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

Supporting Information Appendix

Details of the Conformational Changes Observed in the Simulations of the apo SAM-I Riboswitch Aptamer

The aptamer domain of the SAM-I riboswitch consists of four helical regions (P1–P4) and three single-stranded joining segments (J1/2, J3/4, and J4/1) (SI Appendix, Figure S14A). Its tertiary structure is stabilized by the P1–P4 and P2–P3 coaxial stacking interactions, as well as a pseudoknot between J3/4 and the hairpin loop of P2. The ligand S-adenosylmethionine (SAM) binds to a pocket formed by P1, P3, and J1/2, where an internal loop consisting of residues A45, A46, and U57 recognizes the adenosyl moiety of SAM (SI Appendix, Figure S15A). The crystal structure of the apo state of the SAM-I riboswitch aptamer is highly similar to that of the bound state except for minor differences in the A45-A46-U57 internal loop (SI Appendix, Figure S15A) (1, 2).

Conformations in the Mg-C ensemble (Figure 6C) feature a relatively stable core consisting of the J3/4-P2 pseudoknot and the coaxially stacked P2 and P3 segments (SI Appendix, Figure S14B). The coaxially stacked P1 and P4 segments exhibited large degrees of concerted motion relative to the core (SI Appendix, Figures S14B and S14C). The major difference between Mg-C and K-C (Figure 6D) is the positioning of P4, which is rotated by about 90 degrees relative to the P2-P3-J3/4 core in K-C, making its axis almost perpendicular to that of the P2–P3 region (SI Appendix, Figure S14D). In the K-E ensemble (Figure 6D), two base-triples (G22-C30-A61 and

G23-C29-A62) that mediate the tertiary interactions between P2 and J3/4 are disrupted, thus shifting P3 away from P1 and P4 and resulting in a less compact structure (SI Appendix, Figure S14E).

In addition to the magnesium-dependent global conformational changes, we also observed two major states of the A45-A46-U57 internal loop in both the Mg+ and Mg− simulations. In Conformation A, A46 and U57 form a Watson-Crick base pair, and A45 is stacked against A46 instead of forming the coplanar base-triple as in the crystal structure (SI Appendix, Figure S15A). In Conformation B, the base-triple is present but adopts a different binding mode, with U57 base-paired to the Watson-Crick edge of A45 rather than A46 (SI Appendix, Figure S15A). These two local conformations were interconvertible and together accounted for more than 95% of the trajectory (SI Appendix, Figure S15B). This observation is in agreement with the fact that U57 is capable of disengaging from A45/A46 and binding to SAM when it is present in solution. It also aligns with the results of earlier NMIA probing assays indicating that both A45 and A46 were able to access unconstrained conformations in the absence of the ligand in contrast to U57, which was static (1).

SI Appendix Tables and Figures

Scan	Dihedral	Values (degrees)	
		C2′-endo	C3′-endo
χ	H5T–O5′– C5′– C4′	174	174
	O5′– C5′– C4′– C3′	172	54
	C5′– C4′– C3′– O3′	140	81
	C4′– C3′– O3′– H3T	−148	−148
	O4′– C1′– C2′– C3′	32	−24
	O3′– C3′– C2′– O2′	−35	39
	H3′– C3′– C2′– H2′1	−35	39
	C1′– C2′– O2′– HO′2	−153	−153
γ	H5T–O5′– C5′– C4′	−180	−180
	C5′– C4′– C3′– O3′	140	81
	C4′– C3′– O3′– H3T	−148	−148
	O4′– C1′– C2′– C3′	32	−24
	O3′– C3′– C2′– O2′	−35	39
	H3′– C3′– C2′– H2′1	−35	39
	C1′– C2′– O2′– HO′2	−153	−153
	H4′– C4′– C3′– O3′	22	−37
ζ	H5T–O5′– C5′– C4′	−180	−180
	C5′– C4′– C3′– O3′	140	81
	C4′– C3′– O3′– H3T	−148	−148
	O4′– C1′– C2′– C3′	32	−24
	O3′– C3′– C2′– O2′	−35	39
	H3′– C3′– C2′– H2′1	−35	39
	C1′– C2′– O2′– HO′2	−153	−153
	H4′– C4′– C3′– O3′	22	−37
	O3′– P– O5′– C5′	180	180
	O5′– C5′– C4′– C3′	180	180

Table S1. Constrained dihedral angles in the potential energy surface (PES) scans for the χ, γ, and ζ torsion angles.

Dihedral		k_1	k_2	k_3	k_4	θ_1	θ_2	θ_3	θ_4
χ	C	0.678707	1.195034	−0.892686	−0.310165	111.18	−19.77	28.20	−4.92
	U	0.607129	1.297516	−0.499852	−0.165134	147.46	−13.58	38.13	−12.00
	G	−0.912264	1.378165	−0.425553	0.060088	−72.86	−28.03	27.76	−31.87
	A	0.751335	1.096015	−0.413416	0.064750	98.29	−31.44	50.52	40.99
γ		−0.359947	1.228048	0.780996	–	−60.31	−85.90	−17.79	–
ζ		−0.277219	1.066046	0.188299	–	−26.02	2.98	29.47	–

Table S2. χ, γ, and ζ dihedral terms in the revised force field. The force constants (k) are in kcal mol^{-1} and the phase angles (θ) are in degrees.

