



# RNA force field with accuracy comparable to state-of-the-art protein force fields

Dazhi Tan<sup>a,1</sup>, Stefano Piana<sup>a,2</sup>, Robert M. Dirks<sup>a,3</sup>, and David E. Shaw<sup>a,b,2</sup>

<sup>a</sup>D. E. Shaw Research, New York, NY 10036; and <sup>b</sup>Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032

Edited by Jennifer A. Doudna, University of California, Berkeley, CA, and approved December 21, 2017 (received for review July 26, 2017)

Molecular dynamics (MD) simulation has become a powerful tool for characterizing at an atomic level of detail the conformational changes undergone by proteins. The application of such simulations to RNA structures, however, has proven more challenging, due in large part to the fact that the physical models (“force fields”) available for MD simulations of RNA molecules are substantially less accurate in many respects than those currently available for proteins. Here, we introduce an extensive revision of a widely used RNA force field in which the parameters have been modified, based on quantum mechanical calculations and existing experimental information, to more accurately reflect the fundamental forces that stabilize RNA structures. We evaluate these revised parameters through long-timescale MD simulations of a set of RNA molecules that covers a wide range of structural complexity, including single-stranded RNAs, RNA duplexes, RNA hairpins, and riboswitches. The structural and thermodynamic properties measured in these simulations exhibited dramatically improved agreement with experimentally determined values. Based on the comparisons we performed, this RNA force field appears to achieve a level of accuracy comparable to that of state-of-the-art protein force fields, thus significantly advancing the utility of MD simulation as a tool for elucidating the structural dynamics and function of RNA molecules and RNA-containing biological assemblies.

nucleic acid | molecular dynamics simulations | AMBER | Anton | force field

Long-timescale molecular dynamics (MD) simulations have been successfully used to characterize complex conformational transitions in proteins (1). In principle, MD simulation should also be a powerful tool for characterizing such conformational changes in RNA molecules, whose highly flexible 3D structures enable their diverse cellular functions (2–4). The physical models (“force fields”) used in RNA simulations, however, are widely considered to be much less accurate overall than the force fields used in protein simulations; artifacts have often been observed in simulations of even simple RNA systems (5–10). Improvements to the accuracy of RNA force fields would greatly facilitate the use of MD simulation to study the functional dynamics of biologically significant RNA molecules.

AMBER is currently the most widely used force field for MD simulations of RNA systems (11, 12). Over the last decade, efforts to improve the accuracy of the AMBER RNA force field have focused largely on refining backbone and glycosidic torsion parameters on the basis of quantum mechanical (QM) calculations (11, 13–17). These optimizations resulted in substantial improvements in the description of canonical double-helical RNA structures but did not fully resolve other key problems such as the overestimation of nucleobase stacking and the underestimation of base-pairing strength, which are determined primarily by electrostatics and van der Waals (vdW) interactions (18–20). Chen and García (20) addressed this problem by adjusting the vdW parameters of nucleobase atoms to weaken stacking and strengthen base pairing. Their choice of parameters improved the simulated structural properties of RNA tetraloops, which are widely used for benchmarking the quality of RNA force fields, but did not improve the accuracy of simulations of some other model RNA systems such as ssRNAs (10, 20).

Here, using a combination of ab initio and empirical methods, we have modified electrostatic, vdW, and torsional parameters of the AMBER ff14 RNA force field (12) to more accurately reproduce the energetics of nucleobase stacking, base pairing, and key torsional conformers. We then performed MD simulations of both extended and structured RNA systems, including short and long ssRNAs, RNA duplexes, tetraloops, and riboswitches, using the revised parameters and the recently published TIP4P-D water model (21). The simulated structural and thermodynamic properties of these systems exhibited substantially improved agreement with experimental data, to the point that the accuracy of this revised AMBER RNA force field can be considered comparable to that of state-of-the-art fixed-charge protein force fields. We expect that this force field will enable the use of MD simulation to investigate the complex functional dynamics of a wide range of RNA molecules, as well as RNA-containing macromolecular machines.

## Results

**Reparameterization of the AMBER RNA Force Field.** We adopted a two-step approach to reparameterize the AMBER ff14 RNA force field (12). We first developed revised nucleobase nonbonded parameters to improve the description of base pairing and stacking, then derived revised terms for the  $\gamma$ ,  $\zeta$ , and  $\chi$  torsions.

To tackle base pairing, we calculated the QM dimer interaction energy at the MP2 level as a function of intermolecular separation for six common types of base pairs (see *Methods* for details). This level of QM calculation has been shown to be

## Significance

The complex and often highly dynamic 3D structures of RNA molecules are central to their diverse cellular functions. Molecular dynamics (MD) simulations have played a major role in characterizing the structure and dynamics of proteins, but the physical models (“force fields”) used for simulating nucleic acids are substantially less accurate overall than those used in protein simulations, creating a major challenge for MD studies of RNA. Here, we report an RNA force field capable of describing the structural and thermodynamic properties of RNA molecules with accuracy comparable to state-of-the-art protein force fields. This force field should facilitate the use of MD simulation as a tool for the study of biologically significant RNA molecules and protein–RNA complexes.

Author contributions: D.T., S.P., R.M.D., and D.E.S. designed research; D.T. and R.M.D. performed research; D.T. analyzed data; and D.T., S.P., and D.E.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

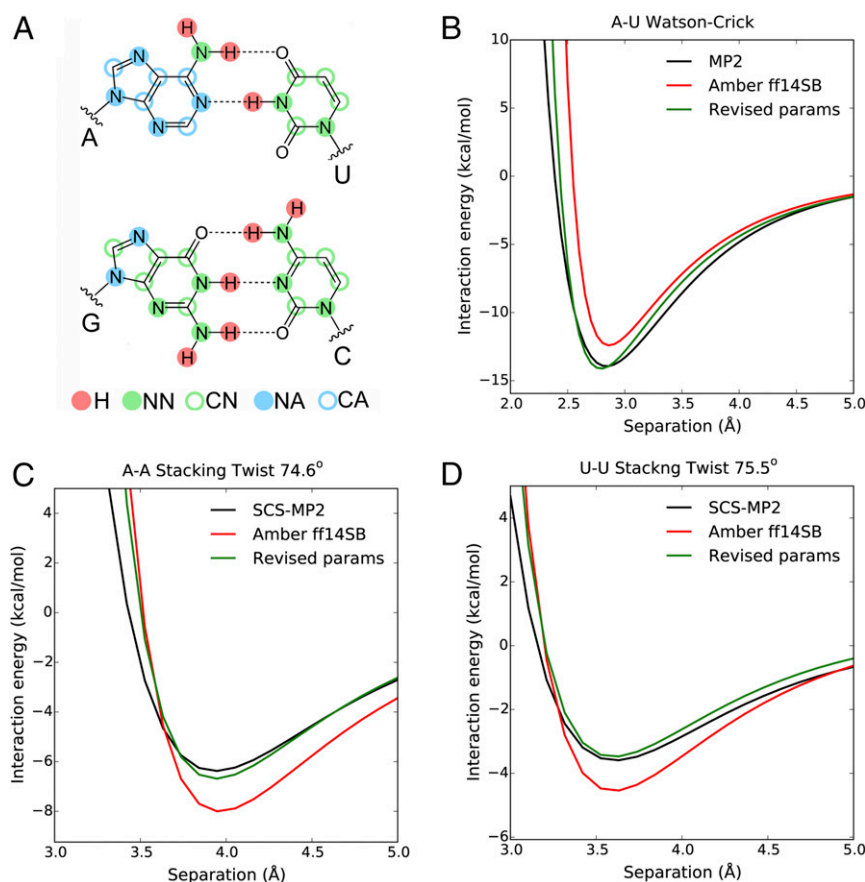
This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>Present address: Silicon Therapeutics, Boston, MA 02210.

<sup>2</sup>To whom correspondence may be addressed. Email: Stefano.Piana-Agostinetti@DEShawResearch.com or David.Shaw@DEShawResearch.com.

<sup>3</sup>Deceased February 3, 2015.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713027115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713027115/-DCSupplemental).



