same data. First, a three-parameter hydrogen-bond counting model, similar to simple phenomenological models that inspired the NN parametrization (10) (SI Appendix, Supporting Methods), achieved rmsd accuracies of 0.29 kcal/mol and 0.45 kcal/mol on canonical stacked pairs and dangling ends, respectively—slightly worse than the RECCES results (0.28 kcal/mol and 0.41 kcal/mol, respectively), despite including fitted parameters that account for conformational entropy loss of base pairs and dangling ends. Second, a prior single-conformation Rosetta method, which uses the same energy function as RECCES-Rosetta but evaluates the score only for a minimized helix conformation (18) achieved accuracies of 0.30 kcal/mol and 0.44 kcal/mol for canonical stacked pairs and dangling ends, respectively—again worse than the RECCES-Rosetta results despite including separately fitted conformational entropy terms. For all three models, the largest deviation was for the stacked pair  $5^{CG}$ , which is less stable than the other stacked pairs with two G-C pairs by 1 kcal/mol; still, even for this parameter, the RECCES-Rosetta calculations were more accurate than the simpler models. These comparisons supported the utility of

**Table 1.** Accuracies of nearest-neighbor parameter predictions

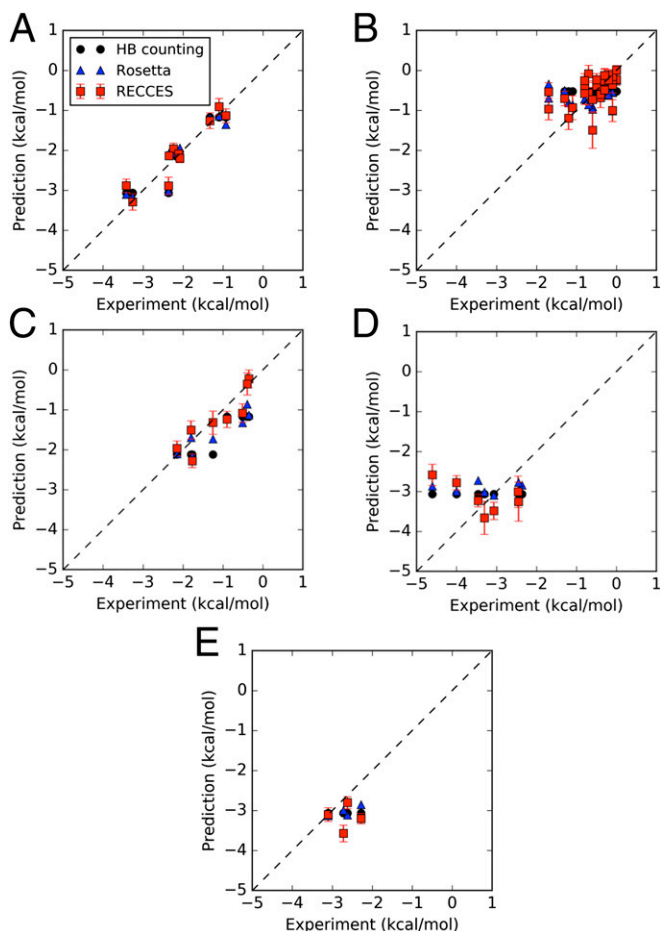
RNA motif category	No. motifs	Rmsd accuracy (kcal/mol)			
		Hydrogen-bond counting	Single-conformation Rosetta	RECCES-Rosetta	RECCES-Rosetta refitted*
Canonical <sup>†</sup>	10	0.29	0.30	0.28	0.41
Dangling <sup>†</sup>	32	0.45	0.44	0.41	0.43
G-U <sup>‡</sup>	8	0.59	0.49	0.32	0.32
iG-iC <sup>‡</sup>	7	0.79	0.85	0.99	1.08
D-U <sup>§</sup>	4	0.48	0.40	0.63	0.46
All	61	0.50	0.49	0.50	0.53

\*The model was trained with all data available, so all entries in the column are training data.

<sup>†</sup>Data used in training the models.

<sup>‡</sup>Data set aside for testing.

<sup>§</sup>Blind test data.