		C2′-endo					**C3′-endo**				
		χ		γ			χ		γ		
		syn	*anti*	*g*+	*g*–	*trans*	*syn*	*anti*	*g*+	*g*–	*trans*
rC	Revised params	11	88	81	3	15	8	92	77	2	19
	Amber ff14	34	63	91	6	3	11	87	91	3	5
rU	Revised params	5	94	83	3	13	7	92	81	1	16
	Amber ff14	16	80	91	7	2	12	85	93	3	4
rG	Revised params	45	52	81	4	14	41	58	79	2	17
	Amber ff14	74	18	94	4	2	60	35	93	3	4
rA	Revised params	26	71	83	3	12	16	82	81	2	16
	Amber ff14	88	7	97	2	1	67	27	95	3	2
rU	QM-COSMO	–	–	84	6	10	–	–	87	0	12

Table S3. Percentage of different χ and γ torsion conformers observed in the explicit-solvent MD simulations of ribonucleosides. The definitions of the conformers are as follows: χ-*syn*, χ ∈ (0°, 100°); χ-*anti*, χ ∈ (−180°, −90°) or χ ∈ (160°, 180°); γ-*g*+, γ ∈ (30°, 90°); γ-*g*–, γ ∈ (−90°, −30°); γ-*trans*, γ ∈ (−180°, −150°) or γ ∈ (150°, 180°). The nucleosides were solvated in cubic 403 Å^3 boxes containing ~2000 water molecules. All simulations were performed at 300 K in the NPT ensemble for 6 μs on Anton.

		ζ *g*+	ζ *g*–	ζ *trans*
ζ-model	Revised params	50	41	9
	Amber ff14	55	34	11
	QM-COSMO	48	39	13

Table S4. Percentage of different ζ torsion conformers observed in the explicit-solvent MD simulations of the ζ-model molecule mimicking the RNA backbone (SI Appendix, Figure S6A). The definitions of the conformers are as follows: ζ-*g*+, ζ ∈ (30°, 90°); ζ-*g*–, ζ ∈ (−90°, −30°); ζ-*trans*, ζ ∈ (−180°, −150°) or ζ ∈ (150°, 180°). The ζ-model molecule was solvated in a cubic 403 Å^3 box containing ~2000 water molecules. The simulations were performed at 300 K in the NPT ensemble for 6 μs on Anton.

^3J-coupling	#	AAAA			CAAU		CCCC				GACC		UUUU	
		NMR	R	A14	NMR	R	NMR	R	A14	A14+ OPC	NMR	R	NMR	R
H5′–P (β)	2	3.8	3.3	3.7	3.7	3.8	3.8	3.6	3.5	3.6	3.7	3.2	2.0	3.6
	3	3.0	3.5	3.5	3.5	3.7	3.9	3.9	4.3	3.5	4.0	3.4	2.0	3.6
	4	3.2	3.2	2.6	4.05	3.1	3.8	3.5	2.5	3.1	4.4	3.5	2.0	3.0
H5″–P (β)	2	1.0	2.3	2.3	2.2	2.0	1.2	2.1	2.1	2.0	0.9	2.3	2.0	2.3
	3	1.0	2.1	2.2	<1	1.9	0.5	1.9	1.8	2.7	2.0	2.2	2.0	2.3
	4	1.0	2.4	3.3	3.3	2.7	1.1	2.3	3.2	2.7	2.0	2.2	2.0	2.8
H4′– H5′ (γ)	1	3.8	2.6	3.0	1.6	2.6	–	2.6	2.6	2.9	4.6	2.6	3.7	2.5
	2	~2	2.5	2.6	~2.8	2.7	<1	2.5	2.4	2.4	~2	2.5	~2.5	2.6
	3	2.0	2.6	2.6	2.0	2.7	2.1	2.6	2.7	2.6	1.5	2.6	~2.5	2.6
	4	2.0	2.7	3.0	1.0	2.5	2.0	2.5	2.7	2.5	1.8	2.5	–	2..4
H4′– H5″ (γ)	1	2.0	2.4	1.3	~3.8	2.9	–	2.2	1.7	1.2	2.3	2.9	2.8	2.4
	2	~1	1.1	1.1	~2.1	1.1	<1	1.1	1.3	1.1	~2	1.0	~2.5	1.1
	3	2.0	0.9	1.0	1.3	0.8	<1	1.0	0.9	1.0	<1	0.9	~2.5	1.2
	4	2.0	0.9	1.2	<1	1.2	1.4	1.1	0.9	1.1	1.3	1.0	–	1.3
H3′–P (ε)	2	8.45	7.6	5.6	9.6	7.8	8.8	8.4	6.4	7.9	9.3	7.8	8.2	6.4
	3	8.7	7.8	6.4	8.4	7.5	9.3	8.8	6.6	8.1	9.1	8.2	7.8	6.2
	4	8.35	7.3	4.9	8.3	6.6	9.3	8.1	5.5	7.4	9.0	8.4	7.7	5.9
H1′– H2′ (ν_1)	1	2.0	1.5	3.5	~1	1.9	<1	1.7	1.5	2.7	1.8	1.4	4.5	2.5
	2	1.0	1.3	2.2	~1	1.4	<1	1.6	3.0	2.6	<1	1.3	5.2	2.3
	3	1.0	1.3	2.0	1.4	1.4	<1	1.4	1.7	2.8	<1	1.5	5.2	2.2
	4	2.75	2.6	3.9	3.7	4.6	2.6	3.0	3.3	3.7	2.5	2.8	3.7	5.2
H2′– H3′ (ν_2)	1	3.4	4.5	5.3	4.5	4.6	3.6	4.5	4.5	4.7	4.4	4.6	4.7	4.8
	2	3.6	4.4	5.2	4.6	4.5	4.3	4.5	4.8	4.7	4.4	4.5	4.6	4.8
	3	4.0	4.4	5.3	4.3	4.7	4.7	4.6	4.6	4.9	4.0	4.6	4.8	4.9
	4	–	5.0	5.4	3.7	5.2	4.9	5.0	5.3	5.2	4.6	5.0	–	5.3
H3′– H4′ (ν_3)	1	5.7	8.1	6.0	8.4	8.1	8.7	8.1	8.1	7.1	7.9	8.0	5.3	7.6
	2	7.3	8.3	6.6	8.3	8.1	–	8.1	6.2	6.9	8.9	8.2	5.5	7.5
	3	8.0	8.3	7.1	8.3	8.1	8.8	8.2	7.8	6.7	8.8	8.2	5.4	7.8
	4	6.0	6.5	4.6	–	4.7	7.2	6.4	5.9	5.7	7.2	6.7	–	4.2
χ^2		–	1.96	5.73	–	2.20	–	1.43	6.19	3.19	–	2.04	–	4.05