**Fig. 1.** (A) The vdW atom types for nucleobase atoms in the revised force field. (B) Interaction energies of the A–U Watson–Crick base pair as functions of intermolecular separations. (C) Interaction energies of the A–A stacked dimer at 74.6° twist angle as functions of intermolecular separations. (D) Interaction energies of the U–U stacked dimer at 75.5° twist angle as functions of intermolecular separations. See *SI Appendix, Figs. S1–S3* for the complete set of data included in the parameterization.

adequate for describing base-pairing energies (22). Comparison of the QM and molecular mechanical (MM) interaction energy profiles shows that the AMBER ff14 parameters underestimate base-pairing strength, especially at distances smaller than 2.7 Å [where the repulsive  $C_{12}$  component of the Lennard-Jones (LJ) potential becomes excessively steep] (Fig. 1*B* and *SI Appendix, Fig. S1*). We adjusted charges and vdW parameters of nucleobase nitrogen and polar hydrogen atoms to minimize the discrepancy between the QM and MM potential energy profiles (Fig. 1*B*, Tables 1 and 2, and *SI Appendix, Fig. S1*). Notably, splitting the nucleobase nitrogen atoms into two different vdW atom types on the basis of their aromatic character (Fig. 1*A*) and reducing the  $\sigma$  and  $\epsilon$  values of polar hydrogens to zero (Table 1) greatly improved the quality of the fit to the QM data.

The AMBER ff14 parameters substantially overestimate the stacking energy with respect to the QM values (18, 20) (Fig. 1*C* and *D*). To improve the description of stacking, we calculated the QM dimer interaction energies of A–A and U–U stacked dimers with varying relative orientations and separations using SCS (MI)-MP2, a semiempirical MP2-based method that allows the computation of accurate stacking energies (19, 23). We then adjusted the vdW parameters of the nucleobase carbon atoms to reproduce the QM interaction energies (Fig. 1*C* and *D* and *SI Appendix, Figs. S2* and *S3*).

Our modifications of the charges and vdW parameters for the nucleobases necessitated an adjustment to the glycosidic ( $\chi$ ) dihedral terms. Following the methods of previous parameterizations of the  $\chi$  torsion (15, 17) (*Methods* and *SI Appendix, Table S1*),

we generated the in vacuo QM potential energy profiles for the four nucleosides with either the C3'-endo or C2'-endo sugar pucker configuration (*SI Appendix, Fig. S4*). The  $\chi$  dihedral terms were then parameterized to minimize the differences between the

**Table 1.** Charges of the nucleobase atoms in the revised force field

| Base | Atom    | Charges    |         |
|------|---------|------------|---------|
|      |         | AMBER ff14 | Revised |
| C    | N4      | −0.9530    | −0.8716 |
|      | H41/H42 | +0.4234    | +0.3827 |
| U    | N3      | −0.3549    | −0.3913 |
|      | H3      | +0.3154    | +0.3518 |
| G    | N1      | −0.4787    | −0.5606 |
|      | H1      | +0.3424    | +0.4243 |
|      | N2      | −0.9672    | −1.0158 |
|      | H21/H22 | +0.4364    | +0.4607 |
| A    | N1      | −0.7615    | −0.7969 |
|      | H2      | +0.0473    | +0.0650 |
|      | N6      | −0.9019    | −1.0088 |
|      | H61/H62 | +0.4115    | +0.4738 |

Unlisted atoms have unchanged charges.

**Table 2. The vdW atom types and parameters for nucleobase atoms in the revised force field**

| vdW atom type | $\sigma$ , Å |         | $\epsilon$ , kcal·mol <sup>-1</sup> |         |
|---------------|--------------|---------|-------------------------------------|---------|
|               | AMBER ff14   | Revised | AMBER ff14                          | Revised |
| H             | 1.0691       | 0.0000  | 0.0157                              | 0.0000  |
| NA            | 3.2500       | 3.2890  | 0.1700                              | 0.1700  |
| NN            |              | 3.3507  |                                     |         |
| CA            | 3.3997       | 3.3004  | 0.0860                              | 0.0636  |
| CN            |              | 3.2850  |                                     | 0.0538  |

Nonpolar hydrogen and oxygen atoms (not listed) have unchanged parameters.

QM and MM potential energy profiles (*SI Appendix, Table S2*). Explicit-solvent MD simulations of mononucleosides showed that the pyrimidine nucleosides adopt the anti  $\chi$  configuration to a greater extent than do purine nucleosides (*SI Appendix, Table S3*), qualitatively matching the results of earlier NMR experiments (24, 25).

The revised force field includes the nonbonded phosphate parameters published by Steinbrecher et al. (26), which have increased  $\sigma$  values for oxygen atoms. This change required the introduction of new backbone  $\gamma$  and  $\zeta$  dihedral terms, which were previously shown to be critical for maintaining the stability of both DNA and RNA duplexes in simulation (14, 16, 27). Initial attempts at parameterizing the  $\gamma$  and  $\zeta$  torsions to match the in vacuo QM potential energy profiles led to erroneous populations of the  $\gamma$  and  $\zeta$  conformers in explicit-solvent simulations. We thus employed a hybrid approach that combined ab initio calculations with the COSMO implicit-solvent model (28) and explicit-solvent MD simulations (see *Methods* for details). The theoretical populations of the *+gauche*, *-gauche*, and *trans* conformers were first derived from the QM-COSMO potential energy profiles, and the dihedral terms were then adjusted to ensure that similar populations were reproduced in the simulations (*SI Appendix, Figs. S5 and S6 and Tables S2, S3, and S4*).

**The Revised Parameters Recover the Solution Conformations of ssRNA Tetranucleotides.** The protein force fields currently in use for MD simulations are typically able to faithfully reproduce the solution conformations of short peptides and attain close agreement ( $\chi^2 < 2$ ) between the simulated and experimental NMR measurements (29–31). In contrast, MD simulations of comparably simple ssRNA tetranucleotides using currently available RNA force fields often result in structures that deviate significantly from the experimentally determined structures. These molecules tend to favor an intercalated conformation in simulation (5–10, 32) (Fig. 2A), but the large discrepancies between the simulated and experimental <sup>3</sup>J scalar-coupling values (*SI Appendix, Table S5*) suggest that this intercalated conformation is probably a simulation artifact.

We evaluated the accuracy of our revised parameters for the AMBER RNA force field through long MD simulations of the AAAA, CAAU, CCCC, GACC, and UUUU ssRNA tetranucleotides, for which extensive NMR data are available. All except UUUU primarily adopted the experimentally observed A-form-like conformations, and intercalation was either completely eliminated or dramatically reduced (Fig. 2B). Consequently, the  $\chi^2$  errors between the simulated and experimental <sup>3</sup>J scalar-coupling values were substantially improved to 1.4–2.2 (*SI Appendix, Table S5*), on par with the performance of most protein force fields. A substantially reduced amount of intercalation was also observed in a recent work (33) using a combination of the AMBER force field and the OPC water model (34), which is similar to the TIP4P-D water model used here, and it was suggested

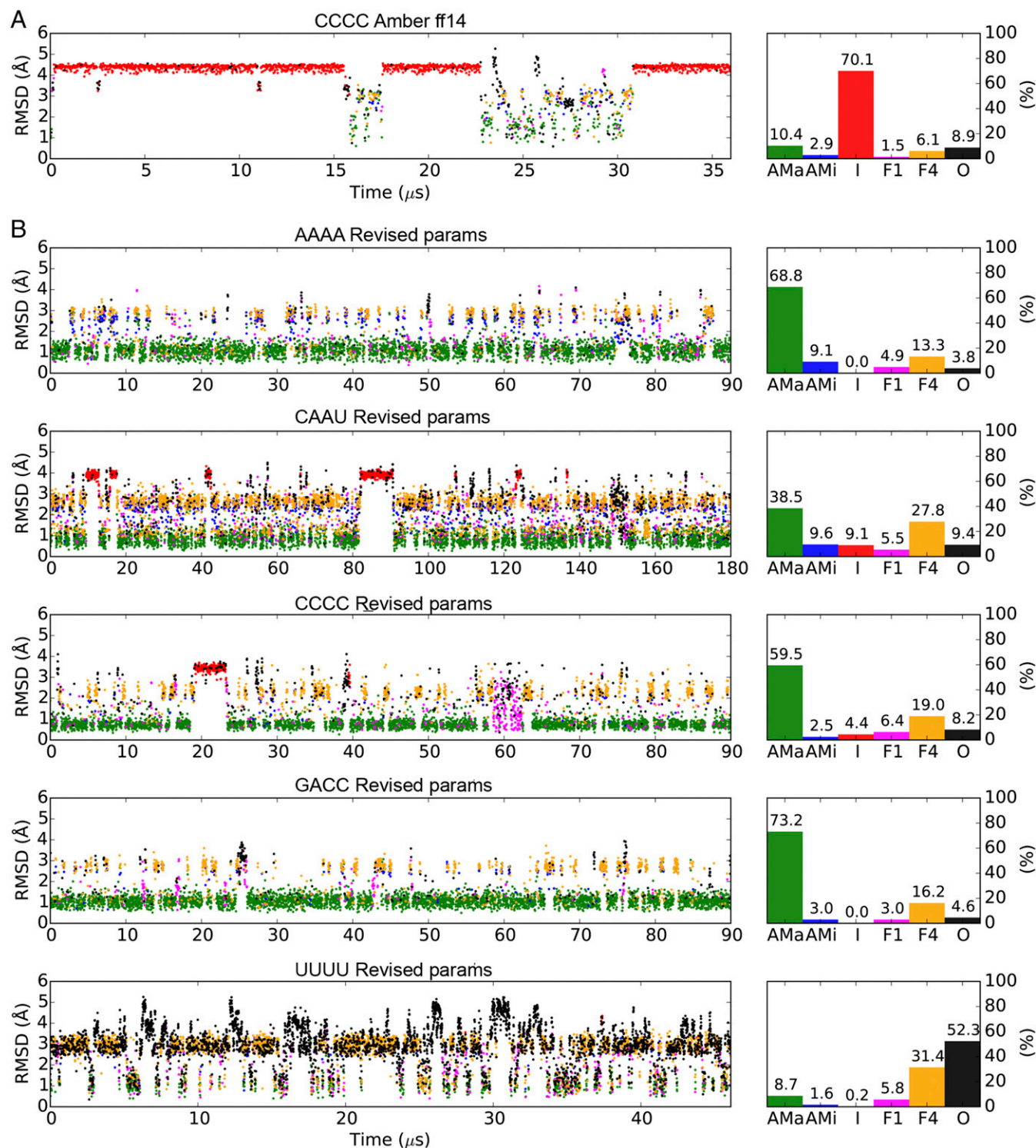
that a substantial fraction of the improvement should be ascribed to the water model. As a further test of this conclusion we performed a simulation with the OPC water model and found only ~20% intercalation (*SI Appendix, Fig. S9*) and much improved agreement with experiment (*SI Appendix, Table S5*) compared with TIP3P. This result is in line with previous findings (33) and highlights the importance of the choice of water model for describing ssRNA.

