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

Dazhi Tan*a,1, Stefano Piana*b,2, Robert M. Dirks*a,3, and David E. Shaw*a,b,2

*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.

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).

1Present address: Silicon Therapeutics, Boston, MA 02210.
2To whom correspondence may be addressed. Email: Stefano.Piana-Agostinetti@DeshawResearch.com or David.Shaw@DeshawResearch.com.
3Deceased February 3, 2015.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713027115/-/DCSupplemental.
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 C12 component of the Lennard-Jones (LJ) potential becomes excessively steep] (Fig. 1B 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. 1B, 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. 1A) and reducing the σ and ε 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 (χ) dihedral terms. Following the methods of previous parameterizations of the χ 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 χ dihedral terms were then parameterized to minimize the differences between the

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

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

Unlisted atoms have unchanged charges.
Table 2. The vdW atom types and parameters for nucleobase atoms in the revised force field

<table>
<thead>
<tr>
<th>vdW atom type</th>
<th>AMBER ff14</th>
<th>Revised AMBER ff14</th>
<th>Revised</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1.0691</td>
<td>0.0000</td>
<td>0.0157</td>
</tr>
<tr>
<td>NA</td>
<td>3.2500</td>
<td>3.2890</td>
<td>0.1700</td>
</tr>
<tr>
<td>NN</td>
<td>3.3507</td>
<td>0.1700</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>3.3997</td>
<td>3.3004</td>
<td>0.0860</td>
</tr>
<tr>
<td>CN</td>
<td>3.2850</td>
<td>0.0538</td>
<td></td>
</tr>
</tbody>
</table>

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 \( \xi \) 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 \( \xi \) torsions to match the in vacuo QM potential energy profiles led to erroneous populations of the \( \gamma \) and \( \xi \) 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 \( \pm \)gauche, \( \pm \)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 \( (\chi < 2) \) 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 solution (5–10, 32) (Fig. 24), but the large discrepancies between the simulated and experimental \( \chi \) angular-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, CCCG, 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. 28). Consequently, the \( \chi \) errors between the simulated and experimental \( \chi \) angular-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 \( \sim 20\% \) 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 \( (\chi \approx 4) \) 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 \( r_{	ext{UA}} \) 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 \( (R_{\text{FRET}}) \) (Fig. 3A and C) (36). The combination of AMBER ff14 and the TIP4P-D water model is also unable to prevent \( r_{	ext{UA}} \) from collapsing (SI Appendix, Fig. S19).

Simulations using our revised RNA parameters and the TIP4P-D water model, however, maintain the extended conformation of \( r_{	ext{UA}} \) and recapitulate the experimental \( R_{\text{FRET}} \) values at low salt concentration \( ([\text{NaCl}] < 0.1 \text{ M}) \) (Fig. 3B and C). In simulations under higher salt concentrations \( ([\text{NaCl}] > 0.1 \text{ M}) \), \( r_{	ext{UA}} \) alternates between extended and double-helical hairpin conformations containing up to 18 U–U base pairs, leading to lower \( R_{\text{FRET}} \) 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: \( \sim 10\% \), 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 gcca-cUUCGGguc, ggcGAAAggc, and ggcGCAAggc hyperstable tetraloops (Figs. 4 and 5 and SI Appendix, Figs. S11 and S12).
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 ($\Delta G_{310}$) calculated from the simulations were slightly more negative than the NN model (root-mean-square error = 1.4 kcal mol$^{-1}$, mean signed...
error = −1.1 kcal mol⁻¹; Fig. 4E), indicating that the revised parameters moderately overstabilize double-helical conformations, consistent with the results of the rU₄₀ simulations. The simulated Tₘ values were 5–50 K higher (mean signed error = +27.6 K) than the values predicted by the NN model. More stable duplexes tended to have larger discrepancies in Tₘ (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 Å 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 Å 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 Tₘ 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) formed 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 μs, at which point guanine escaped from the binding pocket into the solvent (Fig. 6A). The ligand was briefly recaptured by the pocket at 97 μs 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 μs before dissociating from the pocket. At 114 μs 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-μs 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-μs simulation performed with the AMBER ff14SB force field the overall riboswitch structure was maintained, but the binding-pocket structure was not conserved: After 20 μs 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...
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 of the SAM-I riboswitch. The rmsd traces and melting curves of all eight RNA duplexes are in *SI Appendix, Fig. S12*. (D) *Tm* values of the eight RNA duplexes calculated from the MD trajectories. In black are the *Tm* values estimated with the NN model (45). (E) *ΔG* 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 CGACCAG were higher than 310 K, so for these duplexes the *ΔG* values were extrapolated from the fitted ln*F* − 1/*T* curves.