Table S5. Experimental and simulated ^3J scalar coupling data for the ssRNA tetramers. All values are in Hz. R stands for the revised parameters and A14 stands for Amber ff14. OPC indicates the OPC water model (3). The NMR values and the Karplus relations were taken from

(4) and (5). According to (4) and (5), the errors of the Karplus relations used for calculating the χ^2 values are 0.5 Hz for ^3J couplings involving the backbone torsions (β, γ, and ε) and 1.5 Hz for the ribose-ring torsions ($\nu1$, $\nu2$, and $\nu3$).

Tetraloop	T_m (K)		Fraction of the native conformation in the lowest temperature (280 K) ensemble
	Experiment	Simulation	
ggcacUUCGgugcc	347	365 ± 10	0.40
gccGAAAggc	343	353 ± 7	0.66
ggcGCAAgcc	344	357 ± 10	0.46

Table S6. Summary of the MD simulations of three hyperstable tetraloops. The experimental T_m value of the UUCG tetraloop is from (6), and the experimental T_m values of the GAAA and GCAA tetraloops are from (7).

Theoretical profiles	**Experimental profiles**	
	Mg+, apo SAM-I aptamer	Mg+, bound SAM-I aptamer
Mg-C	1.74	2.27
K-C	1.95	3.78
K-E	2.83	7.36
Crystal structure (3QIR)	2.82	1.86

Table S7. χ errors between the theoretical SAXS profiles of the simulated ensembles of the apo SAM-I aptamer and the experimental SAXS profiles. Mg-C stands for the compact ensemble observed in the simulation with Mg^{2+}. K-C and K-E stand for the compact and extended ensembles observed in the simulation without Mg^{2+}, respectively.