**Fig. 3.** Calculations vs. experiment for each NN parameter set. (A) Canonical stacked pairs; (B) single-nucleotide dangling ends; (C) stacked pairs including one G–U pair; (D) stacked pairs including at least one iG–iC pair; (E) stacked pairs including one D–U pair. All panels are drawn with the same axis limits and a line of equality (dashed) to aid cross-panel visual comparison.

the RECCES–Rosetta method compared with less generalizable models. However, the significance in the accuracy improvement was difficult to rigorously evaluate because the models contained different numbers and types of parameters; we therefore turned to independent test sets and blind predictions.

#### Tests on Independent Nearest-Neighbor Parameter Measurements.

Recent comprehensive experimental measurements have updated the NN parameters for stacked pairs involving G–U wobble pairs next to canonical Watson–Crick pairs (11). Because these values were not used in the training of the models herein and because the geometry of G–U wobble pair is distinct from G–C and A–U pairs (Fig. 1B), this set of measurements offered strong tests of modeling accuracy. Furthermore, the expected error in the RECCES–Rosetta calculations from weight uncertainties, based on variation across the large collection of weight sets, was 0.22 kcal/mol (SI Appendix, Table S1), less than the estimated ~0.3 kcal/mol systematic error (SI Appendix, Tables S2 and S4). Both error contributions were significantly less than the full range of predicted NN parameters (2.1 kcal/mol), supporting the strength of this test. The actual rmsd accuracy across these G–U NN measurements was 0.32 kcal/mol for RECCES–Rosetta (Table 1), nearly as accurate as the recovery of training set stacked pairs (0.28 kcal/mol) and comparable to expected systematic errors. Furthermore, this accuracy over G–U-containing stacked pairs outperformed the rmsd values calculated from hydrogen-bond counting and single-conformation

Rosetta scoring methods (0.59 and 0.49 kcal/mol, respectively) by 50–80%, supporting the importance of carrying out detailed physical simulations of the conformational ensemble via RECCES over simpler approaches. Here and below, the predictions and their estimated errors were calculated by computing means and SDs of NN parameters across the full collection of weight sets discovered by RECCES. Compared with this averaging over multiple models, using the single weight set with best fit to the training data gave slightly worse accuracies on the test data (SI Appendix, Table S6) (33).

A more difficult test involved seven previously measured NN parameters of a nonnatural base pair, iG–iC (Fig. 1C) (34). The rmsd for the iG–iC test case was 0.99 kcal/mol, mainly due to two significant outliers:  ${}^5\text{iG}i\text{C}$  and  ${}^5\text{iG}i\text{C}$  (Fig. 3D). The predicted NN parameters for these outliers were larger than experimental values (less stable) by 2.2 and 1.3 kcal/mol, respectively. Nevertheless, over the other five iG–iC NN parameters, the rmsd was 0.51 kcal/mol, and the discrepancies appeared primarily due to a systematic offset in the predictions (Fig. 3D). The accuracy was comparable to the maximum errors expected from weight uncertainties (0.4–0.5 kcal/mol) and similar, in terms of relative accuracies, to the canonical and G–U-containing stacked pairs above. Compared with RECCES–Rosetta, the simpler hydrogen-bond counting and single-conformation Rosetta scoring models gave 15–20% better accuracies (0.79 and 0.85 kcal/mol, respectively; 0.47 and 0.44 kcal/mol, excluding outliers); but both simple models gave near-constant NN parameters (range less than 0.3 kcal/mol) over all stacked pairs, providing no explanation for the 2.2 kcal/mol range in experimental measurements or for the outliers (Fig. 3D). On one hand, the two outliers suggest that some important physical effect is missing or incorrectly implemented in the current calculation procedure (see Discussion). On the other hand, the excellent accuracies over the other iC–iG-containing stacked pairs, along with the performance in the G–U test set, motivated us to continue with blind comparisons.

