

# SCIENTIFIC REPORTS



OPEN

## Capturing RNA Folding Free Energy with Coarse-Grained Molecular Dynamics Simulations

David R. Bell<sup>1</sup>, Sara Y. Cheng<sup>2</sup>, Heber Salazar<sup>1</sup> & Pengyu Ren<sup>1</sup>

Received: 15 August 2016

Accepted: 06 March 2017

Published: 10 April 2017

We introduce a coarse-grained RNA model for molecular dynamics simulations, RACER (RnA CoarsE-grained). RACER achieves accurate native structure prediction for a number of RNAs (average RMSD of 2.93 Å) and the sequence-specific variation of free energy is in excellent agreement with experimentally measured stabilities ( $R^2 = 0.93$ ). Using RACER, we identified hydrogen-bonding (or base pairing), base stacking, and electrostatic interactions as essential driving forces for RNA folding. Also, we found that separating pairing vs. stacking interactions allowed RACER to distinguish folded vs. unfolded states. In RACER, base pairing and stacking interactions each provide an approximate stability of 3–4 kcal/mol for an A-form helix. RACER was developed based on PDB structural statistics and experimental thermodynamic data. In contrast with previous work, RACER implements a novel effective vdW potential energy function, which led us to re-parameterize hydrogen bond and electrostatic potential energy functions. Further, RACER is validated and optimized using a simulated annealing protocol to generate potential energy vs. RMSD landscapes. Finally, RACER is tested using extensive equilibrium pulling simulations (0.86 ms total) on eleven RNA sequences (hairpins and duplexes).

### RNA serves important and diverse functions inside the cell

In 1981, Thomas Cech and colleagues observed self-splicing RNA in a 26 S rRNA precursor<sup>1,2</sup>. In 1983, Sidney Altman found that ribonuclease P could cleave tRNA in the absence of protein<sup>3</sup>. In 2002, it was discovered that even mRNAs could bind small metabolites and regulate protein expression<sup>4–6</sup>. Today, RNA is recognized as extensively active, with roles in regulating genes, preparatory cleavage, metabolite sensing, and immune response. RNAs achieve this diverse activity through intricately regulated structure, with catalytic RNAs such as riboswitches maintaining highly conserved functional regions<sup>7,8</sup>.

### RNA chemistry and the need for accurate structures

RNA structure is a challenge to determine experimentally because it can fold into many different structures. For example, during RNA transcription, synthesized RNA regions fold locally<sup>9</sup>, sampling hairpins and short-range motifs. After transcription completes, RNA molecules are able to fold completely and sample long-range interactions<sup>10</sup>. With the numerous structures available for RNA to fold into, long-lived misfolded RNA intermediates often occur<sup>11–13</sup>. In addition, heterogeneous folding pathways exist for the same RNA sequence<sup>14–20</sup>. As a result, RNA has a highly dynamic folding landscape, which is challenging to capture using techniques such as x-ray crystallography and NMR spectroscopy<sup>21,22</sup>. Further, due to only recent interest in the diversity of RNA function in biology, there is a deficiency in available RNA experimental structures. However, RNA structure is key to understanding its function and for development of RNA-based applications. Due to the lack of available experimental structures of RNA, computational models of RNA are vital to predict RNA structures.

### Secondary structure methods for RNA

Currently, there are a variety of structure prediction methods available to elucidate RNA structure. Secondary structure prediction methods predict base pairing contacts for a given RNA sequence<sup>23</sup>. If homologous sequences exist, comparative sequence analysis<sup>24–27</sup> remains the most accurate secondary structure technique. One of the most popular secondary structure prediction methods is dynamic programming. Using nearest neighbor

<sup>1</sup>Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712, United States.

<sup>2</sup>Department of Physics, University of Texas at Austin, Austin, Texas 78712, United States. Correspondence and requests for materials should be addressed to P.R. (email: pren@mail.utexas.edu)

energies<sup>28</sup> and the sequence of the RNA, dynamic programming methods, such as Mfold<sup>29,30</sup> or ViennaRNA<sup>31–33</sup>, exhaustively compare and build secondary structures to achieve the minimum free energy structure.

However, dynamic programming schemes face certain limitations<sup>34</sup>, such as difficulty predicting pseudoknot structures. Various secondary structure programs<sup>35–37</sup> have been developed to predict the folding of these structures. Recently, it has been shown that incorporating results from the experimental method SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension)<sup>38</sup> can moderately increase accuracy of secondary structure prediction<sup>39–44</sup>. Despite its utility, secondary structure prediction is ultimately limited to 2-D base paired RNA structures. For RNA based therapeutics and *de novo* design, 3-D RNA structure must be determined.

### 3-D structure prediction models

Tertiary or 3-D structure prediction methods use template, graph theory, and physics based modeling to sample and predict relevant 3-D RNA structures<sup>45,46</sup>. Template based modeling uses predefined, small motifs to assemble RNA structures from their sequence. Template based models include the MC-Fold/MC-Sym pipeline<sup>47</sup>, BARNACLE<sup>48</sup>, RSIM<sup>49</sup>, 3dRNA<sup>50</sup>, RNAComposer<sup>51</sup>, Vfold<sup>52–55</sup>, RNA-MoIP<sup>56</sup> and FARNA/FARFAR<sup>57–59</sup> available in the Rosetta package<sup>60</sup>. Similar to template based modeling, ASSEMBLE<sup>61</sup> and RNA2D3D<sup>62</sup> use homologous RNA structures to predict the new RNA structure (with manual refinement available). In graph theory techniques, RNA is depicted topologically to build RNA structures; this improves sampling and even allows for creation of novel RNA motifs. Graph theory techniques<sup>63</sup> are utilized by RAG/RAGTOP<sup>64–67</sup> and others<sup>68–71</sup>. In physics based methods, the RNA is built from sequence into a 3D structure, and these 3D RNA structures are sampled using Monte Carlo or Molecular Dynamics (MD) protocols. Due to the high charge density of RNA and the associated large computational cost to sample structures, many tertiary structure models use coarse-grained representations of RNA<sup>72</sup>.

In coarse-grained (CG) models, atomic sites are grouped together and represented as a “bead” or pseudoatom. Typical coarse-grained models depict a few pseudoatoms per nucleotide. This results in a reduction in the degrees of freedom and lowers the simulation cost of the model, as compared with simulating the all-atom structure. Physics based coarse-grained models with one pseudoatom per nucleotide include YAMMP/YUP<sup>73,74</sup>, an adaptable user input required model, and NAST<sup>75,76</sup>, which assumes ideal helices from secondary structure and uses MD and clustering to build loops. iFoldRNA<sup>77,78</sup>, Denesyuk *et al.*<sup>79,80</sup>, and TOPRNA<sup>81–83</sup> use three pseudoatoms per nucleotide to depict phosphate, sugar, and nucleobase groups. iFoldRNA uses discrete Molecular Dynamics and replica exchange Molecular Dynamics to sample structures, with non-bonded parameters decomposed from nearest neighbor energies. Similarly, the model by Denesyuk *et al.*<sup>79,80</sup> derives its parameters from nearest neighbor energies and experimentally determined structures. TOPRNA captures effects of secondary structure constraints on loop conformations and free energies. HiRE-RNA<sup>84–86</sup> depicts six-seven pseudoatoms per nucleotide with five pseudoatoms along the backbone. SimRNA<sup>87,88</sup>, Bernauer *et al.*<sup>89</sup>, as well as the previous generation and current RACER model studied<sup>90–92</sup>, all represent RNA with five pseudoatoms per nucleotide. SimRNA uses a Monte Carlo sampling algorithm with parameters from statistical potentials. The model by Bernauer *et al.* similarly uses statistics from high-resolution crystal structures for parameterization yet also derives all-atom potentials for structure refinement.

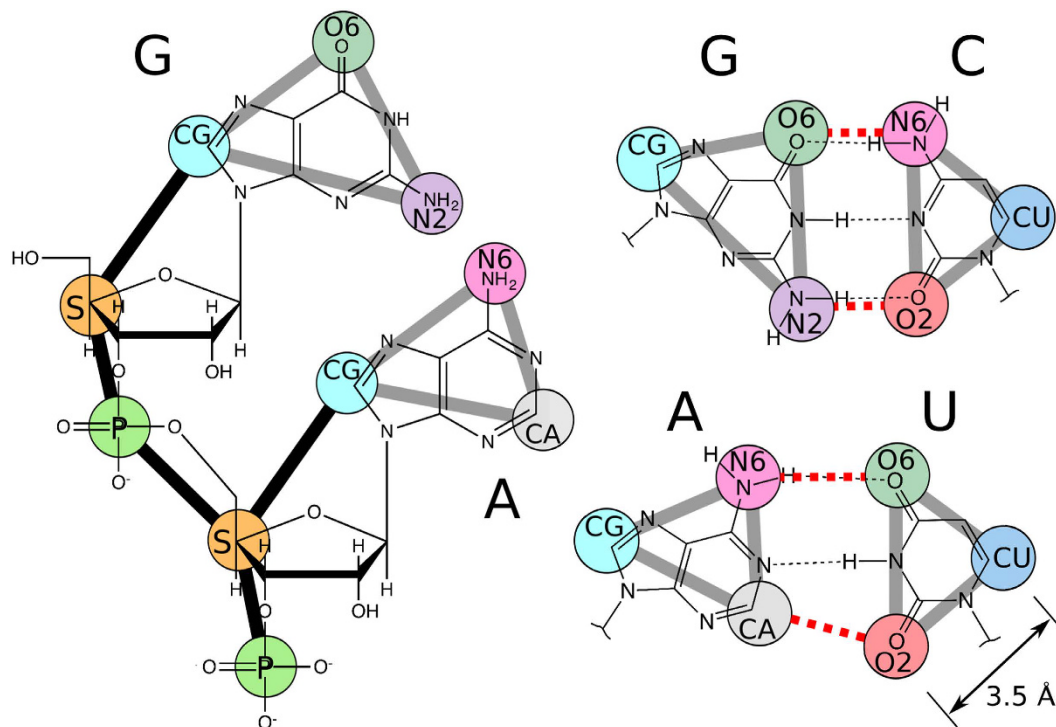
### The RACER RNA Model

The CG RNA model RACER (RnA CoarsE-gRained) developed and applied in this work is a physics-based model, derived from RNA structural statistics, refined using RNA thermodynamics, and applied in molecular dynamics simulations of folding and complexation of RNAs. In the results section, we first introduce the potential energy functions used in the RACER model, with a focus on the newly implemented effective vdW potential. Second, we demonstrate how RACER parameters were optimized using statistical potentials derived from PDB statistics. Additionally, we provide motivation for modeling RNA as a modeling RNA as a 1D molecule and the associated 1D correction we made to the non-bonded PMFs. Third, we show how we validated RACER using simulated annealing simulations for RACER structure prediction capability and generation of funnel free energy landscapes. Fourth, we apply RACER to generate folding free energy predictions for a testing set of RNA hairpins and duplexes, and we compare our results to experiments. In the discussion section, we summarize the changes made to the RACER model and emphasize RACER's ability to capture folding free energies and to predict structures. In the methods section, we show (1) the ability of RACER to map between all-atom and coarse-grained representations for use in multiscale simulations, (2) details on the folding free energy calculations, and (3) implementation instructions for those wishing to use RACER.