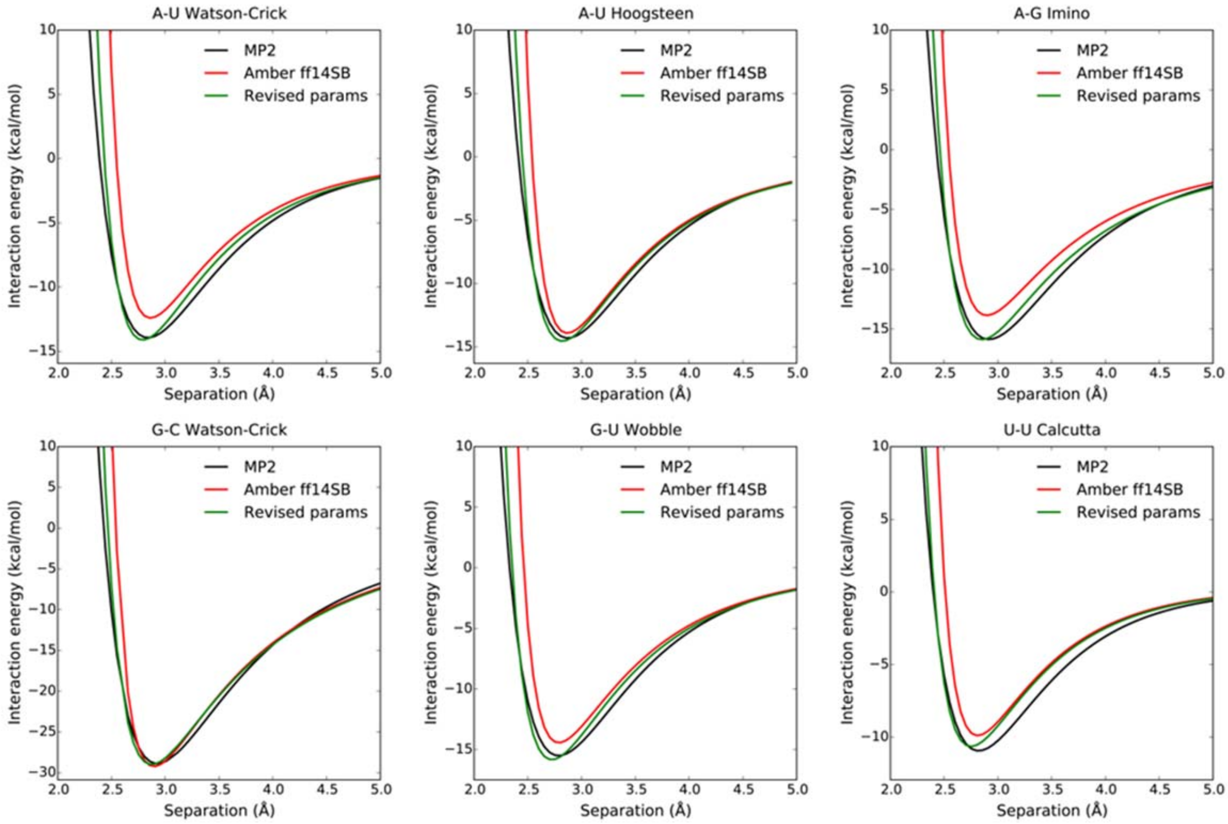


Figure S1. QM and MM dimer interaction energies of six common types of base pairs as