#### Blind Tests Involving Diaminopurine–Uracil Base Pairs.

As a blind test, we applied RECCES–Rosetta to predict the NN parameters for stacked pairs involving a distinct nonnatural base pair, 2,6-diaminopurine paired with uracil (D–U) (Fig. 1D). Predictions of these parameters (SI Appendix, Table S1) suggested a wide range of NN values and confirmed that errors from weight uncertainties were smaller or comparable to other systematic sources of error (0.3 kcal/mol). To test these predictions, we measured NN parameters for the four stacked pairs involving D–U next to G–C pairs, which were expected to have a range of 0.8 kcal/mol. SI Appendix, Table S8, gives construct sequences and experimental folding free-energy values for these constructs, and Table 1 and SI Appendix, Tables S1 and S9, summarize the NN parameter estimation. The rmsd of the RECCES–Rosetta blind predictions was 0.63 kcal/mol (Fig. 3E). The hydrogen-bond counting and single-conformation Rosetta scoring models, which fared worse than RECCES–Rosetta in most tests above, gave rmsds of 0.48 and 0.40 kcal/mol, respectively, better than RECCES–Rosetta by 24–37% (Table 1). This result is similar to what we observed in the iG–iC test case; indeed, the two simple models again produced near constant predictions (range < 0.2 kcal/mol) for the D–U stacked pairs that did not account for the 0.8 kcal/mol range of the measured values (Fig. 3E). Given the blind nature of the test and our attempts to ensure its power to falsify our calculations, this test unambiguously indicated that some physical term is missing in the current Rosetta all-atom energetic model (as well as simpler models). Nevertheless, the results are encouraging: the blind predictions from each of the three models over each of four NN values separately achieved better than 1 kcal/mol accuracy compared with subsequent experimental measurements.

**Post Hoc Fit Across All Data.** Though post hoc tests of models on prior collected data are less rigorous than blind trials, they can

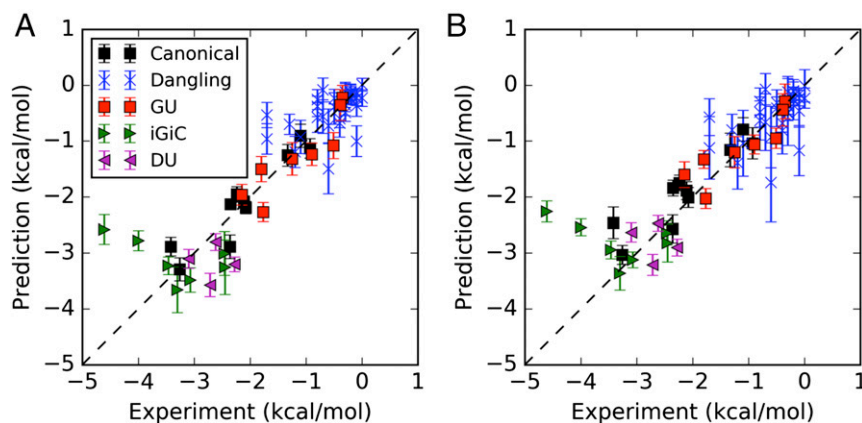
help guide future work. As a final test, we wished to understand possible explanations for the worse accuracy of RECCES–Rosetta in iG–iC test cases and blind D–U trials compared with the G–U test cases. One model for this inaccuracy was that overfitting of energy function weights to the training data worsened predictive power over the new data. Another (not necessarily exclusive) model was that the underlying energy function derived from Rosetta score terms was fundamentally incapable of modeling the available NN data under any weight set with the RECCES procedure. We were able to test these models by carrying out a post hoc global fit of energy function weights over all available NN data (Fig. 4 and Table 1). As expected, we observed better fits to the test data, including an improvement in rmsd accuracy for the four D–U stacked pairs from 0.63 kcal/mol to 0.46 kcal/mol; this result suggests a modest overfitting to the training set in the studies above. However, we observed somewhat worse fits to the training data, including a worsening of rmsd accuracy for the 10 canonical stacked pairs from 0.28 to 0.41 kcal/mol, worse than expected systematic errors in our calculations (0.3 kcal/mol) (*SI Appendix, Tables S2 and S4*) supporting the second model of fundamental energy function inaccuracy. Furthermore, this global fit still failed to account for the two striking outliers involving iG–iC base pairs, again giving evidence for the second model: energetic calculations based on the current Rosetta score function are fundamentally incapable of accounting for all of the data within expected error, even with a post hoc optimized weight set.