### Results

**Model.** *Potential energy functions.* The total potential energy function of the RACER model includes bond stretching, angle bending, torsion, effective vdW, hydrogen bonding, and electrostatics, labeled as  $E_{bond}$ ,  $E_{angle}$ ,  $E_{torsion}$ ,  $E_{vdW\_eff}$ ,  $E_{hb}$ , and  $E_{ele}$  respectively (see Eq. 1). The RACER model is currently implemented in TINKER<sup>93</sup>. In RACER, RNA nucleotides consists of 5 pseudoatoms per nucleotide, with a total of 9 pseudoatom types (shown in Fig. 1). The RACER model used here differs from previous publications<sup>90,91</sup> in that we employ a novel effective vdW potential to better capture the short-range non-bonded interactions among the pseudoatoms, which we found to be essential for correctly capturing the folded state. As a result, we had to re-parameterize the other non-bonded contributors including the electrostatics and hydrogen bonding potential.

$$E = E_{bond} + E_{angle} + E_{torsion} + E_{vdW\_eff} + E_{hb} + E_{ele} \quad (1)$$



**Figure 1. RACER model pseudoatoms overlapping all-atom structure.** RACER bonds are shown in bold lines. (Left) nucleobases with backbone. (Right) GC and AU basepairs with hydrogen bonds shown in red dashed lines. Scale bar is shown in lower right.

**Bonded Potential Energies.** The potential energy functions which retain the same functional form between the previous model and RACER are the bonded potential energy functions. Bond and angle potentials are represented by harmonic terms:  $E_{bond} = k_{bond}(b - b_o)^2$  and  $E_{angle} = k_{angle}(\theta - \theta_o)^2$ . The torsion potential of Eq. 2 uses the first 3 terms of a Fourier series expansion for the torsion potential, where  $\phi$  is the torsion angle, and  $k_n$  and  $\delta_n$  are the spring constant and phase angle of expansion term  $n$ .

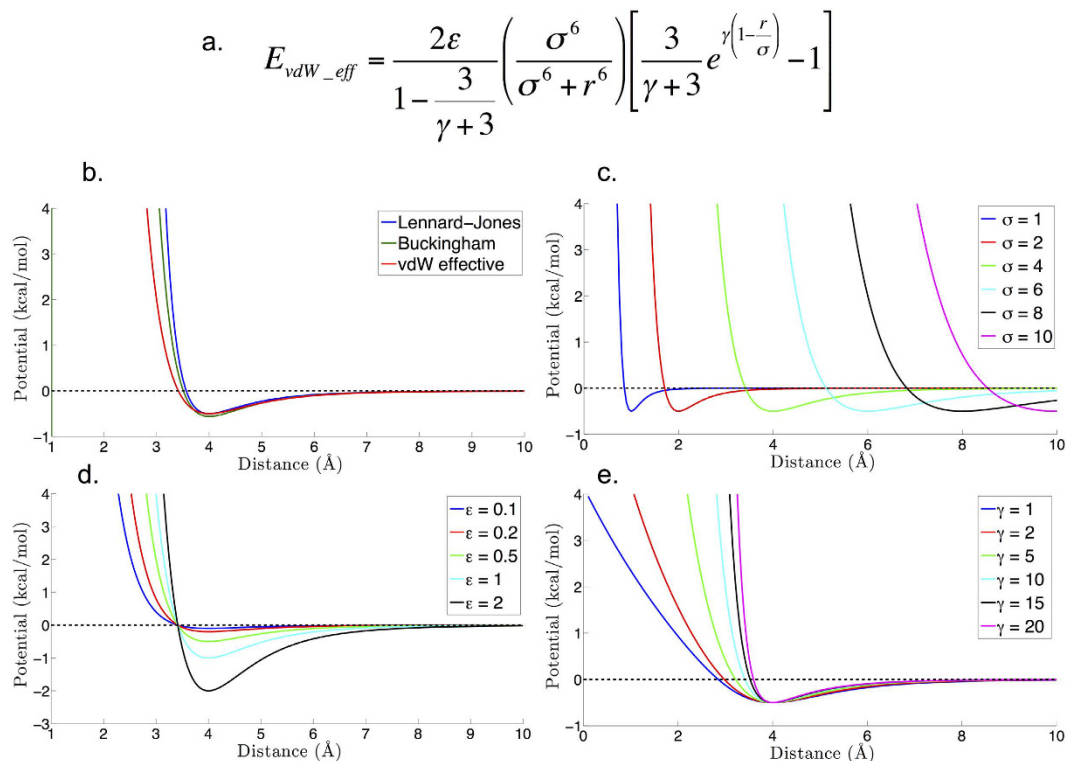
$$E_{torsion}(\phi) = \sum_{n=1}^3 k_n (1 + \cos(n\phi - \delta_n)) \quad (2)$$

**Improved Effective vdW Potential.** The RACER model includes a newly implemented effective potential ( $vdW_{eff}$ ) that significantly improves the fit of RACER to non-bonded statistical potentials. In the previous model<sup>92</sup> the vdW-like non-bonded potential was modeled using a Buckingham function. However, this was found to significantly overestimate repulsion at short distances when compared with statistical potentials. The new effective  $vdW_{eff}$  potential (Eq. 5) allows for tuning the repulsion at short distances through a third parameter  $\gamma$ , enabling a closer fit to the statistical non-bonded potential of mean force (PMF) (Fig. 2).

The  $vdW_{eff}$  does not represent the true vdW interaction, but rather the potential of mean force between a pair of pseudoatoms. However, based on statistical potentials, the non-bonded interactions between most pairs of pseudoatoms we sampled exhibited vdW potential-like behavior. The new functional form for  $vdW_{eff}$  potential taken from ref. 94 is shown in Eq. 3, where  $\varepsilon$  is the minimum well depth and  $\sigma$  is the distance of minimum energy, and  $\gamma$  is a parameter allowing for fine-tuning of the slope of the short-range interaction. Figure 2b presents a comparison between the  $vdW_{eff}$ , Lennard Jones, and Buckingham potentials while Fig. 2c–e show the effects of the three parameters  $\sigma$ ,  $\varepsilon$ , and  $\gamma$  on the  $vdW_{eff}$  potential. The combining rules for unlike pseudoatom types  $i$  and  $j$  in the  $vdW_{eff}$  potential are:  $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ ,  $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ , and  $\gamma_{ij} = (\gamma_i + \gamma_j)/2$ .

$$E_{vdW_{eff}} = \frac{2\varepsilon}{1 - \frac{3}{\gamma+3}} \left( \frac{\sigma^6}{\sigma^6 + r^6} \right) \left[ \frac{3}{\sigma^6 + r^6} e^{\gamma(1 - \frac{r}{\sigma})} - 1 \right] \quad (3)$$

**Hydrogen Bond and Electrostatics Energies.** The hydrogen bond (Eq. 4) and Debye-Huckel electrostatics (Eq. 5) potential energy terms are of the same form as used previously. However, we reparametrized the hydrogen bond and Debye-Huckel potentials with the introduction of the new  $vdW_{eff}$  term. In the hydrogen bond potential  $\varepsilon_{hb,max}$  is the maximum potential found at the hydrogen bond equilibrium distance  $\sigma_{hb,eq} |r_{ji}^-|$  is the magnitude of the vector from atom  $j$  to atom  $i$ , while  $\alpha_k = 2(\theta_i + \theta_j) - \pi$  is a directional component with  $\theta_i$  and  $\theta_j$  defined in



**Figure 2.**  $vdW_{eff}$  potential. (a) Functional form. (b) Effective potential compared to standard Lennard Jones and Buckingham potentials with minimum energy potential  $\varepsilon = 0.5$  kcal/mol, minimum energy distance,  $\sigma = 4$  Å, and gamma of effective potential  $\gamma = 10$ . (c) Effect of changing value of minimum energy distance,  $\sigma$  (d) Effect of changing minimum energy potential,  $\varepsilon$  (e) Effect of changing the short range behavior with parameter  $\gamma$ . For (c–e), unless stated,  $\varepsilon = 0.5$  kcal/mol,  $\sigma = 4$  Å, and  $\gamma = 10$ .