functions of intermolecular separations.

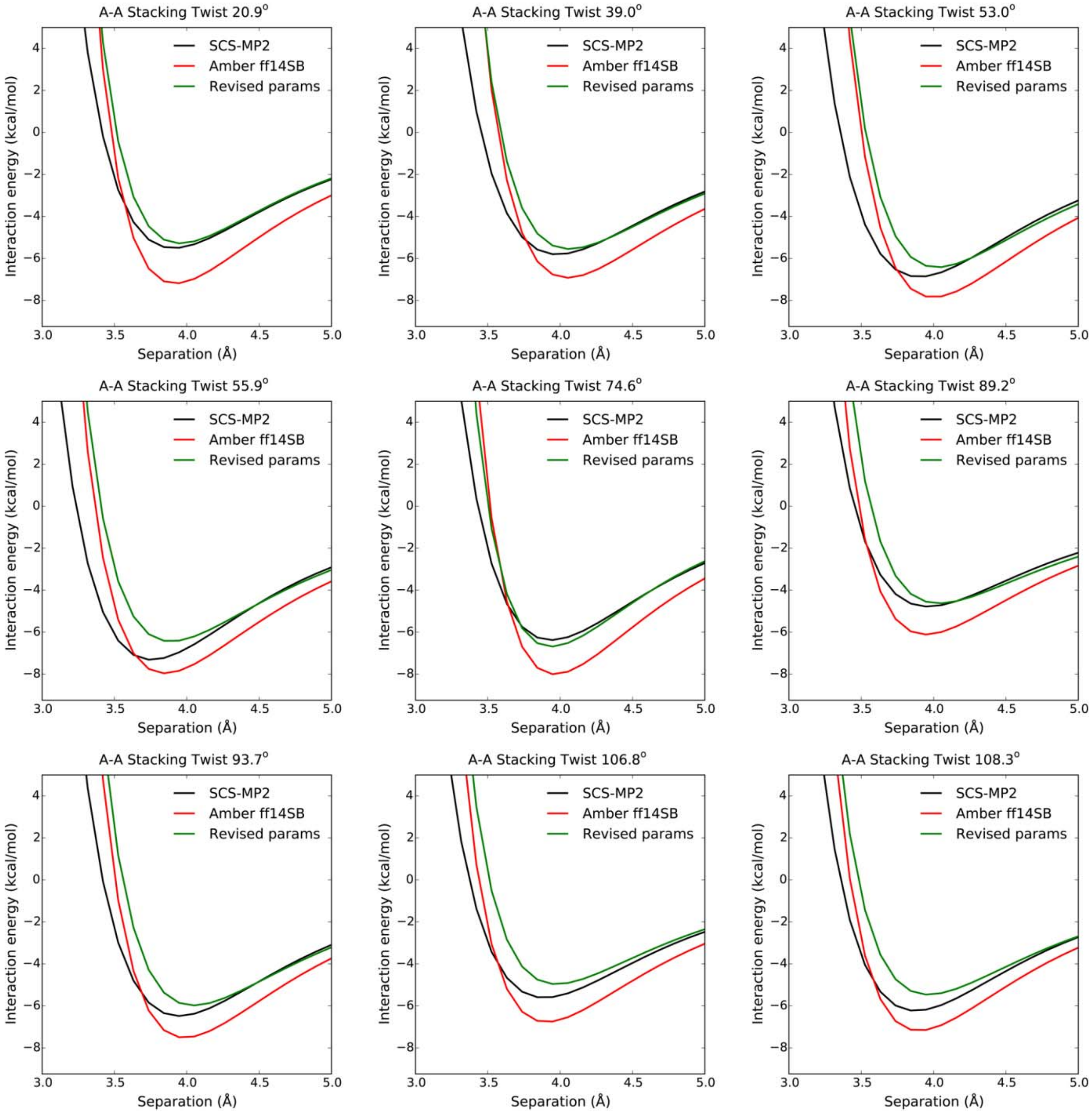
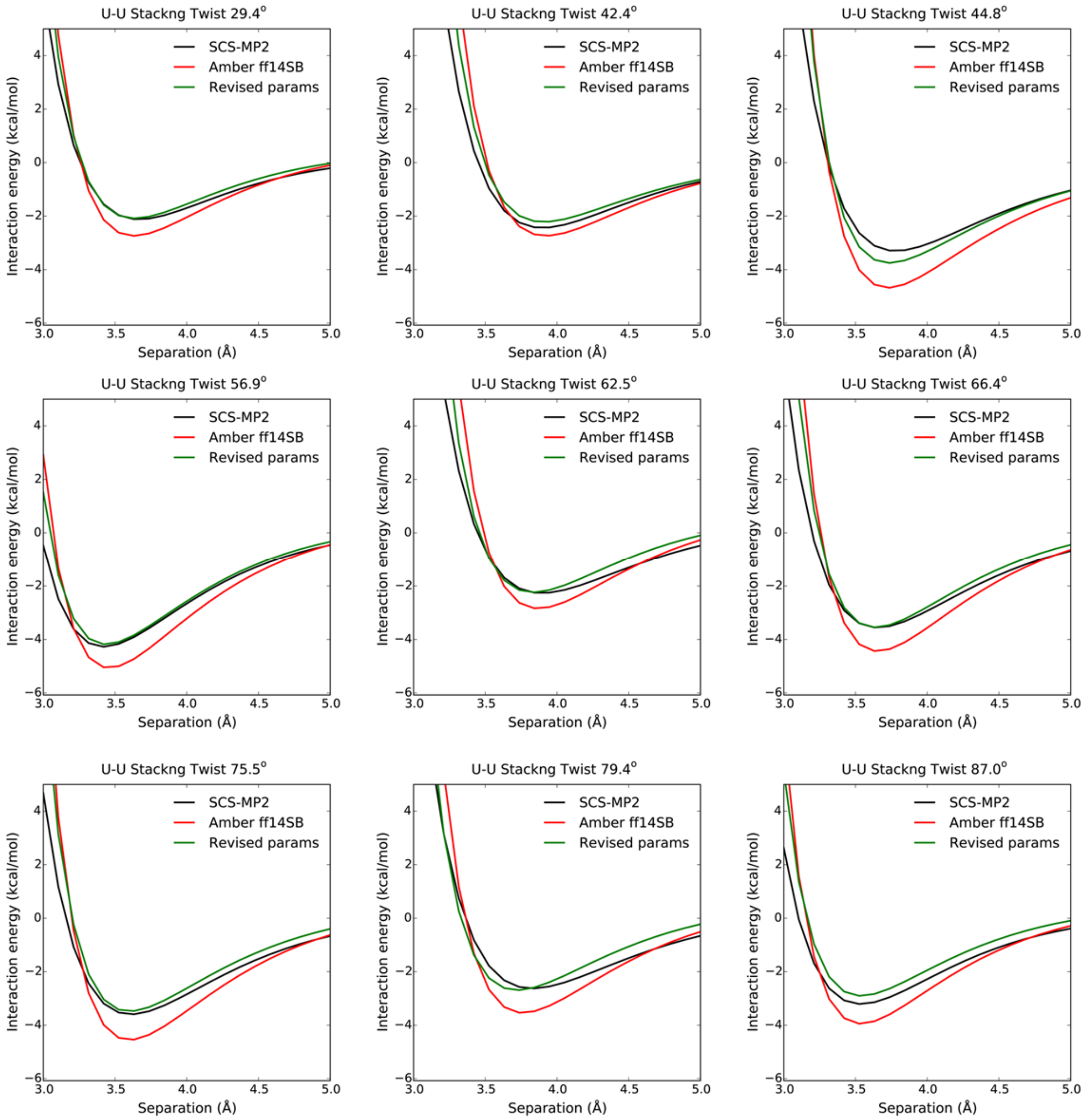