Unlike the other tetranucleotides tested, UUUU predominantly adopts a minimally stacked random-coil conformation in solution (6). In the simulation of UUUU using the revised force field this random-coil conformation was reproduced in the majority of the frames (Fig. 2). The deviation between the simulated and experimental NMR observables for UUUU is larger ( $\chi^2 \approx 4$ ) than those of the other tetranucleotides (*SI Appendix, Table S5*), possibly because UUUU visited the stacked states in a nonnegligible number of simulation frames (Fig. 2), resulting in a less disordered conformational ensemble than was observed experimentally.

**The Revised Parameters Improve the Description of Long Disordered ssRNAs.** MD simulations of disordered proteins with several commonly used protein force fields together with the TIP3P water model result in more compact structures (characterized by much smaller radius of gyration values) than are observed experimentally (21). The TIP4P-D water model, however, largely corrects the overly compact simulated conformations of disordered proteins (21, 35).

Similarly, the AMBER ff14 RNA force field and the TIP3P water model cause the highly extended rU<sub>40</sub> ssRNA to quickly collapse to an ensemble of globule-like structures with end-to-end distance values much smaller than those measured by single-molecule FRET experiments ( $\langle R_{FRET} \rangle$ ) (Fig. 3A and C) (36). The combination of AMBER ff14 and the TIP4P-D water model is also unable to prevent rU<sub>40</sub> from collapsing (*SI Appendix, Fig. S19*). Simulations using our revised RNA parameters and the TIP4P-D water model, however, maintain the extended conformation of rU<sub>40</sub> and recapitulate the experimental  $\langle R_{FRET} \rangle$  values at low salt concentration ([NaCl] < 0.1 M) (Fig. 3B and C). In simulations under higher salt concentrations ([NaCl] ≥ 0.1 M), rU<sub>40</sub> alternates between extended and double-helical hairpin conformations containing up to 18 U–U base pairs, leading to lower  $\langle R_{FRET} \rangle$  values (Fig. 3 and *SI Appendix, Fig. S8*). Experimental evidence supports the existence of such hairpin structures of polyuridylic acids, but only at temperatures of 280–290 K, which are lower than that of the simulations (300 K) (37, 38). The revised force field thus recapitulates well the properties of disordered ssRNAs but likely overstabilizes the double-helical conformation.

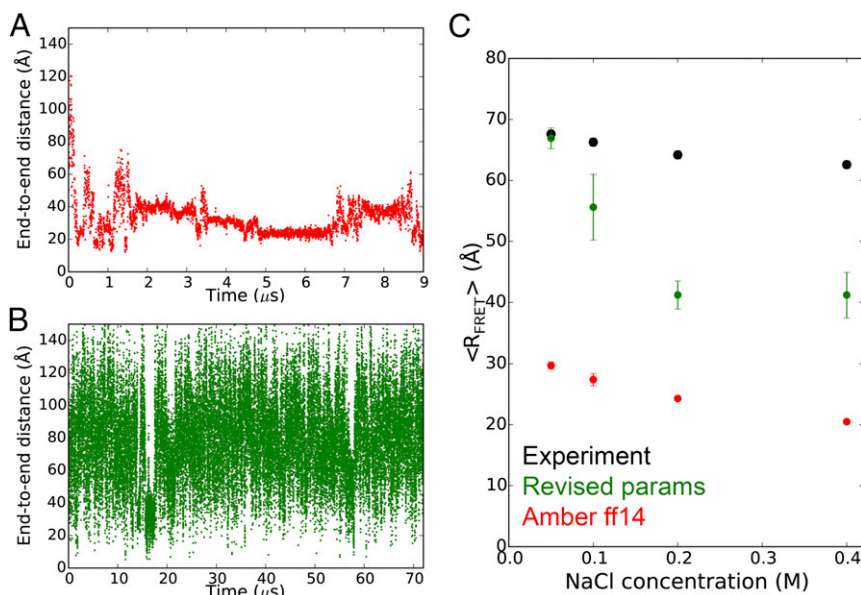
**The Revised Parameters Can Reversibly Fold RNA Duplexes and Tetraloops to Subangstrom Resolution.** A major success of modern protein force fields is that they can fold many small protein domains to within 2 Å rmsd of the crystal structures, and can typically predict their melting temperatures ( $T_m$ ) to within 50 K of the experimentally determined values (39–41). The reversible folding and unfolding of basic RNA structural elements such as duplexes and tetraloops in MD simulations, however, has proven more challenging due to inaccuracies in the RNA force fields. Although replica-exchange MD simulations using either the AMBER ff14 or Chen–García parameters have been shown to fold UUCG and GNRA tetraloops, the native conformations—which are observed in both X-ray and NMR structures, and thus presumably dominate in solution—only accounted for a small fraction (Chen–García: ~10%, AMBER ff14: <1%) of the lowest-temperature (277 K) ensemble (10, 20, 42). Using a simulated tempering protocol (43, 44) and our revised parameter set we were able to simulate the reversible folding of eight RNA duplexes of varying sequences and lengths, as well as the ggca-cUUCGgugcc, gccGAAAggc, and ggcGCAAgcc hyperstable tetraloops (Figs. 4 and 5 and *SI Appendix, Figs. S11 and S12*).



**Fig. 2.** MD simulations of ssRNA tetramers using AMBER ff14 and the TIP3P water model (A) or the revised parameters (B). The scatter plots on the left indicate rmsds of the backbone (including the ribose) nonhydrogen atoms to the initial A-form structures. Based on the stacking patterns of the four nucleotides, the simulated conformations are categorized into six groups: A-form major (AMa, green), A-form minor (AMi, blue), Intercalated (I, red) Nucleotide-4 flipped (F4, yellow), Nucleotide-1 flipped (F1, magenta), and Others (O, black). See *Methods* for the detailed definitions of these conformations. Percentage populations of the six conformations in each simulation are shown on the right.