## Discussion

This study reports, to our knowledge, the first blind test of the predictive power of high-resolution, all-atom modeling methods for RNA folding energetics. We developed a RECCES strategy in the Rosetta framework that rigorously models conformational ensembles of single strand and helical states, is computationally efficient (hours with currently available CPUs), and brackets systematic errors based on comprehensive reweighting tests. Compared with simpler phenomenological methods, RECCES–Rosetta achieved excellent rmsd accuracies for the NN parameters of canonical base pairs, dangling ends, and G–U pairs, but somewhat worse results for NN parameters involving nonnatural base pairs iG–iC and D–U. The latter D–U parameters were measured after the predictions as a blind test. The computational accuracies were better than 1 kcal/mol in all cases, based on rmsd values over each separate set of NN parameters (0.28, 0.41, 0.32, 0.99, and 0.63 kcal/mol for canonical, dangling end, G–U, iG–iC, and D–U parameters, respectively) and also individually for each of the four blind predictions. These rmsd values are significantly

smaller than the 2–3 kcal/mol ranges measured for these sets of NN values (Fig. 4 and *SI Appendix, Table S1*), are comparable to errors in ad hoc fits used in the current NN model for most motifs (11, 25–27), and are generally smaller than molecular dynamics calculations that remain significantly more expensive (21, 22). The generality of the RECCES–Rosetta framework and this level of success in initial tests support the further development of RECCES–Rosetta for nonnatural nucleotides and for motifs more complex than the helical stacked pairs and dangling ends considered herein.

While achieving consistently sub-kcal/mol accuracies, there is room for improvement in the RECCES–Rosetta approach. For example, the modeling does not account for the 1 kcal/mol stability increase of the  ${}^5\text{GC}$  NN parameter relative to  ${}^5\text{CG}$ ; the electrostatic term *stack<sub>elec</sub>* does favor the former, but is not assigned a strong enough weight in the final fits to account for the stability difference. Also, the rmsd accuracies still remain larger than estimated systematic errors (0.3 kcal/mol), particularly for the nonnatural base pairs in the test data, and the discrepancies remain even if those data are included in a post hoc fit of the energy function weights to all available measurements. Our results help bracket which strategies might improve the accuracy and which might not. On one hand, nonnatural pairs present their atomic moieties in different bonded contexts, which might modulate the strengths of hydrogen bonds or other interactions that they form. For example, a previous analysis suggested that the hydrogen bonds in an iG–iC base pair might be stronger than in a G–C base pair by  $\sim 0.4$  kcal/mol (14). Accounting for this effect would be predicted to offset our calculated NN parameters for all iG–iC stacked pairs, without changing their relative ordering, and cannot account for strong outliers. Indeed, if we added an extra fitting term for stabilizing iG–iC pairing, the rmsd accuracy over these data did not significantly improve (0.96 kcal/mol vs. 0.99 kcal/mol without the extra term). On the other hand, several unmodeled factors are sensitive to the ordering of base pairs within stacked pairs and could affect the relative ordering of NN parameters within each set. For example, the current Rosetta all-atom score function models electrostatics through fixed charges with a distance-dependent dielectric and does not explicitly model water or counterions that may differentially stabilize the base pair steps (35, 36). Recent and planned additions of nonlinear Poisson–Boltzmann solvation models, polarizable electrostatic models, and a potential of mean force for water-mediated hydrogen bonding into the Rosetta framework should allow evaluation of whether these physical effects can improve accuracy of NN



**Fig. 4.** Calculations vs. experiment across all NN parameters. Comparisons are based on (A) RECCES–Rosetta weight sets trained on canonical and dangling-end data (same values as in Fig. 3) and (B) “best-case” weight sets fitted post hoc over all available NN parameters, including D–U stacked pairs measured for blind predictions.

parameter calculations to the 0.3 kcal/mol fundamental limit of the RECCES method. If these models can also be expanded to calculate the temperature dependence of solvation, it may also become possible to compare calculated and measured entropies and enthalpies of the NN parameters, which are well measured but may be dominated by solvation effects. In addition, we propose that calculations for recently characterized stacked pairs that give anomalous NN parameters, including some tandem G–U stacked pairs (11) and pseudouridine–A–containing stacked pairs (37, 38), could offer particularly stringent tests.