Figure S2. QM and MM dimer interaction energies of A-A stacked dimers at varying relative orientations as functions of intermolecular separations. The twist angle is defined as the angle between the C1′-N9 bonds of the two adenines.


U-U Stackng Twist 29.4°

U-U Stackng Twist 42.4°

U-U Stackng Twist 44.8°

U-U Stackng Twist 56.9°

U-U Stackng Twist 62.5°

U-U Stackng Twist 66.4°

U-U Stackng Twist 75.5°

U-U Stackng Twist 79.4°

U-U Stackng Twist 87.0°

Figure S3. QM and MM interaction energies of U-U stacked dimers at varying relative orientations as functions of intermolecular separations.

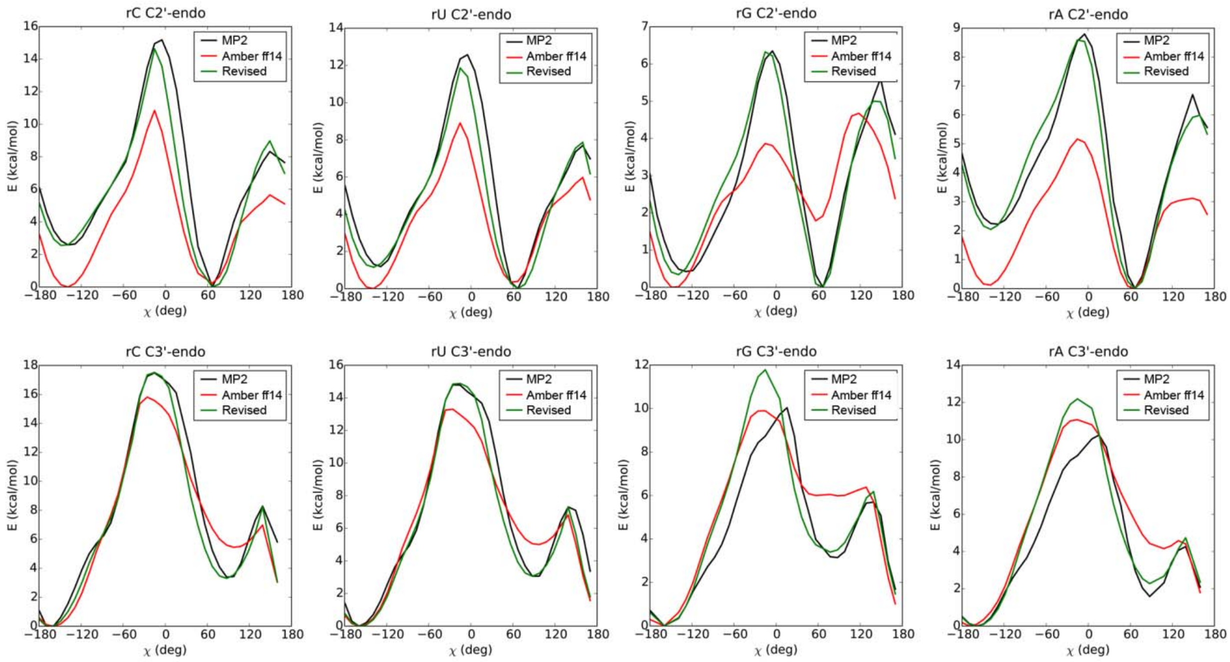


Figure S4. Potential energy profiles of the χ torsion angle in the C, U, G, and A ribonucleosides in vacuum. The vertical axes represent relative potential energies with the lowest point in each scan set to zero. The red curves represent the potential energy profiles without the χ dihedral terms.

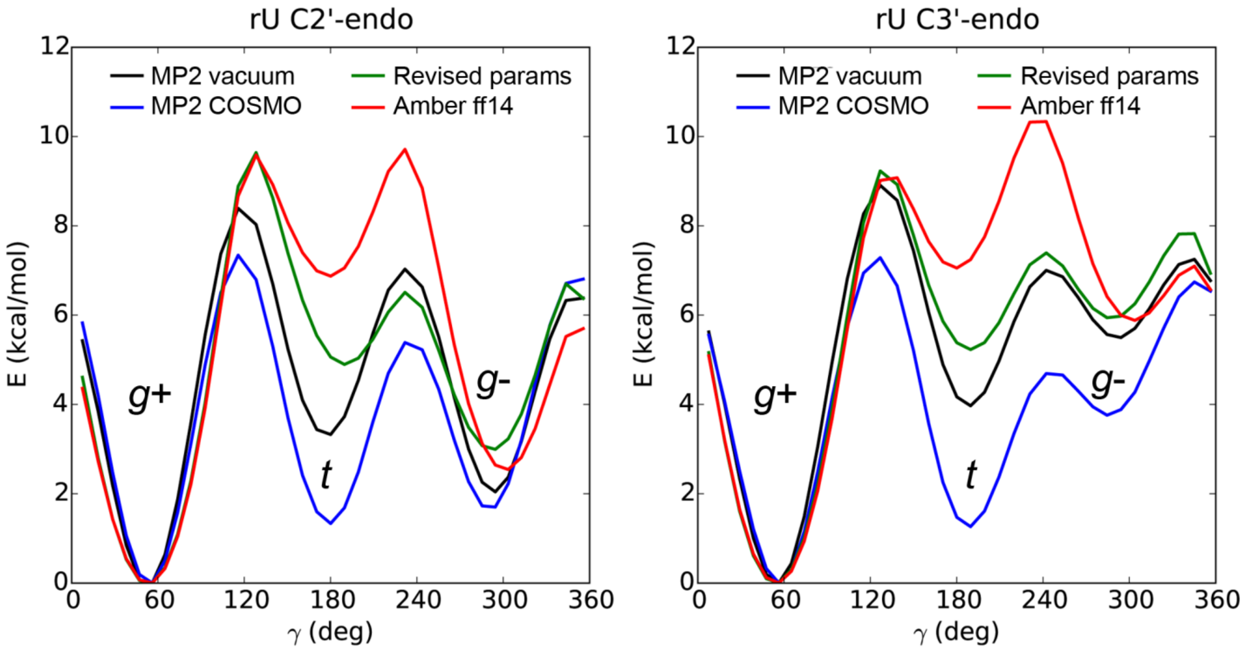


Figure S5. Potential energy profiles of the γ torsion angle in uridine. The vertical axes represent relative potential energies with the lowest point in each scan set to zero. Positions of the +*gauche* (*g*+), −*gauche* (g−), and *trans* (*t*) configurations are indicated. The MM potential energies were calculated in vacuum.