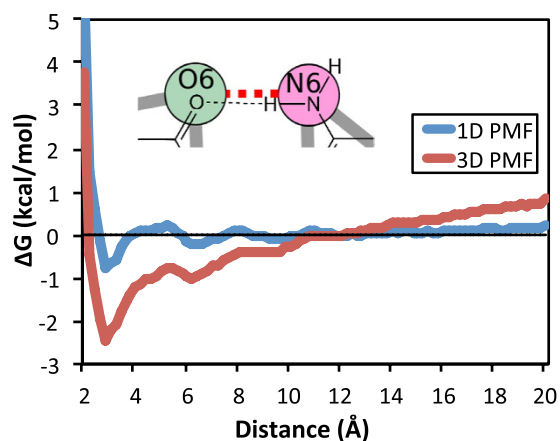
Fig. S1. For hydrogen bond parameterization, the maximum potential  $\varepsilon_{hb,max}$  was increased from 0.5 kcal/mol to 2.0 kcal/mol. Other hydrogen bond parameters including equilibrium distance  $\sigma_{hb,eq}$  of 2.9 Å and cutoff of 6 Å (base edge) remain the same as the previous model. Hydrogen bond potential energy is computed for both canonical (GC, AU) and noncanonical base pairs. For Debye-Huckel Eq. 5,  $q_i$  is the charge of atom  $i$ ,  $r_{ij}$  is the distance between atom  $i$  and atom  $j$ ,  $D$  is the dielectric constant, and  $\xi$  is the Debye length. A dielectric constant  $D$  of 25 was determined to be optimal under the new model potential, compared to 78 from the previous model. In depth discussion of Debye-Huckel and hydrogen bond optimization can be found in the SI.

$$E_{hb} = -\frac{\varepsilon_{hb,max}}{2} (1 - \cos(\alpha_k)) \left( \frac{\sigma_{hb,eq}}{r_{ij}} \right)^3 \quad (4)$$

$$E_{ele} = \frac{q_i q_j}{4\pi D} r_{ij}^{-1} e^{-r_{ij}/\xi} \quad (5)$$

**Model improvement and Parameterization.** *Statistical potentials.* The premise of our parameter optimization was to fit to both RNA structure and experimental free energies. First, we updated model statistical potentials from experimentally determined crystal structures. We downloaded all available Protein Data Bank (PDB, <http://www.rcsb.org/>) RNA structures as of Feb. 10, 2015, (excluding RNA-protein and RNA-DNA combination structures) totaling ~1100 entries. Our previous model fit to statistical potentials used approximately 668 structures. For RACER, our updated parameterization includes an additional ~400 structures, which led to various modifications in the potentials. The method of statistical potentials involves fitting energy functions to statistically derived potential of mean force (PMF) curves. The PMFs are determined by taking the probability distribution  $P(r)$  of occurrences from the PDB structure set and then extracting the free energy  $G(r)$ ,  $G(r) = -k_B T \ln(P(r)/ref)$  with the reference distribution  $ref$  setting the minimum interaction at 0 kcal/mol.

One of the major improvements in the current model is to adopt a new nonbonded effective potential form to capture the intricate short range behavior observed in nonbonded statistical potentials that standard vdW potential forms (including our model's previously used Buckingham potential) cannot capture. The Buckingham and other common vdW functions are too stiff at short range with a steep slope, whereas the nonbonded statistical



**Figure 3.** Comparison of a 1-D and 3-D statistical potential PMF, computed from a 1-D and 3-D radial distribution function respectively. Note how the 3-D PMF continues to diverge at long distances whereas the 1-D PMF falls off to zero. This statistical potential is that of RACER base pseudoatoms O6-N6.

potentials reveal much softer behavior. We have identified a more “flexible”  $\text{vdW}_{\text{eff}}$  potential that better captures this short range behavior, which is critical for local packing of RNA molecules. When we implemented the new potential, it was also necessary to re-parameterize the torsion, electrostatics, and hydrogen bond interactions for consistency.

**1-D PMF for RNA.** In this work, we determined that modeling RNA as one-dimensional rather than a three-dimensional, isotropic molecule is more appropriate when extracting the statistical potentials from PDB structures. This choice is justified as there is an abundance of short, linear helices found in PDB structures of RNA. Additionally, folded RNA typically forms prolate ellipsoids<sup>95</sup>. Similarly, in the PDB structure of 16S rRNA more than half of the nucleotides are base paired<sup>24</sup>. Therefore, treating RNA as a one-dimensional molecule for capture of local interactions is not unreasonable. Additionally, 3D PMFs are more appropriate for systems with isotropic distance distributions, such as molecular liquids<sup>96–98</sup> and proteins<sup>99–101</sup>.

Our motivation for modeling RNA as a 1D molecule came from the observation of divergence of 3D radial distribution functions (RDF) at distances greater than 10 Å, and as a result the potential of mean force (PMF) that was derived from the RDF did not converge to zero at large separation (see Fig. 3). The cause of this divergence at long distances is the inherent volumetric effect of the 3D RDF, while the PDB structures we sample are mostly small and linear. The statistical potentials do include some larger ribosomal structures, but these are too few to cause the observed divergence. Contrary to 3D RDFs, when 1D radial distribution functions were used the PMF asymptotically approached zero for long distances (see Fig. 3), reinforcing the discussion that the set of RNAs used here in statistical potentials can be adequately sampled as linear 1D, rather than 3D RDFs. The main difference between 3D and 1D RDFs is the normalization factor. For 3D RDFs, normalization is done over a volumetric shell  $4\pi r^2 dr$ , whereas 1D RDFs normalizes over an incremental distance,  $dr$ .

Specifically, the non-bonded PMF is evaluated via Boltzmann inversion as  $G(r) = -k_B T \ln(g(r))$  where  $g(r)$  is the radial distribution function, normalized probability function discussed above. When treating RNA as a 3D isotropic molecule, the 3D RDF, as was done previously<sup>90,91</sup>, is given by  $g(r) = n_{ij}(r)/[(N_i N_j / V) 4\pi r^2 dr]$ , where  $n_{ij}(r)$  is the number of atom type  $j$  at distance  $r$  from atom type  $i$ ,  $N_i$  and  $N_j$  are the total number of  $i$  and  $j$  atoms respectively, and  $V$  is the volume of the system. Now we treat RNA as a “1D”, linear molecule to more adequately parameterize the  $\text{vdW}_{\text{eff}}$  potential, and the RDF becomes  $g(r) = n_{ij}(r)/[(N_i N_j / V) dr]$ .

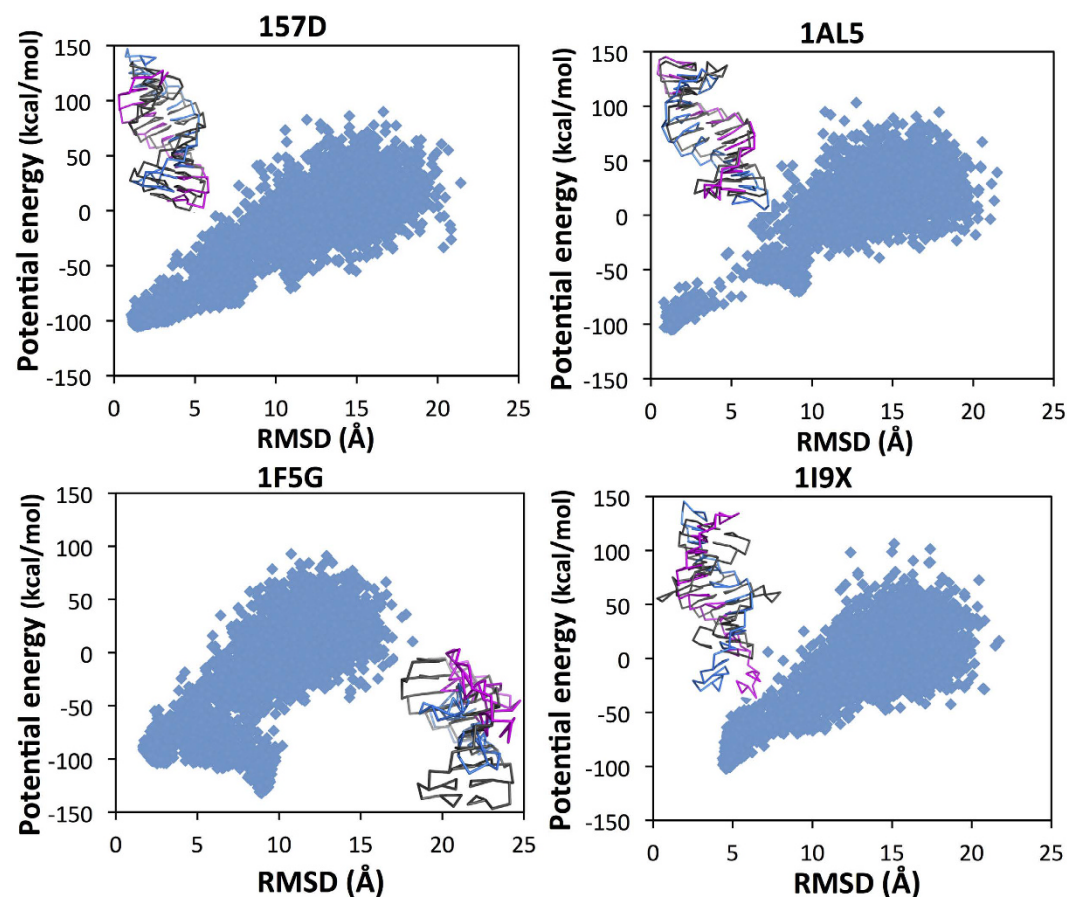
**Structure Prediction.** *Folding RNA by simulated annealing.* We tested RACER with simulated annealing simulations to (1) validate that RACER can accurately fold experimentally determined RNA structures and to (2) ensure the native structure has the lowest energy on its energy landscape. We ran simulated annealing simulations on a testing set of 14 RNAs, duplexes and hairpins, that have known experimentally determined structures<sup>90</sup>. This test set of 14 RNAs was included as part of the 1100 structures used to compute our statistical potentials; however, the contributions of the 14 RNA test set (~1% of training set) to the statistical potentials and thus parameterization is negligible. From annealing simulations on this set of 14 RNAs, RACER is able to predict 13 out of 14 RNA molecules with RMSD < 5 Å, and 6 RNA molecules with RMSD < 2.5 Å. The average RMSD between the predicted lowest-energy structures and native structures is 2.93 Å. This average RMSD is improved from our previously published average RMSD of 3.31 Å; additionally, our model now has the capability to predict free energy landscapes of RNA in addition to structure prediction.