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(++) 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 millisecond of simulation time (2), these results suggest that the K–E ensemble may represent an intermediate on the unfolding pathway.

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 χ, γ, and ζ 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. 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) *Tm* values of the eight RNA duplexes calculated from the MD trajectories. In black are the *Tm* values estimated with the NN model (45). (E) *ΔG* 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 CGACCAG were higher than 310 K, so for these duplexes the *ΔG* values were extrapolated from the fitted ln*F* − 1/*T* curves.
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 yO3L correction for RNA (15) and the bscl 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 UURG long hairpin, the RNA duplexes, and the tetraloops. This overestimation may also contribute to the excessive ordering of the UUUU tetranucleotide and the J1/J2 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 bio-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 $\gamma$ (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:

$$
\epsilon = \sum_{i=1}^{N} (E^{MM} - E^{QM})^2 e^{-\epsilon^{quad}},
$$

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
factor $e^{-E_{QM}}$, where $E_{QM}$ 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. 

---

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.
solvant (SI Appendix, Figs. S5 and S8). A combination of ab initio and empirical methods was thus used to derive the new dihedral terms. The populations of the \( \text{g} \)-gauche, \( \text{g} \)-gauche, and trans conformations 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 \( \gamma \) torsion and three for the \( \chi \) and \( \varphi \) torsions. The new \( m \) and \( n \) values are summarized in SI Appendix, Table S2.

**MD Simulations of ssRNA Tetramers.** The starting A-form structures of five ssRNA tetramers (AAAA, CAUU, CCCC, GACC, and UUUU) were generated with the program Coot (64) and solvated in cubic 40-Å\(^3\) boxes containing 0.15 M NaCl. The TIP4P 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 ff94, which includes the bsol and yOL 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 \( \gamma \)-torsion is \( \text{g} \)-gauche in AMa and trans in AMI. In the Nt1-flipped (F1) and Nt4-flipped (F4) conformations the ribose ring of the central nucleotide adopts the C2-endo configuration and the base is flipped away from the other three. In the intercalated conformation (I), nucleotide \( i \) inserts between and stacks against nucleotides \( i-1 \) and \( i+1 \) (\( j < i < k \)). Conformations other than the ones mentioned above (O) are mainly random coils.

**MD Simulations of rU1354.** The initial linear A-form model of rU1354 was built with the program Coot. The molecule was solvated in 115-Å\(^3\) 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 OS \( \text{atom of U1 and the OS atom of U40, } R \), was converted into FRET efficiency, \( E_{\text{FRET}} \), with the equation \( E_{\text{FRET}} = 1/1 + (R/R_0)^6 \), where \( R_0 \) was set to 56.4 Å according to ref. 56. 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) using Anton (71).

**MD Simulations of Riboswitches.** The structure of the B. subtilis xpt-pbxU 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\(^3\) boxes with 0.5 M NaCl and 0.1 M MgCl\(_2\). 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. tengcongensis 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\(^3\) boxes with either 0.5 M KCl or 0.5 M KCl and 0.1 M MgCl\(_2\).

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) using 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\(^{2+}\) apo or bound SAM-I aptamer retrieved from the BIOISIS database (www.bioisis.net). The fit between the theoretical and experimental scattering curves was evaluated with the \( \chi \) value as defined by

\[
\chi = \frac{\sum_{i=1}^{M} \left( \frac{I_{\exp}(q) - I_{\text{model}}(q)}{\sigma_{\text{mod}}(q)} \right)^2}{\sum_{i=1}^{M} \sigma_{\text{mod}}(q)^2},
\]

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

**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.