Continuing work in modeling RNA energetics will benefit from further blind trials, perhaps in a community-wide setting analogous to the ongoing RNA-puzzle structure prediction trials (39, 40). The prediction of two kinds of parameters could serve as future blind tests. First, based on the results herein, nonnatural base pairs offer good test cases and require the same amount of computational power as canonical base pair NN parameter estimation. Alternative approaches based on, e.g., molecular dynamics, should also be applicable to these cases. We have completed RECCES–Rosetta predictions for additional stacked pairs involving iG–iC and D–U pairs, as well as for inosine–cytosine (I–C) base pairs (Fig. 1E and *SI Appendix, Table S1*), but are waiting to make experimental measurements until there are comparison values from other groups and approaches. Second, future blind trials might involve predicting energetics of RNA motifs more complex than those considered herein, such as apical loops, internal loops, multihelix junctions, and tertiary interactions. For these cases, an expansion of the

RECCES approach in which physically realistic candidate conformations of each motif are first estimated with structure prediction (18, 41) and then subjected to rigorous RECCES-based free-energy calculations may offer predictive power. Such an approach may also allow calculations of next-NN effects and development of rapid approximations to estimate conformational entropy of candidate conformations, which would be useful for structure prediction and design (*SI Appendix, Fig. S2*). A new generation of high-throughput RNA biochemistry platforms (42–44) offers the prospect of both training these next-generation energetic prediction algorithms and carrying out blind tests with many thousands of measurements.

## Materials and Methods

Details of NN parameter estimation with RECCES (including basic equations, simulation parameters, and energy function) and with simple single-conformation methods, as well as methods used to experimentally estimate NN parameters for helices with D–U base pairs, are presented in *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank J. Yesselman for help in generating the partial charges for nonnatural RNA bases; and P. Sripakdeevong, K. Beauchamp, W. Greenleaf, and members of R.D.'s laboratory for useful discussions. Calculations were performed using the Texas Advanced Computing Center Stampede cluster through Extreme Science and Engineering Discovery Environment (XSEDE) Allocation Project MCB120152, and the Stanford BioX<sup>3</sup> cluster. This work is supported by a Howard Hughes Medical Institute International Student Research Fellowship (to F.-C.C.); a Stanford BioX Graduate Student Fellowship (to F.-C.C.); a Burroughs-Wellcome Career Award at Scientific Interface (to R.D.); and NIH Grants NIGMS R21 GM102716 and R01 GM102519 (to R.D.).