The simulated annealing protocol involved running MD sequentially for 5 ns at temperatures in order of 298(K), 400, 1000, 900, 800, 700, 600, 500, 400, 298 K, for a total simulation time of 50 ns, with structures saved every 10 ps. Given the high temperatures used, we used a 1fs time step for annealing simulations. Results for structure prediction using simulated annealing are given in Table 1. These predicted RMSD values are calculated between PDB structures and the minimum potential energy structures found by RACER.



PDB ID	157D	1AL5	1DQF	1F5G	1I9X	1KD5	1LNT	1QCU	1ZIH	2AO5	2JXQ	2K7E	353D	472D	Avg.
RMSD (Å)	1.45	1.26	3.04	8.91	4.56	2.88	2.65	0.96	3.11	1.83	1.13	3.58	2.24	3.47	2.93 ± 0.54

**Table 1.** Predicted (minimum energy) RMSD values compared to Protein Data Bank (PDB) structures from simulated annealing.



**Figure 4.** Representative energy landscapes from annealing for two RNAs that are accurately predicted: accurately predicted: 157D and 1AL5 (top), and two RNAs that are unfavorably predicted: 1F5G and 1I9X (bottom). For each RNA, the RACER minimum free energy structure is shown in blue and magenta aligned to the PDB structure shown in black. Five thousand structures over 50 ns are shown for each RNA; each structure is energy minimized before plotting. Note the funnel toward low energy and low RMSD structures. The RMSD of lowest energy structure for 157D is 1.45 Å, 1AL5 is 1.26 Å, 1F5G is 8.91 Å, and 1I9X is 4.56 Å.

**Energy landscapes.** Analyzing the energy landscapes of the 14 RNAs in our training set was an important part of our optimization. RNAs are complex molecules that may adopt stable and long lived misfolded structures. However, it is assumed the final native structures, at least *in vitro*, should have the lowest free energy for the given environment<sup>102</sup>. Here, annealing simulations are used to generate a large number of unfolded structures for each RNA. Each of these structures is then energy minimized to 0 K. The energy and RMSD (with respect to the native structure) of each structure are used to characterize the energy landscape. The energy-RMSD landscapes for all 14 RNAs are given in SI, Table S1.

The energy vs RMSD landscapes for all 14 RNAs show clear “funnel” shapes skewed toward the native structure. As examples, we present the energy landscapes for two favorably predicted structures (157D and 1AL5, 1.45 Å and 1.26 Å RMSD respectively) in Fig. 4a,b, and the energy landscapes for the two most unfavorably predicted structures (1F5G and 1I9X, 8.91 Å and 4.56 Å RMSD respectively), where the lowest energy structures have large RMSD in Fig. 4c,d.

RACER predicted structures for PDB ID: 157D, 1AL5, 1F5G, and 1I9X are shown in Fig. 4. RACER predicted structures 157D and 1AL5 agree well with experiment (inset in Fig. 4a,b). The RACER predicted structure for 1F5G (8.91 Å RMSD) has collapsed into a torus-like structure, with very little backbone twist (inset in Fig. 4c,d). A possible explanation for this observed behavior is the non-canonical base pairing present in 1F5G. While RACER can capture non-canonical base pairing through the hydrogen bond potential, these hydrogen

bonds need further calibration relative to canonical interactions. The RACER predicted structure for 1I9X (4.56 Å RMSD) forms an extended helix compared to the crystal structure. This is likely due to two bases flipped out of the helix in the crystal structure, while RACER incorporates these bases back into the helix. In the crystal structure for 1I9X, several water molecules stabilize these bases. In the RACER model, this stabilization is challenging to capture due to the implicit treatment of solvent via the Debye-Huckel potential.

Additionally, energy landscapes allow us to identify possible meta-stable intermediates, which are high-RMSD (~8 Å) “local” funnels observed in plots for 1DQF and 1QCU. The meta-stable structure of 1DQF at the local minimum, shown in Fig. S2 resembles the toroidal structure observed for 1F5G, but for 1QCU an extended, base stacking meta-stable structure is observed. For 1DQF, the local funnel structure has increased torsional potential energy (~30 kcal/mol) over the global-minimum structure, although both have similar vdW<sub>eff</sub> and hydrogen bond potentials. For 1QCU, the local funnel structure has a more stabilizing hydrogen bond potential (~–15 kcal/mol) than the global-minimum structure; however, in the global-minimum structure, the Debye-Huckel electrostatics and vdW<sub>eff</sub> potentials compensate hydrogen bonds to result in an overall more stabilizing intermolecular energy than the local-funnel structure. It is important to note that for both of these RNAs, the RACER global-minimum structure is very close to the experimental structures.

In the process of validating and optimizing our model by energy landscape analysis, we noticed the importance of a dedicated hydrogen bond potential for base pairing, as the vdW<sub>eff</sub> potential is not well suited for distinguishing between base stacking and base pairing interactions<sup>103</sup>. The hydrogen bond potential allows for directional base pairing and helps in separating the base stacking and base pairing interactions effectively.

**Equilibrium Pulling Simulations.** *Experimental free energies.* To test RACER, we focused on capturing experimental melting free energies of canonical helices<sup>104</sup> and hairpins<sup>105</sup>. We used RACER to perform equilibrium pulling simulations, and we compared free energy differences to two sets of experimental thermodynamic data: RNA melting free energies from Turner and coworkers<sup>28</sup> and folding free energies from single molecule force experiments. Five hairpins of size 10, 10, 12, 14, and 18 nt and five duplexes of size 6, 6, 8, 8, and 10 base pairs were selected from melting free energy experiments, and the TAR RNA hairpin was chosen to compare RACER to single molecule force experiments. Hairpin sequences 30, 11, 33, 47, and 19 from the Supplementary Information of<sup>105</sup> are referred to here as h1, h2, h3, h4, and h5, and duplex sequences 35, 48, 71, 78, and 90 of<sup>104</sup> are referred to here as d1, d2, d3, d4, and d5. TAR is a 52 nt, 21 bp hairpin with two internal loops.

In melting free energy experiments, a solution of RNAs of known sequence are heated while measuring UV absorption. As helical and single stranded RNAs absorb light at different wavelengths, the absorption will change over heating as the RNA denatures. By fitting a curve to absorption vs temperature the melting free energy can be determined<sup>106–108</sup>. Turner and co-workers have published a compendium of melting free energies for small RNA motifs and structures using nearest neighbor energy parameters and RNA secondary structure prediction<sup>28,104,109</sup>. Additionally, we compared our model to RNA single molecule force experiments.

In single molecule force experiments, folded RNA molecules are unfolded by mechanical force using techniques such as optical tweezers or atomic force microscopy. Using the end-to-end extension as a reaction coordinate, the free energy of unfolding can be determined from position vs. time data. A recent single molecule research study of the trans activation response (TAR) element of HIV extracted the free energy of folding at zero force under the assumption of the worm-like chain model<sup>110</sup>. Here we study the same TAR RNA as used in the single molecule force experiments.

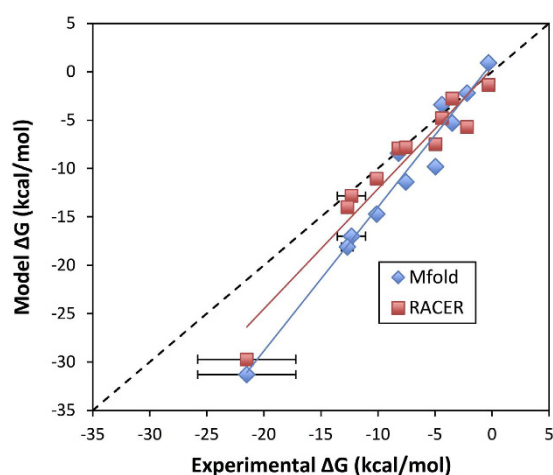
Melting and pulling experiments for all RNAs were simulated by umbrella sampling simulations pulling the RNAs apart from their ends (see Fig. S3 for example simulation setup showing end-to-end reaction coordinate). Free energy values were then computed using the Weighted Histogram Analysis Method (WHAM) software distributed by Alan Grossfield<sup>111</sup>. Details of these simulations are included in the Methods section. Although exact energy landscapes at equilibrium for both TAR and melting free energy helices are unknown, folding free energies can be computed according to Eq. 6. The folded free energy,  $\Delta G$ , is found by integrating over all folded conformations at end-to-end extension  $r$  with free energy  $\Delta\omega$ . Folded free energy is then normalized to volumetric entropy, with standard state volume  $V_{ref}$  of  $1660 \text{ \AA}^3$ .  $kT$  is the Boltzmann constant multiplied by temperature (298 K).

$$\Delta G = kT \ln(8\pi^2 V_{ref}) - kT \ln \left( \int 4\pi r^2 dr e^{-\frac{\Delta\omega(r)}{kT}} \right) \quad (6)$$

*Unfolding free energies from RACER MD simulations.* The free energies computed from equilibrium pulling MD simulations (WHAM) using RACER are in excellent agreement with experimental measurements, with a correlation coefficient ( $R^2$ ) of 0.93 for 11 RNAs tested (Table 2 and Fig. 5). For additional comparison, we also included the melting free energies from Mfold, a widely-used secondary structure prediction program that has been parameterized using the experimental melting thermodynamic data (Mfold predicted structures are shown in Fig. S4). The unfolding free energies evaluated by RACER and Mfold<sup>30</sup> are presented in Table 2 along experimental values and the length of each MD simulation. The correlation plots for RACER and Mfold show both models have close  $R^2$  correlation coefficients of 0.93 and 0.96 respectively. However, Mfold’s linear fit has a slightly higher slope (1.5) than RACER (1.2) as Mfold over predicts the stability of the duplexes. Note that RACER is a 3D particle based physical model developed for molecular dynamics simulations, whereas Mfold predicts secondary structures from sequences based on nearest neighbor energy parameters. In RACER we explicitly compute the entropy contributions to the free energy through molecular dynamics sampling.