In the simulations of the duplexes at least eight reversible folding events were observed in each trajectory, allowing the estimation of thermodynamic quantities that can be compared with the corresponding values predicted by the relatively accurate,

experimentally derived nearest-neighbor (NN) model (45). The folding free energies of the duplexes at 310 K ( $\Delta G_{310\text{K}}$ ) calculated from the simulations were slightly more negative than the NN model (root-mean-square error = 1.4 kcal mol<sup>-1</sup>, mean signed



**Fig. 3.** The end-to-end distance of the rU<sub>40</sub> long ssRNA simulated with AMBER ff14 and TIP3P (A) or the revised parameters and TIP4P-D (B) under 0.05 M NaCl at 300 K. See *SI Appendix, Figs. S7 and S8* for other NaCl concentrations. (C) The  $\langle R_{FRET} \rangle$  values of rU<sub>40</sub> under different NaCl concentrations. The experimental values were reported in ref. 36. See *Methods* for the procedure of calculating the simulated  $\langle R_{FRET} \rangle$  values.

error =  $-1.1 \text{ kcal mol}^{-1}$ ; Fig. 4E), indicating that the revised parameters moderately overstabilize double-helical conformations, consistent with the results of the rU<sub>40</sub> simulations. The simulated  $T_m$  values were 5–50 K higher (mean signed error =  $+27.6 \text{ K}$ ) than the values predicted by the NN model. More stable duplexes tended to have larger discrepancies in  $T_m$  (Fig. 4D). This observation is related to an underestimation of the enthalpy of duplex formation in simulation, which leads to a less-steep melting curve, thus shifting the midpoint toward a higher temperature range. Underestimation of folding enthalpy has also been observed in protein-folding simulations (46). Simulations of the tetraloops each registered several reversible folding events to  $<1 \text{ \AA}$  rmsd of the experimental structures from completely unfolded conformations (Fig. 5). In particular, the single-stranded loop regions of the tetraloops repeatedly adopted conformations that were  $<0.8 \text{ \AA}$  rmsd from the experimental structures (*SI Appendix, Fig. S11*), suggesting that their native conformations have been successfully recapitulated in the simulations. The fraction of the native conformation of each tetraloop at the lowest-temperature (280 K) ensemble was at least 40% (Fig. 5 and *SI Appendix, Table S6*), a significant improvement compared with the values reported previously (10, 20, 42). As we observed for the RNA duplexes, the  $T_m$  values of the double-stranded segments of the tetraloops (Fig. 5) were 20–30 K higher than the experimental values (*SI Appendix, Table S6*) (47).

#### Reversible Ligand Binding to a Guanine Riboswitch Aptamer Domain.

The level of accuracy achieved in describing the conformational ensemble of relatively simple RNA systems encouraged us to assess whether the revised parameters can maintain the stability of larger, more complex RNA molecules such as the aptamer domains of riboswitches, which can fold into 3D structures having a degree of complexity comparable to that of protein domains (3).

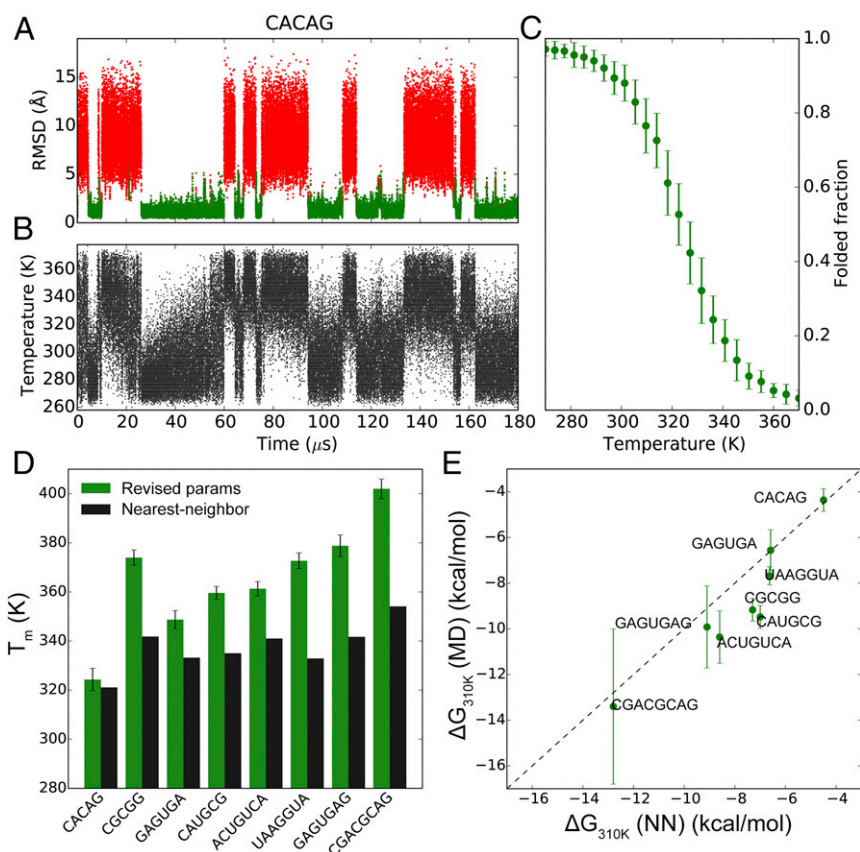
The aptamer domain of the *Bacillus subtilis xpt* guanine riboswitch binds guanine and its analog hypoxanthine with high affinity (48, 49). Crystal structures show that the aptamer adopts a tuning-fork-like fold containing three double-helical regions (P1–P3) (*SI Appendix, Fig. S13A*) (50, 51). P2 and P3 are parallel to each other and are anchored together through the kissing-loop interactions between their hairpin loops L2 and L3. The single-stranded segments connecting the stems (J1/2, J2/3, and J3/1) form

a three-way junction that acts as the binding pocket, where the ligand is specifically recognized by residues U22, U51, and C74 through multiple hydrogen bonds (*SI Appendix, Fig. S13B*) (50, 51).

In the simulation of the guanine-bound aptamer, the guanine-binding pocket accurately maintained its native conformation for 81  $\mu\text{s}$ , at which point guanine escaped from the binding pocket into the solvent (Fig. 6A). The ligand was briefly recaptured by the pocket at 97  $\mu\text{s}$  but adopted a binding pose distinct from the crystal structure (Fig. 6A and *SI Appendix, Fig. S13C*) and stayed in this nonnative bound conformation for 13  $\mu\text{s}$  before dissociating from the pocket. At 114  $\mu\text{s}$  guanine again entered the binding pocket, and this time the native bound conformation was readopted and maintained until the end of the simulation (Fig. 6A). This result shows that the simulation is able to identify the correct binding pocket and binding pose in this complex system. Because the experimental affinity of guanine to the aptamer is less than 10 nM (49), the fact that we observed reversible ligand dissociation and reassociation within a 180- $\mu\text{s}$  simulation starting from the bound structure indicates that the affinity of the ligand is probably underestimated in simulation, which may result from a suboptimal choice of parameters for the ligand (*Methods*). In a 90- $\mu\text{s}$  simulation performed with the AMBER ff14SB force field the overall riboswitch structure was maintained, but the binding-pocket structure was not conserved: After 20  $\mu\text{s}$  U51 rotated away from guanine and did not participate in the binding (*SI Appendix, Fig. S16*).

Examining the conformational flexibility of the aptamer domain in its guanine-bound and apo states in simulation, we found the conformations of the three double-helical regions to be stable, whereas the single-stranded segments exhibited considerable conformational freedom (Fig. 6B). J2/3 was the most dynamic part of the molecule, and the binding of guanine considerably reduced its flexibility (Fig. 6B). The simulated global dynamics of the apo and bound aptamer domain are qualitatively consistent with the results of previous *N*-methylisatoic anhydride probing assays, with the exception that the J1/2 segment was less flexible in the simulations compared with the experimental results (52).

**Magnesium-Dependent Dynamics of the Apo SAM-I Riboswitch Aptamer Domain.** As a model for a complex system that undergoes large-scale conformational changes, we focused on the aptamer domain of



**Fig. 4.** MD simulations of RNA duplexes. (A) Nonhydrogen rmsd trace of the CACAG RNA duplex. The folded (green) and unfolded (red) states of the duplex are indicated. (B) The time evolution of temperatures in the simulated tempering MD simulation of the CACAG duplex. (C) Melting curve of the CACAG duplex. The rmsd traces and melting curves of all eight RNA duplexes are in *SI Appendix, Fig. S12*. (D)  $T_m$  values of the eight RNA duplexes calculated from the MD trajectories. In black are the  $T_m$  values estimated with the NN model (45). (E)  $\Delta G_{310K}$  values calculated from the MD trajectories (by fitting to the nonlinear van't Hoff equation) compared with those estimated with the NN model. The lowest temperatures in the simulated tempering simulations of GAGUGAG and CGACGCAG were higher than 310 K, so for these duplexes the  $\Delta G_{310K}$  values were extrapolated from the fitted  $\ln K_{eq} - 1/T$  curves.

the *Thermoanaerobacter tengcongensis* SAM-I riboswitch. Stoddard et al. (53) used small-angle X-ray scattering (SAXS) experiments to demonstrate that the solution conformation of the SAM-I riboswitch is heavily influenced by magnesium. In the absence of magnesium the aptamer domain adopts extended and disordered structures regardless of the availability of the ligand. The addition of magnesium alone, in the absence of the ligand, is able to facilitate the formation of tertiary interactions within the apo aptamer, resulting in an ensemble of compact, yet heterogeneous, conformations. Ligand binding then further immobilizes the secondary structural elements and leads to a stable and well-defined structure (53).