- Bloomfield VA, Crothers DM, Tinoco I (2000) *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, CA).
- Atkins JF, Gesteland RF, Cech TR (2010) *RNA Worlds: From Life's Origins to Diversity in Gene Regulation* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
- Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819–823.
- Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823–826.
- Hannon GJ (2002) RNA interference. *Nature* 418(6894):244–251.
- Cantor CR, Tinoco I, Jr (1965) Absorption and optical rotatory dispersion of seven trinucleoside diphosphates. *J Mol Biol* 13(1):65–77.
- Kent JL, et al. (2014) Non-nearest-neighbor dependence of stability for group III RNA single nucleotide bulge loops. *RNA* 20(6):825–834.
- Lavery R, et al. (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res* 38(1):299–313.
- Vanegas PL, Horwitz TS, Znosko BM (2012) Effects of non-nearest neighbors on the thermodynamic stability of RNA GNRA hairpin tetraloops. *Biochemistry* 51(11):2192–2198.
- Xia T, et al. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37(42):14719–14735.
- Chen JL, et al. (2012) Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry* 51(16):3508–3522.
- Serra MJ, Turner DH (1995) Predicting thermodynamic properties of RNA. *Methods Enzymol* 259:242–261.
- 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(5):911–940.
- Yildirim I, Turner DH (2005) RNA challenges for computational chemists. *Biochemistry* 44(40):13225–13234.
- Liu B, Diamond JM, Mathews DH, Turner DH (2011) Fluorescence competition and optical melting measurements of RNA three-way multibranch loops provide a revised model for thermodynamic parameters. *Biochemistry* 50(5):640–653.
- Mathews DH, Turner DH (2002) Experimentally derived nearest-neighbor parameters for the stability of RNA three- and four-way multibranch loops. *Biochemistry* 41(3):869–880.
- Diamond JM, Turner DH, Mathews DH (2001) Thermodynamics of three-way multibranch loops in RNA. *Biochemistry* 40(23):6971–6981.
- Das R, Karanicas J, Baker D (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* 7(4):291–294.
- Sarzynska J, Nilsson L, Kulinski T (2003) Effects of base substitutions in an RNA hairpin from molecular dynamics and free energy simulations. *Biophys J* 85(6):3445–3459.
- Deng N-J, Cieplak P (2007) Molecular dynamics and free energy study of the conformational equilibria in the UUUU RNA Hairpin. *J Chem Theory Comput* 3(4):1435–1450.
- Deng N-J, Cieplak P (2010) Free energy profile of RNA hairpins: A molecular dynamics simulation study. *Biophys J* 98(4):627–636.
- Spasic A, Serafini J, Mathews DH (2012) The Amber ff99 force field predicts relative free energy changes for RNA helix formation. *J Chem Theory Comput* 8(7):2497–2505.
- Réblová K, et al. (2010) An RNA molecular switch: Intrinsic flexibility of 23S rRNA helices 40 and 68 5'-UAA/5'-GAN internal loops studied by molecular dynamics methods. *J Chem Theory Comput* 10(6):910–929.
- Van Nostrand KP, Kennedy SD, Turner DH, Mathews DH (2011) Molecular mechanics investigation of an adenine-adenine non-canonical pair conformational change. *J Chem Theory Comput* 7(11):3779–3792.
- Znosko BM, Silvestri SB, Volkman H, Boswell B, Serra MJ (2002) Thermodynamic parameters for an expanded nearest-neighbor model for the formation of RNA duplexes with single nucleotide bulges. *Biochemistry* 41(33):10406–10417.
- Blose JM, et al. (2007) Non-nearest-neighbor dependence of the stability for RNA bulge loops based on the complete set of group I single-nucleotide bulge loops. *Biochemistry* 46(51):15123–15135.
- Sheehy JP, Davis AR, Znosko BM (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* 16(2):417–429.
- Roost C, et al. (2015) Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J Am Chem Soc* 137(5):2107–2115.
- Zhao BS, He C (2015) Pseudouridine in a new era of RNA modifications. *Cell Res* 25(2):153–154.
- Ng EW, et al. (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat Rev Drug Discov* 5(2):123–132.
- Sipa K, et al. (2007) Effect of base modifications on structure, thermodynamic stability, and gene silencing activity of short interfering RNA. *RNA* 13(8):1301–1316.
- Proctor DJ, et al. (2004) Folding thermodynamics and kinetics of YNMG RNA hairpins: Specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry* 43(44):14004–14014.
- Opitz D, Maclin R (1999) Popular ensemble methods: An empirical study. *J Artif Intell Res* 11:169–198.
- Chen X, Kierzek R, Turner DH (2001) Stability and structure of RNA duplexes containing isoguanosine and isocytidine. *J Am Chem Soc* 123(7):1267–1274.
- Auffinger P, Westhof E (2000) Water and ion binding around RNA and DNA (C,G) oligomers. *J Mol Biol* 300(5):1113–1131.
- Chen Z, Znosko BM (2013) Effect of sodium ions on RNA duplex stability. *Biochemistry* 52(42):7477–7485.
- Hudson GA, Bloomingdale RJ, Znosko BM (2013) Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA* 19(11):1474–1482.
- Kierzek E, et al. (2014) The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* 42(5):3492–3501.
- Cruz JA, et al. (2012) RNA-puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 18(4):610–625.
- Miao Z, et al. (2015) RNA-puzzles round II: Assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* 21(6):1066–1084.
- Sripakdeevong P, Kladwang W, Das R (2011) An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc Natl Acad Sci USA* 108(51):20573–20578.
- Buenrostro JD, et al. (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat Biotechnol* 32(6):562–568.
- Tome JM, et al. (2014) Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling. *Nat Methods* 11(6):683–688.
- Ozer A, Pagano JM, Lis JT (2014) New technologies provide quantum changes in the scale, speed, and success of SELEX methods and aptamer characterization. *Mol Ther Nucleic Acids* 3:e183.