Hairpin	Length (nt)	Expt. $\Delta G$ (kcal/mol)	Mfold $\Delta G$ (kcal/mol)	RACER $\Delta G$	Length per window (kcal/mol)
h1	10	$-3.5 \pm 0.3^{105,118}$	-5.3	$-2.8 \pm 0.12$	1 $\mu$ s
h2	10	$-0.3 \pm 0.1^{105,119}$	+0.9	$-1.4 \pm 0.12$	1 $\mu$ s
h3	12	$-4.4 \pm 0.2^{105,120}$	-3.4	$-4.8 \pm 0.14$	1 $\mu$ s
h4	14	$-2.2 \pm 0.08^{105,121}$	-2.2	$-5.7 \pm 0.15$	1 $\mu$ s
h5	18	$-8.2 \pm 0.2^{105,122}$	-8.4	$-7.9 \pm 0.22$	1 $\mu$ s
TAR	52	$\approx -21.5 \pm 4.3^{110}$	-31.3	$-29.7 \pm 0.36$	0.1 $\mu$ s
Duplex	Length (bp)				
d1	6	$-7.56 \pm 0.3^{104}$	-11.4	$-7.8 \pm 0.15$	1 $\mu$ s
d2	6	$-4.95 \pm 0.2^{104}$	-9.8	$-7.5 \pm 0.14$	1 $\mu$ s
d3	8	$-12.32 \pm 1.2^{104}$	-17.0	$-12.8 \pm 0.17$	1 $\mu$ s
d4	8	$-10.11 \pm 0.2^{104,106}$	-14.7	$-11.1 \pm 0.18$	1 $\mu$ s
d5	10	$-12.69 \pm 0.5^{104}$	-18.1	$-14.0 \pm 0.19$	1 $\mu$ s
					Total: 860 $\mu$ s

**Table 2. Unfolding free energy values for RNAs from experiment (Expt.), Mfold predicted, and RACER predicted.** Molecule length in nucleotides or basepairs and the simulation length per window are also shown. Error is taken from a Monte Carlo bootstrap error analysis as implemented in the WHAM program by Grossfield<sup>111</sup>.

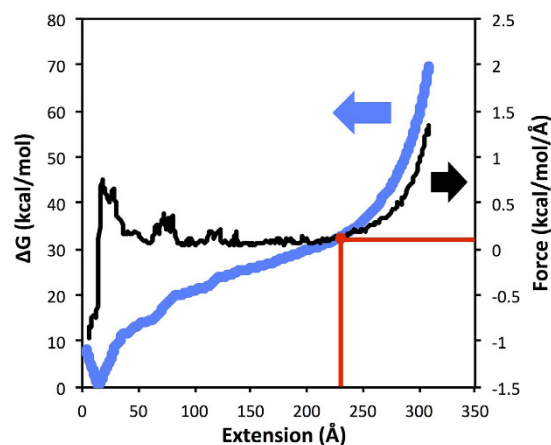


**Figure 5. Correlation plot between predicted free energy from RACER and experimental free energy in kcal/mol.** RACER simulation predicted free energy is compared with Mfold minimum free energy as well as unity slope (dashed line). RACER and Mfold have the same correlation free energy predictive capability ( $R^2 = 0.93$  to experimental free energies, experiment), but RACER has a slope closer to unity (slope = 1.2), while Mfold over-stabilizes the free energy of larger RNAs (slope = 1.5). Error bars present on RACER data come from a Monte Carlo bootstrap error analysis as implemented in WHAM by Alan Grossfield<sup>111</sup> (most errors are within the data point).

**Pulling generated RNA structures.** Ensemble model structures for folded states are shown in Figs S5 and S6. In the folded states, TAR, h4 and h5 are observed to form helices while h1–h3 form base pairs and stacking interactions but without regular helical structure. For duplexes, the two RNA strands form canonical base pairs resulting in proper helices. The terminal nucleotides of d5 are observed to break base pairing with one nucleotide rotating out of the helix while the other remains stacked, but this is also observed in experiment<sup>112</sup>.

In pulling experiments, free energy vs end-to-end extension plots show two distinct energy minima corresponding to folded and unfolded states<sup>113–115</sup>. In the RACER model unfolded (extended) states remain stabilized by  $vdW_{\text{eff}}$  base stacking interactions, so the location of unfolded free energy is difficult to determine directly from free energy landscapes of RNAs. While the free energy landscapes predicted by RACER show an energy well around the folded state, there is a flat to monotonically increasing curve observed at large extensions (Figs 6 and 7, blue curve, also see Fig. S7). The location of the unfolded state is paramount to computing the folded free energy  $\Delta G$  using Eq. 6. To determine unfolded state location, we plotted the gradient of the free energy, the ‘force’ as a function of extension (Figs 6 and 7, black curve). From these force vs. extension plots, the predicted free energy of the unfolded state was taken to be the free energy value where the force is very low ( $\sim 0.1$  kcal/mol/Å), i.e. before the RNA reaches the over-stretched regime (Figs 6 and 7, red lines). A 4 Å running average of ‘force’ over extension was used to eliminate noise (Figs 6 and 7). Histogram figures showing equal sampling of





**Figure 6.** The equilibrium pulling free energy profile (blue) of TAR hairpin computed with WHAM using the RACER model (see Method section details). The unfolded state is determined as the state right before the force (derivative of the free energy, curve shown in black) sharply increases from low ( $<0.1$  kcal/mol/Å) to high due to overstretching,  $0.1$  kcal/mol/Å and the location of the unfolded state are denoted by the red lines. The calculated folding free energy for TAR is  $-29.7 \pm 0.36$  kcal/mol, compared to the experimental value of  $\approx -21.5$  kcal/mol. A  $4$  Å running average of force (black curve) is shown to eliminate noise.

the pulling windows are included in Figs S8 and S11. Additionally, the uncertainty of the free energy landscape as computed by a Monte Carlo bootstrap error analysis in the WHAM program by Alan Grossfield<sup>111</sup> is shown as a range in Figs S12 and S14.

## Discussion

**Statistical potential summary.** RACER, a coarse-grained RNA model, can accurately predict native structures and capture RNA folding free energy. The functional forms and parameters in RACER were determined by systematic optimization against native structures and melting free energies for a number of RNA molecules. We found that the statistical potentials<sup>92</sup> used in the previous model were over stabilizing and the 3D PMFs diverged at long distances. As a result, we treat RNA as a one-dimensional rather than three-dimensional molecule, and use a 1D RDF when fitting to PMFs. Our optimization procedure led us to incorporate a more general effective van der Waals potential energy function ( $vdW_{\text{eff}}$ ) to describe the interactions among pseudoatoms.

As a result of implementing a new non-bonded potential energy, we have also reparametrized both electrostatic and hydrogen bond potential energy functions. As the RNA backbone is highly charged, a Debye-Huckel electrostatics term is included for each phosphate pseudoatom; a dielectric of 25 was chosen in order to capture both folded and unfolded RNA structures. A directional hydrogen bond potential was reparametrized in order to accurately distinguish base pairing (hydrogen bond, some  $vdW_{\text{eff}}$ ) and base stacking ( $vdW_{\text{eff}}$ ) interactions. We found that the hydrogen bond potential was pivotal to accurate folding free energies as both folded and unfolded RNA have base stacking interactions, while only folded RNA have base pairing (hydrogen bond) interactions.

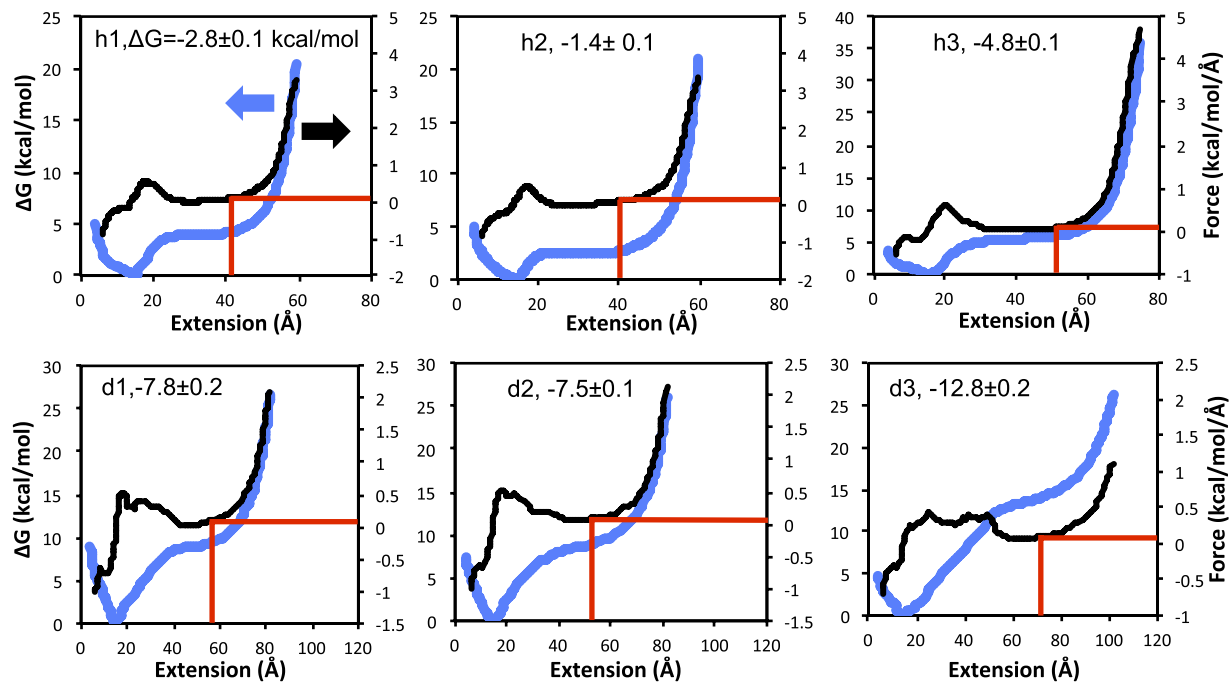
**Thermodynamic summary.** For a structure prediction model, thermodynamic accuracy is important to ensure that the energy landscape correctly represents RNAs with varying size and sequence. Our energy landscape analysis suggests that even relatively small RNAs may have complex energy landscapes, and there are many RNA structures at low potential energy. Therefore, explicit consideration of entropy through techniques such as MD is crucial to capture the free energy landscapes of RNA structures.