We removed the ligand from the crystal structure of the ligand-bound aptamer domain (54) to generate the starting structure for our MD simulations. The theoretical SAXS profile of the starting structure suggests that it is representative of the solution conformation in the Mg(+), SAM-bound state ( $\chi = 1.86$ ) while being distinct from that of the Mg(+), apo ensemble ( $\chi = 2.82$ ) (*SI Appendix, Table S7*). After 30  $\mu$ s of simulation with magnesium we observed an ensemble of compact conformations, Mg-C (Fig. 6C), whose theoretical SAXS profile instead exhibits better agreement with the Mg(+), apo profile ( $\chi = 1.74$ ) than with the Mg(+), bound profile ( $\chi = 2.27$ ) (Fig. 6E and F and *SI Appendix, Table S7*). The simulated conformational transition of the SAM-I riboswitch from the bound to the apo state upon the removal of the ligand in the presence of magnesium is thus consistent with that observed experimentally.

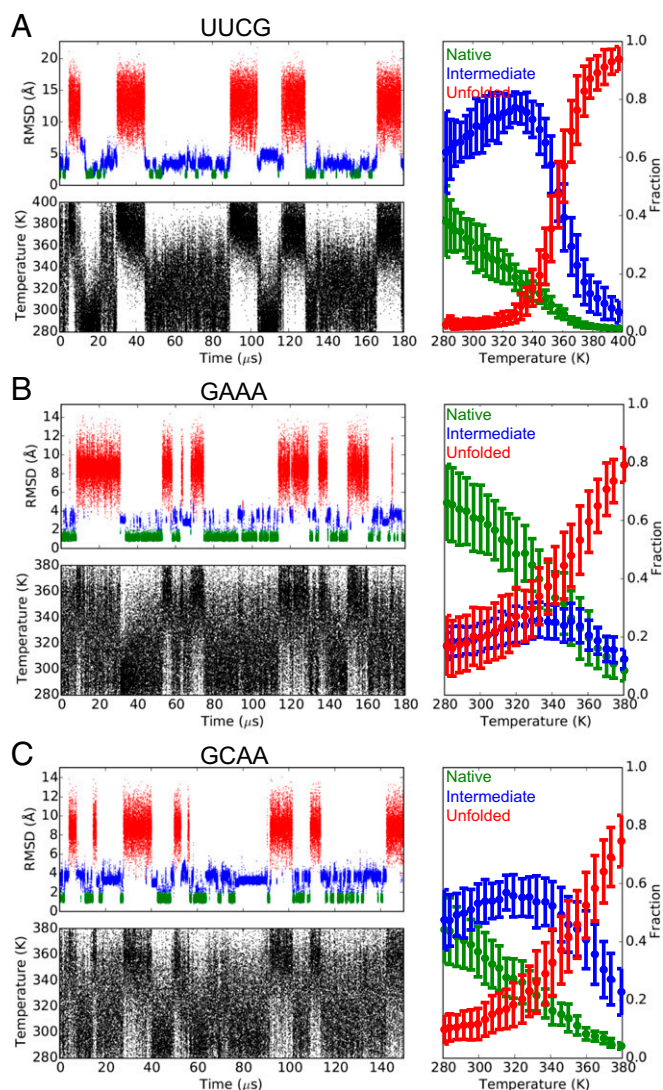
In a simulation with potassium as the sole counter ion the aptamer domain first sampled compact structures (K-C) not

observed in the Mg-C ensemble (see *SI Appendix* for a more detailed description of the conformational changes in this riboswitch) and later adopted more extended conformations (K-E) featuring disrupted tertiary interactions (Fig. 6D and *SI Appendix*). The SAXS profile of the K-E ensemble displays high intensities in the larger scattering angles, which is substantially different from both Mg(+) experimental profiles (Fig. 6E and F and *SI Appendix, Table S7*) but shares similarities with those measured in the absence of magnesium (53). Since the complete unfolding of the aptamer domain might require milliseconds of simulation time (2), these results suggest that the K-E ensemble may represent an intermediate on the unfolding pathway.

In contrast to what would be expected based on the experimental results, we found that simulations of the SAM-I riboswitch performed with the AMBER ff14SB force field were not strongly influenced by Mg(+): The overall structure of the riboswitch was well maintained in 90- $\mu$ s simulations performed either with or without Mg(+) (*SI Appendix, Fig. S17*).

## Discussion

We have reparameterized the AMBER RNA force field to rebalance the stacking and base-pairing interactions, which are the primary stabilizing forces of RNA structures. In particular, we have modified the charges and vdW parameters of the nucleobase atoms, as well as the  $\chi$ ,  $\gamma$ , and  $\zeta$  torsion potentials. Moreover, we have used the TIP4P-D water model (21), in which there is a better balance between the strengths of electrostatic and dispersive interactions, and where the dispersive interactions are not too weak relative to



**Fig. 5.** The nonhydrogen rmsd traces, the time evolutions of temperatures, and the melting curves of the *ggcacUUCGgugcc* (A), *gccGAAAagc* (B), and *ggcGCAAagc* (C) RNA tetraloops. The native (green), intermediate (blue), and unfolded (red) states of the tetraloops are indicated. Error bars were estimated using a blocking analysis.

those in biomolecules. Similarly to its effect in simulations of disordered proteins, this water model helps to alleviate the problem of overly stable stacking interactions (33).

This revised set of parameters was tested for its ability to reproduce the structural and thermodynamic properties of RNA systems of varying structural complexity. The simulations were able to closely recapitulate the conformational ensembles of long and short single-stranded polynucleotides, the structures (to sub-angstrom rmsd) and stabilities (to within  $1.5 \text{ kcal}\cdot\text{mol}^{-1}$ ) of tetraloops and RNA duplexes, the conformational dynamics of complex RNA folds, and the binding pose of a small ligand. The level of accuracy we observed is comparable to that of state-of-the-art protein force fields, which can typically reproduce the conformational ensembles of small peptides and disordered proteins (21, 29, 30, 46), predict the structures of small fast-folding proteins to less than 2-Å resolution (39–41), accurately describe the dynamics of folded proteins (55, 56), and, in favorable cases, predict the binding poses of ligands (57). This substantial improvement in accuracy should allow MD simulation to play an expanded role as a tool for elucidating the structure and dynamics of biologically important RNA molecules and when used in combi-

nation with protein force fields may also facilitate the study of protein-RNA complexes.

We expect that the revised nucleobase nonbonded parameters will be largely transferable to DNA, but the absence of the 2'-hydroxyl group in deoxynucleotides has nonnegligible effects on the torsion profiles. It will thus be necessary to introduce DNA-specific dihedral terms to make this parameter set suitable for DNA simulations. [A similar distinction between DNA and RNA force fields already exists in the most recent AMBER16 package, which uses the  $\chi$ OL3 correction for RNA (15) and the bsc1 update for DNA (16).]

Like any MM force field, the parameterization described here has limitations. One issue is that it appears to modestly over-stabilize the helical regions of RNAs, resulting in higher melting temperatures for the rU<sub>40</sub> long hairpin, the RNA duplexes, and the tetraloops. This over-stabilization may also contribute to the excessive ordering of the UUUU tetranucleotide and the J1/2 segment of the guanine riboswitch aptamer domain. An area that we did not investigate is the accuracy of nonbonded parameters for charged species like ions and the phosphate group. The phosphate vdW parameters published by Steinbrecher et al. (26) were originally developed to ensure that the solvation free energies of bi-organic phosphates in TIP3P water matched the experimental values, and most of the current ion parameters were developed to be compatible with the standard TIP3P or TIP4P water models (58–60). In particular, it has been observed that the CHARMM22 parameters for magnesium, which were used in simulating the riboswitches, result in an exchange rate of water from the first solvation shell that is orders of magnitude smaller than the experimental estimate (61). Although it is unlikely that this limitation would significantly affect the results of the simulations in this work that included magnesium (the ions were expected to remain hydrated and not interact directly with the RNA), it would be desirable to derive new ion and phosphate parameters for better compatibility with the TIP4P-D water model.