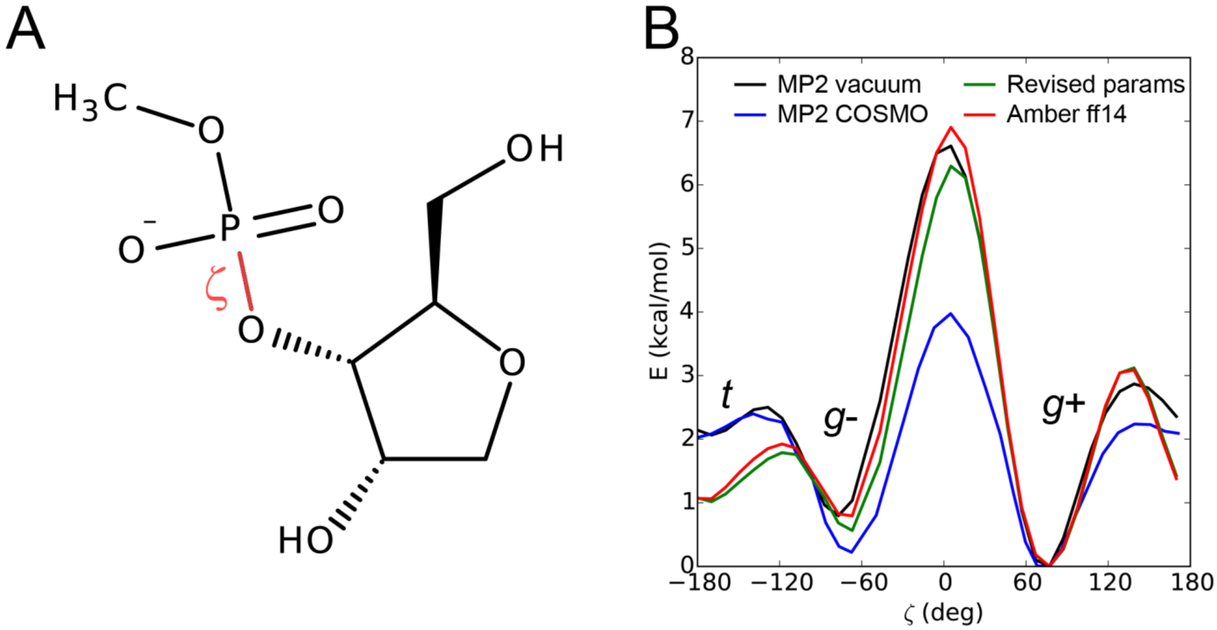


Figure S6. (A) The model molecule mimicking the RNA backbone used for the PES scan of the ζ torsion angle. (B) Potential energy profiles of the ζ torsion angle. The vertical axes represent relative potential energies with the lowest point in each scan set to zero. Positions of the +*gauche* (*g*+), −*gauche* (*g*−), and *trans* (*t*) configurations are indicated. The MM potential energies were calculated in vacuum.