Folding free energy values for six RNA hairpins of size 10–52 nts and five duplexes of size 6–10 bp were determined by umbrella sampling simulations with WHAM-computed free energy. For hairpins, we determined that umbrella sampling simulations with a reaction coordinate of end-to-end extension is appropriate for capturing folding free energy. For duplexes, the same protocol is found to be appropriate, with the addition of a restraint preventing the single strands from long-lasting intra-strand interactions (e.g. hairpin-like structures). Pulling free energy landscapes of hairpins and duplexes clearly revealed the folded state and we used the gradient (force) of pulling free energy to define the location of the unfolded state.

Given the low computational cost of RACER, over 0.8 ms of umbrella sampling and simulated annealing simulations are presented. Overall, the MD-calculated free energy results using the RNA model are in excellent agreement ( $R^2 = 0.93$ ) with experimental folding free energy values while preserving accurate structure prediction. In this work, we present RACER, a novel RNA coarse-grained model that captures both RNA structure and thermodynamics for increased utility to RNA folding investigations.

## Methods

**Mapping from all-atom to coarse-grained structures.** A notable feature of our model is the ability to map to and from all-atom experimental crystal structures. Each of our model's pseudoatoms represents an atomic site in nucleotides; for example, the sugar pseudoatom is assigned the C4' atom position on ribose. Moreover, our model captures the planarity of the nucleobase with three pseudoatoms. Given a novel (structure undetermined)



**Figure 7.** The equilibrium pulling free energy profile (blue) of hairpins h1–h3 (top) and duplexes d1–d3 (bottom) computed with WHAM using the RACER model (h4–h5 and d4–d5 are given in Fig. S7). Umbrella sampling pulling simulations were run for 1  $\mu$ s for each window, with a 1  $\text{\AA}$  window separation. The unfolded state is determined as the state right before the force (derivative of the free energy, curves shown in black) sharply increases from low ( $<0.1$  kcal/mol/ $\text{\AA}$ ) to high due to overstretching. 0.1 kcal/mol/ $\text{\AA}$  and the location of the unfolded state are denoted by the red lines. The folding free energy ( $\Delta G_f$ , kcal/mol) is included for each RNA. A 4  $\text{\AA}$  running average of force (black curves) is shown to eliminate noise.

RNA sequence, our model can first predict the three-dimensional structure in coarse-grained coordinates and then map to all-atom coordinates with further minimization, producing an equivalent to an all-atom experimentally determined structure. As a result, our RNA model is well suited to perform multiscale simulations in the future.

**Pulling methods.** Melting and pulling experiments are modeled by using umbrella simulations pulling the RNA molecule apart from its terminal ends. A harmonic potential of 1 kcal/mol/ $\text{\AA}^2$  spring constant is used to restrain the RNA ends at the sugar pseudoatoms (C4' sugar atomic site). Simulation extensions ran from 5.5  $\text{\AA}$  up to fully extended lengths (59.5, 76.5, 86.5, 106.5, and 307.5  $\text{\AA}$  for 10, 12, 14, 18, and 52 nt hairpins assuming 5.9  $\text{\AA}$  per nt contour length) with a spacing of 1  $\text{\AA}$  between windows.

Duplexes are similarly pulled apart from the sugar pseudoatoms at one terminal end with a 1 kcal/mol/ $\text{\AA}^2$  spring constant; the other terminal end is restrained between two terminal sugar pseudoatoms with a 1 kcal/mol/ $\text{\AA}^2$  spring constant. Duplex extensions ranged from 5.5 Angstroms up to fully extended lengths (80.5, 100.5, and 124.5 Angstroms for 6, 8, and 10 base pair duplexes respectively) with umbrella window spacing of 1  $\text{\AA}$ . For the duplexes and shorter hairpins of size 10 and 18 nt, 1  $\mu$ s of Molecular Dynamics was run for each window. For the TAR hairpin, 100 ns was found to be sufficient given the longer end-to-end extension (more windows) needed. We used a 4 fs time step for pulling simulations. From the umbrella simulations, the free energy landscapes were computed by the Weighted Histogram Analysis Method<sup>116</sup> (WHAM) using the program distributed by Alan Grossfield<sup>111</sup>.

**Computational efficiency of the RACER Model.** All annealing and pulling simulations (total of 0.86 ms) were computed on a local computer cluster. For all simulations discussed below a 4 fs time step was used, and the CPUs used are an early generation Intel Xeon E5345 2.33 GHz CPU. Using one CPU core for each simulation, 1  $\mu$ s of simulation of the 10 nt hairpin h1 took 22 hours, 1  $\mu$ s of simulation of the 18 nt hairpin h3 took ~60 hours, and 100 ns of simulation of the 52 nt hairpin TAR took ~48 hours. Additionally, 1  $\mu$ s simulation of duplex d35 required 30 hours, while 1  $\mu$ s for duplex d90 required 74 hours. Recently, RACER has been implemented with OpenMP allowing parallelization to multiple cores. In the future, we will implement our model on GPUs, using the software package OpenMM<sup>117</sup>. Implementation of RACER on GPUs will allow for even better efficiency. As a result of the improved computational efficiency offered by the coarse-graining, it will be possible to simulate RNAs at physiologically relevant timescales.

**Implementation and parameters.** The TINKERMD implemented RACER model is available free of charge at <http://biomol.bme.utexas.edu/tinker-openmm/index.php/TINKER-OPENMM:Development-rna>. The parameters and conversion programs are included in the distribution. Conversion tutorials are posted online at <http://biomol.bme.utexas.edu/tinker-openmm/index.php/TINKER-OPENMM:Tutorials-rna>.

## References

- Cech, T. R., Zaug, A. J. & Grabowski, P. J. splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496, [http://dx.doi.org/10.1016/0092-8674\(81\)90390-1](http://dx.doi.org/10.1016/0092-8674(81)90390-1) (1981).
- Kruger, K. *et al.* Self-Splicing Rna - Auto-Excision and Auto-Cyclization of the Ribosomal-Rna Intervening Sequence of Tetrahymena. *Cell* **31**, 147–157, doi: 10.1016/0092-8674(82)90414-7 (1982).
- Guerriertakada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety Of ribonuclease-P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857, doi: 10.1016/0092-8674(83)90117-4 (1983).
- Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* **111**, 747–756, doi: 10.1016/s0092-8674(02)01134-0 (2002).
- Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043–1049, doi: 10.1016/s1074-5521(02)00224-7 (2002).
- Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956, doi: 10.1038/nature01145 (2002).
- Breaker, R. R. Prospects for Riboswitch Discovery and Analysis. *Mol. Cell* **43**, 867–879, doi: 10.1016/j.molcel.2011.08.024 (2011).
- Serganov, A. & Nudler, E. A Decade of Riboswitches. *Cell* **152**, 17–24, doi: 10.1016/j.cell.2012.12.024 (2013).
- Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA-Publ. RNA Soc.* **19**, 1461–1473, doi: 10.1261/rna.037390.112 (2013).
- Russell, R. In *Biophysics of RNA Folding Biophysics for the Life Sciences* (ed. R. Russell) Ch. 1, 1–10 (Springer-Verlag New York, 2013).
- Mitchell, D., Jarmoskaite, I., Seval, N., Seifert, S. & Russell, R. The Long-Range P3 Helix of the Tetrahymena Ribozyme Is Disrupted during Folding between the Native and Misfolded Conformations. *J. Mol. Biol.* **425**, 2670–2686, doi: 10.1016/j.jmb.2013.05.008 (2013).
- Mitchell, D. & Russell, R. Folding Pathways of the Tetrahymena Ribozyme. *J. Mol. Biol.* **426**, 2300–2312, doi: 10.1016/j.jmb.2014.04.011 (2014).
- Russell, R. *et al.* The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J. Mol. Biol.* **363**, 531–544, doi: 10.1016/j.jmb.2006.08.024 (2006).
- Russell, R. *et al.* Exploring the folding landscape of a structured RNA. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 155–160, doi: 10.1073/pnas.221593598 (2002).
- Thirumalai, D. & Hyeon, C. In *Non-Protein Coding RNAs* (eds Nils G. Walter, Sarah A. Woodson & Robert T. Batey) 27–47 (Springer Berlin Heidelberg, 2009).
- Silverman, S. K., Deras, M. L., Woodson, S. A., Scaringe, S. A. & Cech, T. R. Multiple Folding Pathways for the P4–P6 RNA Domain. *Biochemistry* **39**, 12465–12475, doi: 10.1021/bi000828y (2000).
- Woodson, S. A. Recent insights on RNA folding mechanisms from catalytic RNA. *Cell. Mol. Life Sci.* **57**, 796–808, doi: 10.1007/s000180050042 (2000).
- Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. *Nature Reviews Molecular Cell Biology* **5**, 908–919, doi: 10.1038/nrm1497 (2004).
- Bokinsky, G. & Zhuang, X. W. Single-molecule RNA folding. *Accounts Chem. Res.* **38**, 566–573, doi: 10.1021/ar040142o (2005).
- Gell, C. *et al.* Single-Molecule Fluorescence Resonance Energy Transfer Assays Reveal Heterogeneous Folding Ensembles in a Simple RNA Stem-Loop. *J. Mol. Biol.* **384**, 264–278, doi: 10.1016/j.jmb.2008.08.088 (2008).
- Uhlenbeck, O. C. Keeping RNA happy. *RNA-Publ. RNA Soc.* **1**, 4–6 (1995).
- Uhlenbeck, O. C. RNA biophysics has come of age. *Biopolymers* **91**, 811–814, doi: 10.1002/bip.21269 (2009).
- Schuster, P. Prediction of RNA secondary structures: from theory to models and real molecules. *Rep. Prog. Phys.* **69**, 1419–1477, doi: 10.1088/0034-4885/69/5/r04 (2006).
- Cannone, J. J. *et al.* The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *Bmc Bioinformatics* **3**, doi: 10.1186/1471-2105-3-2 (2002).
- Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R. & Stadler, P. F. RNAalifold: improved consensus structure prediction for RNA alignments. *Bmc Bioinformatics* **9**, 13, doi: 10.1186/1471-2105-9-474 (2008).
- Hofacker, I. L., Fekete, M. & Stadler, P. F. Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* **319**, 1059–1066, doi: 10.1016/s0022-2836(02)00308-x (2002).
- Knudsen, B. & Hein, J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Research* **31**, 3423–3428, doi: 10.1093/nar/gkg614 (2003).
- Turner, D. H. & Mathews, D. H. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research* **38**, D280–D282, doi: 10.1093/nar/gkp892 (2010).
- Markham, N. & Zuker, M. In *Bioinformatics Vol. 453 Methods in Molecular Biology*<sup>TM</sup> (ed. Jonathan M. Keith) Ch. 1, 3–31 (Humana Press, 2008).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415, doi: 10.1093/nar/gkg595 (2003).
- Hofacker, I. L. *et al.* Fast folding and comparison of RNA secondary structures. *Mon. Chem.* **125**, 167–188, doi: 10.1007/bf00818163 (1994).
- Hofacker, I. In *Comparative Genomics Vol. 395 Methods in Molecular Biology*<sup>TM</sup> (ed. Nicholas H. Bergman) Ch. 33, 527–543 (Humana Press, 2008).
- Lorenz, R. *et al.* ViennaRNA package 2.0. *Algorithms for Molecular Biology* **6**, 1–14, doi: 10.1186/1748-7188-6-26 (2011).
- Doshi, K. J., Cannone, J. J., Cobaugh, C. W. & Gutell, R. R. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 1–22, doi: 10.1186/1471-2105-5-105 (2004).
- Bellaousov, S. & Mathews, D. H. ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. *RNA* **16**, 1870–1880, doi: 10.1261/rna.2125310 (2010).
- Ren, J., Rastegari, B., Condon, A. & Hoos, H. H. HotKnots: Heuristic prediction of RNA secondary structures including pseudoknots. *RNA* **11**, 1494–1504, doi: 10.1261/rna.7284905 (2005).
- Sato, K., Kato, Y., Hamada, M., Akutsu, T. & Asai, K. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* **27**, i85–i93, doi: 10.1093/bioinformatics/btr215 (2011).
- Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2[prime]-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protocols* **1**, 1610–1616, doi: 10.1038/nprot.2006.249 (2006).