## Methods

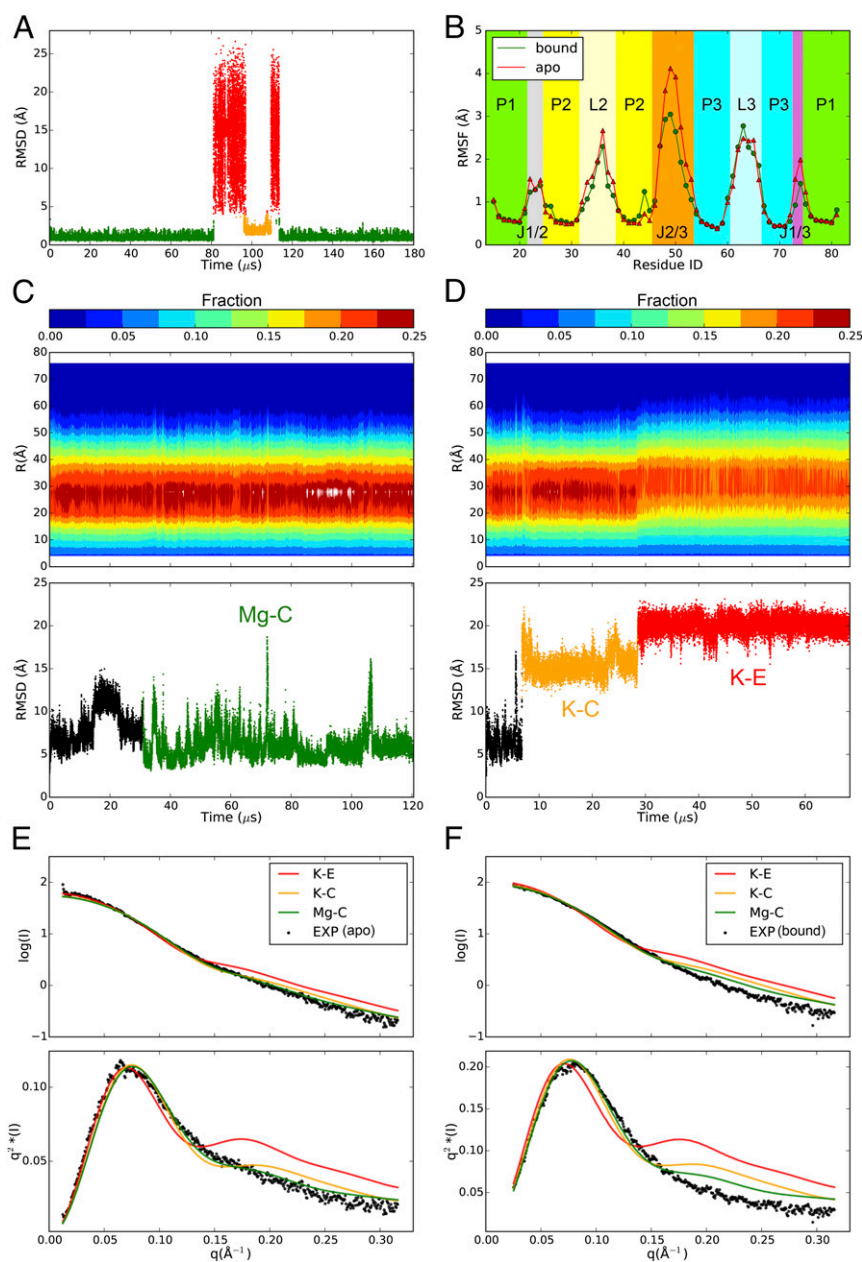
**Ab Initio Calculations.** Initial geometries of Watson–Crick A–U, Watson–Crick G–C, Hoogsteen A–U, Wobble G–U, Calcutta U–U, and Imino G–A base pairs, as well as the A–A and U–U stacked dimers with different orientations, were extracted from high-resolution crystal structures in the Protein Data Bank (PDB). The intermolecular separations of each nucleobase dimer were varied from 2.0 Å to 5.0 Å in 0.05-Å increments. For the base pairs, in vacuo QM counterpoise-corrected interaction energies were calculated at the MP2 level of theory, using density-fitting approximations (62), with an augmented triple-zeta basis set (aug-cc-pVTZ) using the MOLPRO program (63). The QM interaction energies of the stacked dimers were calculated with SCS(MI)-MP2, a semiempirical MP2-based method that is able to estimate the strength of nucleobase stacking only slightly less accurately than the much more computationally demanding CCSD(T) method (19, 23).

In vacuo potential energy surface (PES) scans of the  $\chi$  (O4'-C1'-N9-C4 in purines and O4'-C1'-N1-C6 in pyrimidines) torsion were performed for the four ribonucleosides. PES scans of the  $\gamma$  (O5'-C5'-C4'-C3') torsion for uridine and the  $\zeta$  (C3'-O3'-P'-O5') torsion for a model molecule mimicking the RNA backbone (SI Appendix, Fig. S6A) were performed either in vacuum or with the COSMO continuum solvent model (28). We left the  $\alpha$  torsion unchanged, since the AMBER ff14 RNA force field reproduces the QM PES scan well. The torsions were varied between  $-180^\circ$  and  $180^\circ$  in  $10^\circ$  increments. Each structure was relaxed at the MP2 level with a number of other dihedral angles constrained to ensure the smoothness of the PESs (SI Appendix, Table S1).

**Parameter Fitting.** Charges of the hydrogen bond donors in base pairs and the N1 and H2 atoms in adenine, as well as the LJ parameters of nucleobase nitrogen and polar hydrogen atoms, were optimized to minimize the following function:

$$\varepsilon = \sum_{i=1}^N (E_i^{QM} - E_i^{MM})^2 e^{-E_i^{QM}},$$

where  $E^{MM}$  and  $E^{QM}$  are the MM and QM interaction energies of the base-paired dimers. The differences between  $E^{MM}$  and  $E^{QM}$  were weighted by the



**Fig. 6.** MD simulations of riboswitches. (A) The nonhydrogen rmsd trace of the binding pocket of the ligand-bound guanine riboswitch aptamer. Green, orange, and red colors indicate the native bound conformation, the nonnative bound conformation, and the unbound conformation, respectively. (B) Root-mean-square fluctuation (RMSF) values of the C1' atoms in the simulations of the bound (green; only the frames where guanine was bound with the native conformation were included in the calculation) and apo (red) guanine riboswitch. The C1' atoms in the double-helical regions were aligned against for the calculation. (C and D) The distribution of all nonhydrogen pairwise interatomic distances (*Top*) and the nonhydrogen rmsd to the crystal structure (PDB ID code 3IQR) (*Bottom*) calculated for the MD simulations of the apo SAM-I aptamer with magnesium (C) or only potassium (D). (E and F) The theoretical SAXS profiles, shown as scattering curves (*Top*) and Kratky plots (*Bottom*), of the simulated Mg-C, K-C, and K-E ensembles compared with the experimental SAXS profiles of the apo (E) and bound (F) SAM-I aptamer in the presence of magnesium.

factor  $e^{-E_i^{QM}}$ , where  $E_i^{QM}$  is in kilocalories per mole. We found that the agreement between the MM and the QM interaction energies greatly improved by the optimization of the  $\sigma$  and  $\varepsilon$  values of the polar hydrogen atoms, and by optimizing the vdW parameters of the aromatic ring N atoms (N1, N3, N7, and N9) in adenine and the N atoms in the five-member ring of guanine (N7 and N9) separately from all of the other nucleobase N atoms (Table 2). The LJ parameters of nucleobase carbon atoms were subsequently optimized based on the interaction energies of the stacked dimers.

New  $\chi$  dihedral terms were obtained through fitting to the QM potential energy profiles with the same minimization method (*SI Appendix*,

*Fig. S3*). The force field potential energy,  $E^{MM}$ , was calculated with the following equation:

$$E^{MM}(\varphi) = E^{MM(\text{no dih})} + \sum_{m=1}^M k_m (1 + \cos(m\varphi - \theta_m)).$$

In this equation,  $E^{MM(\text{no dih})}$  was calculated with the dihedral term to be optimized removed from the MM force field.  $k_m$  and  $\theta_m$  are the parameters of the fit and represent the force constant and equilibrium dihedral angle for the  $m$ th term in the cosine expansion.

For the backbone  $\gamma$  and  $\zeta$  torsions, considerable differences were observed between the profiles calculated in vacuum and with the COSMO continuum



solvent (SI Appendix, Figs. S5 and S6). A combination of ab initio and empirical methods was thus used to derive the new dihedral terms. The populations of the *+gauche*, *-gauche*, and *trans* configurations were first calculated based on the MP2-COSMO profiles, and the dihedral terms were adjusted to ensure that similar populations were recapitulated in explicit-solvent MD simulations (SI Appendix, Tables S3 and S4).

The total number of terms,  $M$ , in the cosine expansion was four for the  $\chi$  torsion and three for the  $\gamma$  and  $\zeta$  torsions. The new  $k_m$  and  $\theta_m$  values are summarized in SI Appendix, Table S2.