39. Lusvarghi, S., Sztuba-Solinska, J., Purzycka, K. J., Rausch, J. W. & Le Grice, S. F. J. RNA Secondary Structure Prediction Using High-throughput SHAPE. e50243, doi: 10.3791/50243 (2013).
40. Leonard, C. W. *et al.* Principles for Understanding the Accuracy of SHAPE-Directed RNA Structure Modeling. *Biochemistry* **52**, 588–595, doi: 10.1021/bi300755u (2013).
41. Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. Understanding the Errors of SHAPE-Directed RNA Structure Modeling. *Biochemistry* **50**, 8049–8056, doi: 10.1021/bi200524n (2011).
42. Sükkösd, Z., Swenson, M. S., Kjems, J. & Heitsch, C. E. Evaluating the accuracy of SHAPE-directed RNA secondary structure predictions. *Nucleic Acids Research* **41**, 2807–2816, doi: 10.1093/nar/gks1283 (2013).
43. Lorenz, R., Luntzer, D., Hofacker, I. L., Stadler, P. F. & Wolfinger, M. T. SHAPE directed RNA folding. *Bioinformatics* **32**, 145–147, doi: 10.1093/bioinformatics/btv523 (2016).
44. Hajdin, C. E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences* **110**, 5498–5503, doi: 10.1073/pnas.1219988110 (2013).
45. Laing, C. & Schlick, T. Computational approaches to RNA structure prediction, analysis, and design. *Current Opinion in Structural Biology* **21**, 306–318, doi: http://dx.doi.org/10.1016/j.sbi.2011.03.015 (2011).
46. Laing, C. & Schlick, T. Computational approaches to 3D modeling of RNA. *J. Phys.-Condes. Matter* **22**, 18, doi: 10.1088/0953-8984/22/28/283101 (2010).
47. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51–55, doi: http://www.nature.com/nature/journal/v452/n7183/supinfo/nature06684\_S1.html (2008).
48. Frellsen, J. *et al.* A Probabilistic Model of RNA Conformational Space. *Plos Computational Biology* **5**, 11, doi: 10.1371/journal.pcbi.1000406 (2009).
49. Bida, J. P. & Maher, L. J. Improved prediction of RNA tertiary structure with insights into native state dynamics. *RNA-Publ. RNA Soc.* **18**, 385–393, doi: 10.1261/rna.027201.111 (2012).
50. Zhao, Y. J. *et al.* Automated and fast building of three-dimensional RNA structures. *Sci Rep.* **2**, 6, doi: 10.1038/srep00734 (2012).
51. Popenda, M. *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Research* **40**, 12, doi: 10.1093/nar/gks339 (2012).
52. Cao, S. & Chen, S.-J. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA* **11**, 1884–1897, doi: 10.1261/rna.2109105 (2005).
53. Cao, S. & Chen, S. J. Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA-Publ. RNA Soc.* **15**, 696–706, doi: 10.1261/rna.1429009 (2009).
54. Cao, S. & Chen, S. J. Physics-Based De Novo Prediction of RNA 3D Structures. *J. Phys. Chem. B.* **115**, 4216–4226, doi: 10.1021/jp112059y (2011).
55. Xu, X. J., Zhao, P. N. & Chen, S. J. Vfold: A Web Server for RNA Structure and Folding Thermodynamics Prediction. *PLoS One* **9**, 7, doi: 10.1371/journal.pone.0107504 (2014).
56. Reinhartz, V., Major, F. & Waldispühl, J. Towards 3D structure prediction of large RNA molecules: an integer programming framework to insert local 3D motifs in RNA secondary structure. *Bioinformatics* **28**, i207–i214, doi: 10.1093/bioinformatics/bts226 (2012).
57. Das, R. & Baker, D. Automated de novo prediction of native-like RNA tertiary structures. *Proceedings of the National Academy of Sciences* **104**, 14664–14669, doi: 10.1073/pnas.0703836104 (2007).
58. Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nature Methods* **7**, 291–294, doi: 10.1038/nmeth.1433 (2010).
59. Cheng, C. Y., Chou, F.-C. & Das, R. In *Methods in Enzymology* Vol. 553 (eds Chen Shi-Jie & H. Burke-Aguero Donald) 35–64 (Academic Press, 2015).
60. Leaver-Fay, A. *et al.* In *Methods in Enzymology* Vol. 487 (eds L. Johnson Michael & Brand Ludwig) 545–574 (Academic Press, 2011).
61. Jossinet, F., Ludwig, T. E. & Westhof, E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057–2059, doi: 10.1093/bioinformatics/btq321 (2010).
62. Martinez, H. M., Maizel, J. V. & Shapiro, B. A. RNA2D3D: A program for Generating, Viewing, and Comparing 3-Dimensional Models of RNA. *Journal of Biomolecular Structure and Dynamics* **25**, 669–683, doi: 10.1080/07391102.2008.10531240 (2008).
63. Kim, N., Petingi, L. & Schlick, T. Network Theory Tools for RNA Modeling. *WSEAS transactions on mathematics* **9**, 941–955 (2013).
64. Kim, N. *et al.* Graph-based sampling for approximating global helical topologies of RNA. *Proceedings of the National Academy of Sciences* **111**, 4079–4084, doi: 10.1073/pnas.1318893111 (2014).
65. Kim, N., Zahran, M. & Schlick, T. Computational prediction of riboswitch tertiary structures including pseudoknots by RAGTOP: a hierarchical graph sampling approach. *Methods in enzymology* **553**, 115–135, doi: 10.1016/bs.mie.2014.10.054 (2015).
66. Zahran, M., Sevim Bayrak, C., Elmetwaly, S. & Schlick, T. RAG-3D: a search tool for RNA 3D substructures. *Nucleic Acids Research*, doi: 10.1093/nar/gkv823 (2015).
67. Izzo, J. A., Kim, N., Elmetwaly, S. & Schlick, T. RAG: An update to the RNA-As-Graphs resource. *Bmc Bioinformatics* **12**, 17, doi: 10.1186/1471-2105-12-219 (2011).
68. Fulle, S. & Gohlke, H. Statics of the Ribosomal Exit Tunnel: Implications for Cotranslational Peptide Folding, Elongation Regulation, and Antibiotics Binding. *J. Mol. Biol.* **387**, 502–517, doi: http://dx.doi.org/10.1016/j.jmb.2009.01.037 (2009).
69. Gillespie, J., Mayne, M. & Jiang, M. RNA folding on the 3D triangular lattice. *BMC Bioinformatics* **10**, 1–17, doi: 10.1186/1471-2105-10-369 (2009).
70. Kerpedjiev, P., Höner zu Siederdisen, C. & Hofacker, I. L. Predicting RNA 3D structure using a coarse-grain helix-centered model. *RNA* **21**, 1110–1121, doi: 10.1261/rna.047522.114 (2015).
71. Lamiable, A., Quessette, F., Vial, S., Barth, D. & Denise, A. An Algorithmic Game-Theory Approach for Coarse-Grain Prediction of RNA 3D Structure. *Ieee-Acm Transactions on Computational Biology and Bioinformatics* **10**, 193–199, doi: 10.1109/tcbb.2012.148 (2013).
72. Dawson, W. K., Maciejczyk, M., Jankowska, E. J. & Bujnicki, J. M. Coarse-grained modeling of RNA 3D structure. *Methods*, doi: http://dx.doi.org/10.1016/j.ymeth.2016.04.026.
73. Malhotra, A., Tan, R. K. Z. & Harvey, S. C. Modeling large RNAs and ribonucleoprotein-particles using molecular mechanics techniques. *Biophys. J.* **66**, 1777–1795 (1994).
74. Tan, R. K. Z., Petrov, A. S. & Harvey, S. C. YUP: A molecular simulation program for coarse-grained and multiscaled models. *Journal of Chemical Theory and Computation* **2**, 529–540, doi: 10.1021/ct050323r (2006).
75. Jonikas, M. A., Radmer, R. J. & Altman, R. B. Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models. *Bioinformatics* **25**, 3259–3266, doi: 10.1093/bioinformatics/btp576 (2009).
76. Jonikas, M. A. *et al.* Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* **15**, 189–199, doi: 10.1261/rna.1270809 (2009).
77. Krokhotin, A., Houlihan, K. & Dokholyan, N. V. iFoldRNA v2: folding RNA with constraints. *Bioinformatics*, doi: 10.1093/bioinformatics/btv221 (2015).
78. Sharma, S., Ding, F. & Dokholyan, N. V. iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* **24**, 1951–1952, doi: 10.1093/bioinformatics/btn328 (2008).



79. Denesyuk, N. A. & Thirumalai, D. Coarse-Grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **117**, 4901–4911, doi: 10.1021/jp401087x (2013).
80. Denesyuk, N. A. & Thirumalai, D. How do metal ions direct ribozyme folding? *Nat Chem* **7**, 793–801, doi: 10.1038/nchem.2330, <http://www.nature.com/nchem/journal/v7/n10/abs/nchem.2330.html-supplementary-information> (2015).
81. Mustoe, A. M., Al-Hashimi, H. M. & Brooks, C. L. Coarse Grained Models Reveal Essential Contributions of Topological Constraints to the Conformational Free Energy of RNA Bulges. *The Journal of Physical Chemistry B* **118**, 2615–2627, doi: 10.1021/jp411478x (2014).
82. Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Topological constraints are major determinants of tRNA tertiary structure and dynamics and provide basis for tertiary folding cooperativity. *Nucleic Acids Research* **42**, 11792–11804, doi: 10.1093/nar/gku807 (2014).
83. Mustoe, A. M. *et al.* Noncanonical Secondary Structure Stabilizes Mitochondrial tRNA<sup>Ser</sup>(UCN) by Reducing the Entropic Cost of Tertiary Folding. *J. Am. Chem. Soc.* **137**, 3592–3599, doi: 10.1021/ja5130308 (2015).
84. Cragolini, T., Derreumaux, P. & Pasquali, S. Coarse-Grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* **117**, 8047–8060, doi: 10.1021/jp400786b (2013).
85. Pasquali, S. & Derreumaux, P. HiRE-RNA: A High Resolution Coarse-Grained Energy Model for RNA. *The Journal of Physical Chemistry B* **114**, 11957–11966, doi: 10.1021/jp102497y (2010).
86. Cragolini, T., Laurin, Y., Derreumaux, P. & Pasquali, S. Coarse-Grained HiRE-RNA Model for ab Initio RNA Folding beyond Simple Molecules, Including Noncanonical and Multiple Base Pairings. *Journal of Chemical Theory and Computation* **11**, 3510–3522, doi: 10.1021/acs.jctc.5b00200 (2015).
87. Boniecki, M. J. *et al.* SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Research* **44**, e63, doi: 10.1093/nar/gkv1479 (2016).
88. Magnus, M., Boniecki, M. J., Dawson, W. & Bujnicki, J. M. SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. *Nucleic Acids Research*, doi: 10.1093/nar/gkw279 (2016).
89. Bernauer, J., Huang, X., Sim, A. Y. L. & Levitt, M. Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. *RNA* **17**, 1066–1075, doi: 10.1261/rna.2543711 (2011).
90. Xia, Z., Bell, D. R., Shi, Y. & Ren, P. RNA 3D Structure Prediction by Using a Coarse-Grained Model and Experimental Data. *The Journal of Physical Chemistry B* **117**, 3135–3144, doi: 10.1021/jp400751w (2013).
91. Xia, Z., Gardner, D. P., Gutell, R. R. & Ren, P. Y. Coarse-Grained Model for Simulation of RNA Three-Dimensional Structures. *J. Phys. Chem. B* **114**, 13497–13506, doi: 10.1021/jp104926t (2010).
92. Xia, Z. & Ren, P. In *Biophysics of RNA Folding* Vol. 3 *Biophysics for the Life Sciences* (ed. Rick Russell) Ch. 4, 53–68 (Springer New York, 2013).
93. TINKER Molecular Modeling Package v. 6.3 (<http://dasher.wustl.edu/tinker>).
94. Wang, L.-P., Chen, J. & Van Voorhis, T. Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data. *Journal of Chemical Theory and Computation* **9**, 452–460, doi: 10.1021/ct300826t (2013).
95. Hyeon, C., Dima, R. I. & Thirumalai, D. Size, shape, and flexibility of RNA structures. *The Journal of Chemical Physics* **125**, 194905, doi: <http://dx.doi.org/10.1063/1.2364190> (2006).
96. Saunders, M. G. & Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annual Review of Biophysics* **42**, 73–93, doi: 10.1146/annurev-biophys-083012-130348 (2013).
97. Müller-Plathe, F. Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back. *ChemPhysChem* **3**, 754–769, doi: 10.1002/1439-7641(20020916)3:9<754::AID-CPHC754>3.0.CO;2-U (2002).
98. Tschöp, W., Kremer, K., Batoulis, J., Bürger, T. & Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica* **49**, 61–74, doi: 10.1002/(SICI)1521-4044(199802)49:2/3<61::AID-APOL61>3.0.CO;2-V (1998).
99. Zhao, F. & Xu, J. A Position-Specific Distance-Dependent Statistical Potential for Protein Structure and Functional Study. *Structure* **20**, 1118–1126, doi: <http://dx.doi.org/10.1016/j.str.2012.04.003> (2012).
100. Zhou, H. & Zhou, Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science* **11**, 2714–2726, doi: 10.1110/ps.0217002 (2002).
101. Shen, M.-y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Science: A Publication of the Protein Society* **15**, 2507–2524, doi: 10.1110/ps.062416606 (2006).
102. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **181**, 223–230, doi: 10.1126/science.181.4096.223 (1973).
103. Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* **34**, 564–574, doi: 10.1093/nar/gkj454 (2006).
104. Xia, T. B. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **37**, 14719–14735, doi: 10.1021/bi9809425 (1998).
105. Mathews, D. H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7287–7292, doi: 10.1073/pnas.0401799101 (2004).
106. Freier, S. M. *et al.* Improved Free-Energy Parameters for Predictions of Rna Duplex Stability. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 9373–9377, doi: 10.1073/pnas.83.24.9373 (1986).
107. Borer, P. N., Dengler, B., Tinoco, I. Jr. & Uhlenbeck, O. C. Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.* **86**, 843–853, doi: [http://dx.doi.org/10.1016/0022-2836\(74\)90357-X](http://dx.doi.org/10.1016/0022-2836(74)90357-X) (1974).
108. Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3746–3750, doi: 10.1073/pnas.83.11.3746 (1986).
109. Xia, T. B., McDowell, J. A. & Turner, D. H. Thermodynamics of nonsymmetric tandem mismatches adjacent to G center dot C base pairs in RNA. *Biochemistry* **36**, 12486–12497, doi: 10.1021/bi971069v (1997).
110. Li, P. T. X., Collin, D., Smith, S. B., Bustamante, C. & Tinoco, I. Probing the mechanical folding kinetics of TAR RNA by hopping, force-jump, and force-ramp methods. *Biophys. J.* **90**, 250–260, doi: 10.1529/biophysj.105.068049 (2006).
111. WHAM: The Weighted Histogram Analysis Method v. 2.0.9 (<http://membrane.urmc.rochester.edu/content/wham>).
112. Burkard, M. E., Kierzek, R. & Turner, D. H. Thermodynamics of unpaired terminal nucleotides on short RNA helices correlates with stacking at helix termini in larger RNAs. *J. Mol. Biol.* **290**, 967–982, doi: <http://dx.doi.org/10.1006/jmbi.1999.2906> (1999).
113. Woodside, M. T. *et al.* Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science* **314**, 1001–1004, doi: 10.1126/science.1133601 (2006).
114. Woodside, M. T. *et al.* Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences* **103**, 6190–6195, doi: 10.1073/pnas.0511048103 (2006).
115. Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. & Bustamante, C. Reversible Unfolding of Single RNA Molecules by Mechanical Force. *Science* **292**, 733–737, doi: 10.1126/science.1058498 (2001).
116. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021, doi: 10.1002/jcc.540130812 (1992).
117. Eastman, P. *et al.* OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation* **9**, 461–469, doi: 10.1021/ct300857j (2013).
118. Dale, T., Smith, R. & Serra, M. J. A test of the model to predict unusually stable RNA hairpin loop stability. *RNA* **6**, 608–615 (2000).



119. Giese, M. R. *et al.* Stability of RNA Hairpins Closed by Wobble Base Pairs. *Biochemistry* **37**, 1094–1100, doi: 10.1021/bi972050v (1998).
120. Antao, V. P. & Tinoco, I. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Research* **20**, 819–824, doi: 10.1093/nar/20.4.819 (1992).
121. Serra, M. J., Lyttle, M. H., Axenson, T. J., Schadt, C. A. & Turner, D. H. RNA hairpin loop stability depends on closing base pair. *Nucleic Acids Research* **21**, 3845–3849 (1993).
122. Groebe, D. R. & Uhlenbeck, O. C. Characterization of Rna Hairpin Loop Stability. *Nucleic Acids Research* **16**, 11725–11735, doi: 10.1093/nar/16.24.11725 (1988).

## Acknowledgements

We are grateful for support from the Robert A. Welch Foundation (Grant F-1691 to P.R.) and the National Institutes of Health (Grants GM106137 and GM114237 to P.R.).

## Author Contributions

P.R. designed and supervised the research. D.R.B., S.Y.C., and P.R. prepared the manuscript. D.R.B., H.S., and P.R. parameterized the model. D.R.B., and S.Y.C. conducted the pulling and annealing simulations. D.R.B., S.Y.C., and P.R. analyzed the simulations.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Bell, D. R. *et al.* Capturing RNA Folding Free Energy with Coarse-Grained Molecular Dynamics Simulations. *Sci. Rep.* **7**, 45812; doi: 10.1038/srep45812 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017