**MD Simulations of ssRNA Tetramers.** The starting A-form structures of five ssRNA tetramers (AAAA, CAAU, CCCC, GACC, and UUUU) were generated with the program Coot (64) and solvated in cubic 40-Å<sup>3</sup> boxes containing 0.15 M NaCl. The TIP4P-D water model (21) and the CHARMM22 force field (65) were used to parameterize the water molecules and ions, respectively. As a comparison, CCCC was also parameterized with AMBER ff14, which includes the bsc0 and  $\chi$ OL modifications (14, 15), and the TIP3P water model (66). The systems were first equilibrated at 275 K for 50 ps using the Desmond software (67), and production simulations were performed at 275 K in the NPT ensemble (68–70) using Anton (71). Conformations of the ssRNA tetramers were categorized based on their base-stacking patterns, which were analyzed by the program DSSR (72). In both the A-form major (AMa) and A-form minor (AMi) conformations the four nucleobases are sequentially stacked. The main difference between AMa and AMi is that the 3'-end backbone  $\alpha$ -torsion is *+gauche* in AMa and *trans* in AMi. In the Nt1-flipped (F1) and Nt4-flipped (F4) conformations the ribose ring of the terminal nucleotide adopts the C2'-endo configuration and the base is flipped away from the other three. In the intercalated conformation (I), nucleotide  $j$  inserts between and stacks against nucleotides  $i$  and  $i + 1$  ( $j < i$  or  $j > i + 1$ ). Conformations other than the ones mentioned above (O) are mainly random coils.

**MD Simulations of rU<sub>40</sub>.** The initial linear A-form model of rU<sub>40</sub> was built with the program Coot. The molecule was solvated in 115-Å<sup>3</sup> boxes containing 0.05 M, 0.1 M, 0.2 M, or 0.4 M NaCl. Production runs were performed at 300 K in the NPT ensemble with Anton. The distance between the O5' atom of U1 and the O3' atom of U40,  $R$ , was converted into FRET efficiency,  $E_{FRET}$ , with the equation  $E_{FRET} = 1/(1 + (R/R_0)^6)$ , where  $R_0$  was set to 56.4 Å according to previous experimental results (36). The FRET-averaged end-to-end distance was calculated using  $\langle R_{FRET} \rangle = R_0(1/E_{FRET} - 1)^{1/6}$ . SEM of  $\langle E_{FRET} \rangle$  was estimated using a blocking procedure (73).

**MD Simulations of RNA Duplexes.** The double-stranded models of eight non-self-complementary RNA duplexes were generated with the program Coot. The systems were solvated in ~55-Å<sup>3</sup> boxes with 1 M KCl. Simulated tempering simulations following the protocol described in ref. 43 were performed on Anton. Exponentially distributed 24-rung temperature ladders spanning 100 K were used, with a temperature update interval of 18 ps.

The  $T_m$  value and the folding free energy at 310 K ( $\Delta G_{310 K}$ ) of an RNA duplex at 1 M standard concentration were estimated using the two-state model (45). The average folded fraction,  $f$ , in each rung was calculated based on the criterion that more than 50% of the original base pair hydrogen bonds should be present in the folded conformation. The data were then fitted to the nonlinear van't Hoff equation:

$$\ln K_{eq} = \ln \frac{2f}{(1-f)^2 C_t} = \frac{\Delta H_0 - T_0 \Delta C_p}{R} \left( \frac{1}{T_0} - \frac{1}{T} \right) + \frac{\Delta C_p}{R} \ln \frac{T}{T_0} + \ln K_0,$$

where  $C_t$  is the total concentration of single strands in the system,  $\Delta C_p$  is the temperature-independent heat capacity change,  $T_0$  is an arbitrarily selected

reference temperature, and  $\Delta H_0$  and  $K_0$  are the enthalpy and equilibrium constants at  $T_0$  (45, 74).

**MD Simulations of RNA Tetraloops.** The initial structures of the ggca-cUUCGgugcc, gccGAAAggc, and ggcGCAAgcc tetraloops were extracted from PDB entries 2KOC, 1JJ2, and 1ZIH, respectively. The systems were neutralized with potassium cations, and the simulated tempering protocols were similar to those of the RNA duplex simulations, except that the simulation of ggca-cUUCGgugcc used a 30-rung temperature ladder spanning 120 K.

The simulated structure of a tetraloop was categorized in the native, intermediate, or unfolded conformation based on the following criteria. (i) If not all of the Watson–Crick base pairs in the stem are formed, the structure is considered to be unfolded. (ii) If all of the Watson–Crick base pairs in the stem are formed, the overall rmsd value is lower than 2.5 Å, and all of the signature H-bond interactions in the loop in the experimental structure are formed, the structure is considered to be native. (iii) Structures that do not fall into either the unfolded or the native categories are designated intermediate.

**MD Simulations of Riboswitches.** The structure of the *B. subtilis* xpt-pbuX guanine riboswitch was extracted from PDB entry 4FE5 (50). The original ligand hypoxanthine in 4FE5 was either replaced by guanine or removed. The guanine-bound and apo molecules were solvated in 88-Å<sup>3</sup> boxes with 0.5 M NaCl and 0.1 M MgCl<sub>2</sub>. The nonbonded parameters of the atoms in the guanine base were transferred from our revised RNA force field, except that the charge of atom H9 was adjusted to make the net charge of the ligand equal to zero.

The structure of the SAM-bound *T. tencongensis* SAM-I riboswitch was extracted from PDB entry 3IQR (53), and the SAM ligand was removed for the simulations. Residue A94, which was mutated to G in the crystal structure, was mutated back to A. The molecule was solvated in 91-Å<sup>3</sup> boxes with either 0.5 M KCl or 0.5 M KCl and 0.1 M MgCl<sub>2</sub>.

The systems were first equilibrated at 293 K for 50 ps using Desmond (67), and production simulations were performed at 293 K in the NPT ensemble (68–70) on Anton (71).

**Calculation of the Theoretical SAXS Profiles.** Two hundred frames were randomly picked from each of the Mg–C, K–C, and K–E ensembles and converted to individual PDB files. The FoXS program (75) was used to calculate the SAXS profiles of these three model ensembles. The resulting theoretical scattering curves were then fitted to the experimental scattering curves of the Mg(+) apo or bound SAM-I aptamer retrieved from the BIOISIS database ([www.bioisis.net](http://www.bioisis.net)). The fit between the theoretical and experimental scattering curves was evaluated with the  $\chi$  value as defined by

$$\chi = \sqrt{\frac{1}{M} \sum_{i=1}^M \left( \frac{l_{exp}(q_i) - c l_{model}(q_i)}{\sigma_{exp}(q_i)} \right)^2},$$

where  $l_{exp}(q_i)$  and  $l_{model}(q_i)$  are the experimental and modeled SAXS intensities measured at wavevector  $q_i$ , respectively,  $c$  is the scaling factor to be fitted, and  $\sigma_{exp}(q_i)$  is the measured error in  $l_{exp}(q_i)$ .

**ACKNOWLEDGMENTS.** We thank Michael Eastwood and John Klepeis for helpful discussions and comments on the manuscript, Je-Luen Li, Kim Palmo, and Andrew Taube for help with QM calculations, and Rebecca Bish-Cornelissen and Berkman Frank for editorial assistance.

- Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE (2012) Biomolecular simulation: A computational microscope for molecular biology. *Annu Rev Biophys* 41:429–452.
- Mustoe AM, Brooks CL, Al-Hashimi HM (2014) Hierarchy of RNA functional dynamics. *Annu Rev Biochem* 83:441–466.
- Garst AD, Edwards AL, Batey RT (2011) Riboswitches: Structures and mechanisms. *Cold Spring Harb Perspect Biol* 3:a003533.
- Breaker RR (2012) Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* 4:a003566.
- Tubbs JD, et al. (2013) The nuclear magnetic resonance of CCCC RNA reveals a right-handed helix, and revised parameters for AMBER force field torsions improve structural predictions from molecular dynamics. *Biochemistry* 52:996–1010.
- Condon DE, et al. (2015) Stacking in RNA: NMR of four tetramers benchmark molecular dynamics. *J Chem Theory Comput* 11:2729–2742.
- Bergonzo C, et al. (2014) Multidimensional replica exchange molecular dynamics yields a converged ensemble of an RNA tetranucleotide. *J Chem Theory Comput* 10:492–499.
- Henriksen NM, Roe DR, Cheatham TE, 3rd (2013) Reliable oligonucleotide conformational ensemble generation in explicit solvent for force field assessment using reservoir replica exchange molecular dynamics simulations. *J Phys Chem B* 117:4014–4027.
- Schrodt MV, Andrews CT, Elcock AH (2015) Large-scale analysis of 48 DNA and 48 RNA tetranucleotides studied by 1  $\mu$ s explicit-solvent molecular dynamics simulations. *J Chem Theory Comput* 11:5906–5917.
- Bergonzo C, Henriksen NM, Roe DR, Cheatham TE, 3rd (2015) Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* 21:1578–1590.
- Sponer J, et al. (2014) Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J Phys Chem Lett* 5:1771–1782.
- Maier JA, et al. (2015) ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713.
- Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197.
- Pérez A, et al. (2007) Refinement of the AMBER force field for nucleic acids: Improving the description of  $\omega/\chi$  conformers. *Biophys J* 92:3817–3829.
- Zgarbová M, et al. (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J Chem Theory Comput* 7:2886–2902.

16. Ivani I, et al. (2016) Parmbsc1: A refined force field for DNA simulations. *Nat Methods* 13:55–58.
17. Yildirim I, Stern HA, Kennedy SD, Tubbs JD, Turner DH (2010) Reparameterization of RNA  $\chi$  torsion parameters for the AMBER force field and comparison to NMR spectra for cytidine and uridine. *J Chem Theory Comput* 6:1520–1531.
18. Banáš P, et al. (2012) Can we accurately describe the structure of adenine tracts in B-DNA? Reference quantum-chemical computations reveal overstabilization of stacking by molecular mechanics. *J Chem Theory Comput* 8:2448–2460.
19. Morgado CA, Jurečka P, Svozil D, Hobza P, Šponer J (2009) Balance of attraction and repulsion in nucleic-acid base stacking: CCSD(T)/complete-basis-set-limit calculations on uracil dimer and a comparison with the force-field description. *J Chem Theory Comput* 5:1524–1544.
20. Chen AA, García AE (2013) High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc Natl Acad Sci USA* 110:16820–16825.
21. Piana S, Donchev AG, Robustelli P, Shaw DE (2015) Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J Phys Chem B* 119:5113–5123.
22. Šponer J, Hobza P (1997) MP2 and CCSD(T) study on hydrogen bonding, aromatic stacking, and nonaromatic stacking. *Chem Phys Lett* 267:263–270.
23. Morgado CA, Jurečka P, Svozil D, Hobza P, Šponer J (2010) Reference MP2/CBS and CCSD(T) quantum-chemical calculations on stacked adenine dimers. Comparison with DFT-D, MP2.5, SCS(MI)-MP2, M06-2X, CBS(SCS-D) and force field descriptions. *Phys Chem Chem Phys* 12:3522–3534.
24. Stolarski R, Hagberg C-E, Shugar D (1984) Studies on the dynamic syn-anti equilibrium in purine nucleosides and nucleotides with the aid of <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy. *Eur J Biochem* 138:187–192.
25. Rosemeyer H, et al. (1990) Syn-anti conformational analysis of regular and modified nucleosides by 1D <sup>1</sup>H NOE difference spectroscopy: A simple graphical method based on conformationally rigid molecules. *J Org Chem* 55:5784–5790.
26. Steinbrecher T, Latzer J, Case DA (2012) Revised AMBER parameters for bioorganic phosphates. *J Chem Theory Comput* 8:4405–4412.
27. Gil-Ley A, Bottaro S, Bussi G (2016) Empirical corrections to the Amber RNA force field with target metadynamics. *J Chem Theory Comput* 12:2790–2798.
28. Klamt A, Schüürmann G (1993) COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans 2*, 799–805.
29. Beauchamp KA, Lin Y-S, Das R, Pande VS (2012) Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J Chem Theory Comput* 8:1409–1414.
30. Li S, et al. (2015) Molecular dynamics simulations of 441 two-residue peptides in aqueous solution: Conformational preferences and neighboring residue effects with the Amber ff99SB-ildn-NMR force field. *J Chem Theory Comput* 11:1315–1329.
31. Best RB, Buchete N-V, Hummer G (2008) Are current molecular dynamics force fields too helical? *Biophys J* 95:L07–L09.
32. Yildirim I, Stern HA, Tubbs JD, Kennedy SD, Turner DH (2011) Benchmarking AMBER force fields for RNA: Comparisons to NMR spectra for single-stranded r(GACC) are improved by revised  $\chi$  torsions. *J Phys Chem B* 115:9261–9270.
33. Bergonzo C, Cheatham TE, 3rd (2015) Improved force field parameters lead to a better description of RNA structure. *J Chem Theory Comput* 11:3969–3972.
34. Izadi S, Anandakrishnan R, Onufriev AV (2014) Building water models: A different approach. *J Phys Chem Lett* 5:3863–3871.
35. Mercadante D, et al. (2015) Kirkwood-Buff approach rescues overcollapse of a disordered protein in canonical protein force fields. *J Phys Chem B* 119:7975–7984.
36. Chen H, et al. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc Natl Acad Sci USA* 109:799–804.
37. Thrierr JC, Dourlent M, Leng M (1971) A study of polyuridylic acid. *J Mol Biol* 58:815–830.
38. Lipsett MN (1960) Evidence for helical structure in polyuridylic acid. *Proc Natl Acad Sci USA* 46:445–446.
39. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE (2011) How fast-folding proteins fold. *Science* 334:517–520.
40. Jiang F, Zhou C-Y, Wu Y-D (2014) Residue-specific force field based on the protein coil library. RSFF1: Modification of OPLS-AA/L. *J Phys Chem B* 118:6983–6998.
41. Lane TJ, Shukla D, Beauchamp KA, Pande VS (2013) To milliseconds and beyond: Challenges in the simulation of protein folding. *Curr Opin Struct Biol* 23:58–65.
42. Kůhrová P, Banáš P, Best RB, Šponer J, Otyepka M (2013) Computer folding of RNA tetraloops? Are we there yet? *J Chem Theory Comput* 9:2115–2125.
43. Pan AC, Weinreich TM, Piana S, Shaw DE (2016) Demonstrating an order-of-magnitude sampling enhancement in molecular dynamics simulations of complex protein systems. *J Chem Theory Comput* 12:1360–1367.
44. Marinari E, Parisi G (1992) Simulated tempering: A new Monte Carlo scheme. *Europhys Lett* 19:451–458.
45. Xia T, et al. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37:14719–14735.
46. Piana S, Klepeis JL, Shaw DE (2014) Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Curr Opin Struct Biol* 24:98–105.
47. Sheehy JP, Davis AR, Znosko BM (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* 16:417–429.
48. Johansen LE, Nygaard P, Lassen C, Agersø Y, Saxild HH (2003) Definition of a second *Bacillus subtilis* pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL). *J Bacteriol* 185:5200–5209.
49. Gilbert SD, Stoddard CD, Wise SJ, Batey RT (2006) Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol* 359:754–768.
50. Batey RT, Gilbert SD, Montange RK (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432:411–415.
51. Serganov A, et al. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* 11:1729–1741.
52. Stoddard CD, Gilbert SD, Batey RT (2008) Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* 14:675–684.
53. Stoddard CD, et al. (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure* 18:787–797.
54. Montange RK, Batey RT (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441:1172–1175.
55. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–1958.
56. Huang J, MacKerell AD, Jr (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem* 34:2135–2145.
57. Shan Y, et al. (2011) How does a drug molecule find its target binding site? *J Am Chem Soc* 133:9181–9183.
58. Joung IS, Cheatham TE, 3rd (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* 112:9020–9041.
59. Jensen KP, Jorgensen WL (2006) Halide, ammonium, and alkali metal ion parameters for modeling aqueous solutions. *J Chem Theory Comput* 2:1499–1509.
60. Chen AA, Pappu RV (2007) Parameters of monovalent ions in the AMBER-99 force-field: Assessment of inaccuracies and proposed improvements. *J Phys Chem B* 111:11884–11887.
61. Allnér O, Nilsson L, Villa A (2012) Magnesium ion-water coordination and exchange in biomolecular simulations. *J Chem Theory Comput* 8:1493–1502.
62. Werner H-J, Manby FR, Knowles PJ (2003) Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J Chem Phys* 118:8149–8160.
63. Werner H-J, Knowles PJ, Knizia G, Manby FR, Schütz M (2012) Molpro: A general-purpose quantum chemistry program package. *Wiley Interdiscip Rev Comput Mol Sci* 2:242–253.
64. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
65. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
66. Jorgensen WL (1981) Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J Am Chem Soc* 103:335–340.
67. Bowers KJ, et al. (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the ACM/IEEE SC 2006 Conference* (ACM, New York), p 43.
68. Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys* 81:511–519.
69. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A Gen Phys* 31:1695–1697.
70. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101:4177–4189.
71. Shaw DE, et al. (2009) Millisecond-scale molecular dynamics simulations on Anton. *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (ACM, New York), pp 1–11.
72. Lu X-J, Bussemaker HJ, Olson WK (2015) DSSR: An integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* 43:e142.
73. Flyvbjerg J, Petersen HG (1989) Error estimates on averages of correlated data. *J Chem Phys* 91:461–466.
74. Naghibi H, Tamura A, Sturtevant JM (1995) Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc Natl Acad Sci USA* 92:5597–5599.
75. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* 105:962–974.