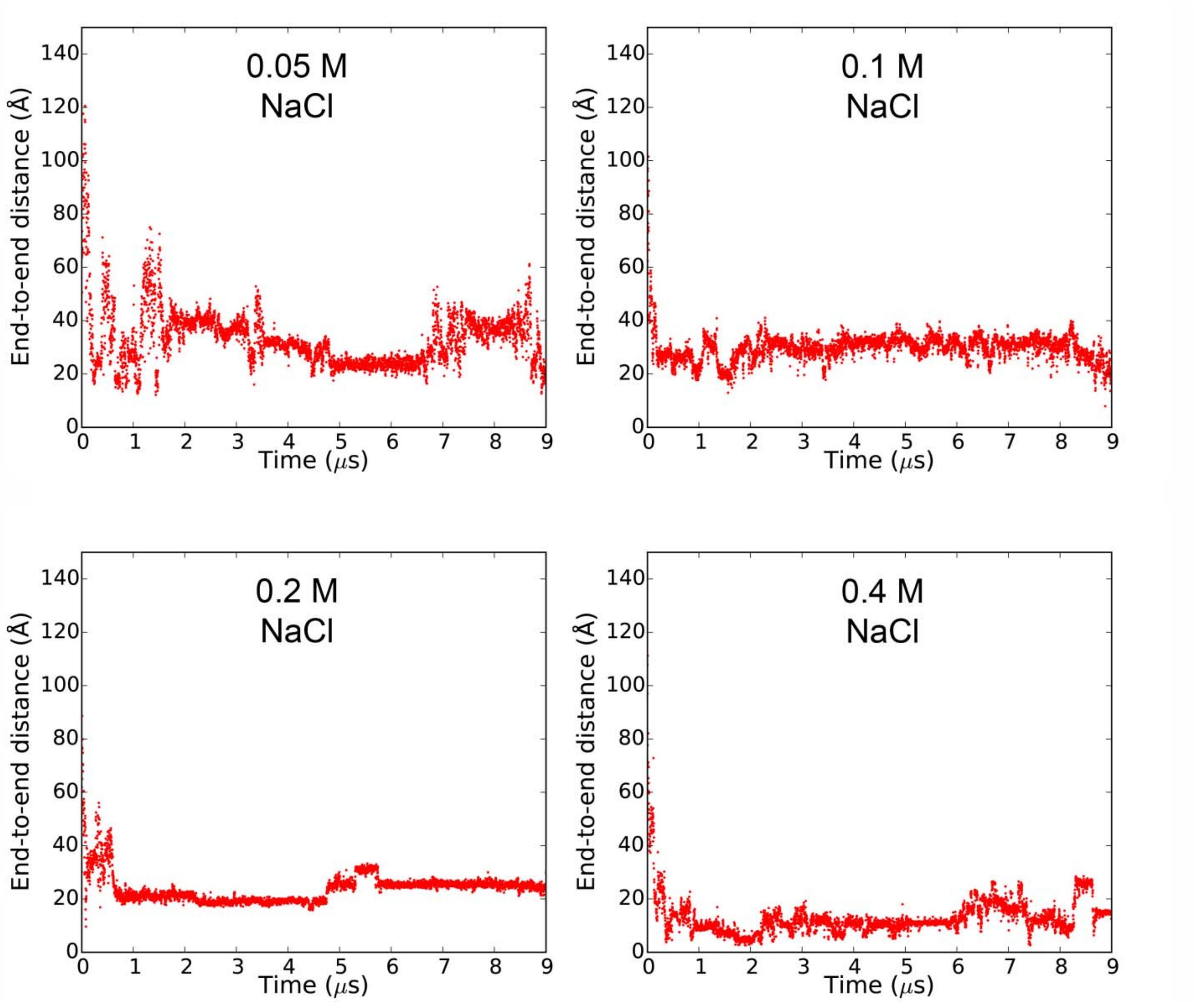


Figure S7. The end-to-end distance of rU_{40} in simulations using the Amber ff14 force field at 300 K under different NaCl concentrations.

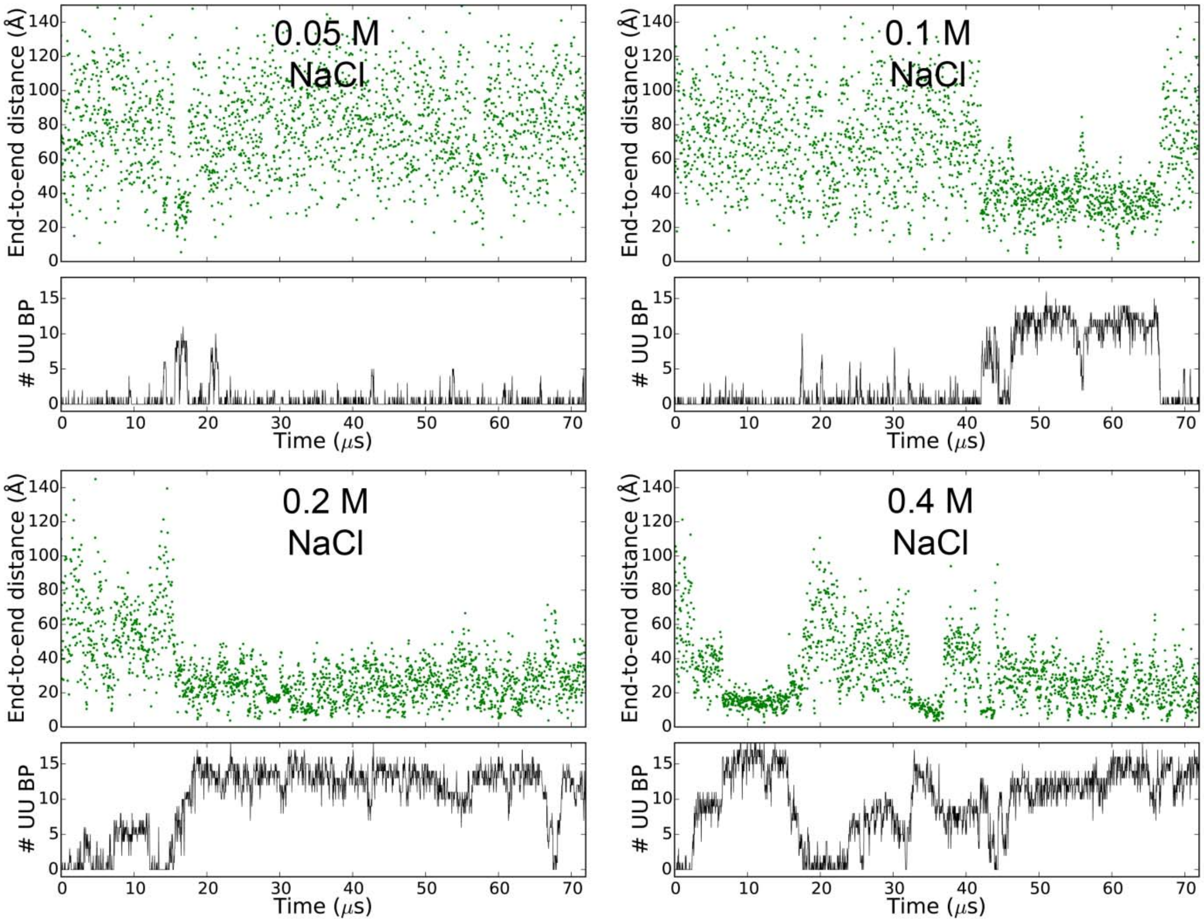


Figure S8. The end-to-end distance of rU_{40} and the number of U-U base-pairs in rU_{40} in simulations using the revised parameters at 300 K under different NaCl concentrations. U-U base-pairs were detected using the program DSSR (8).

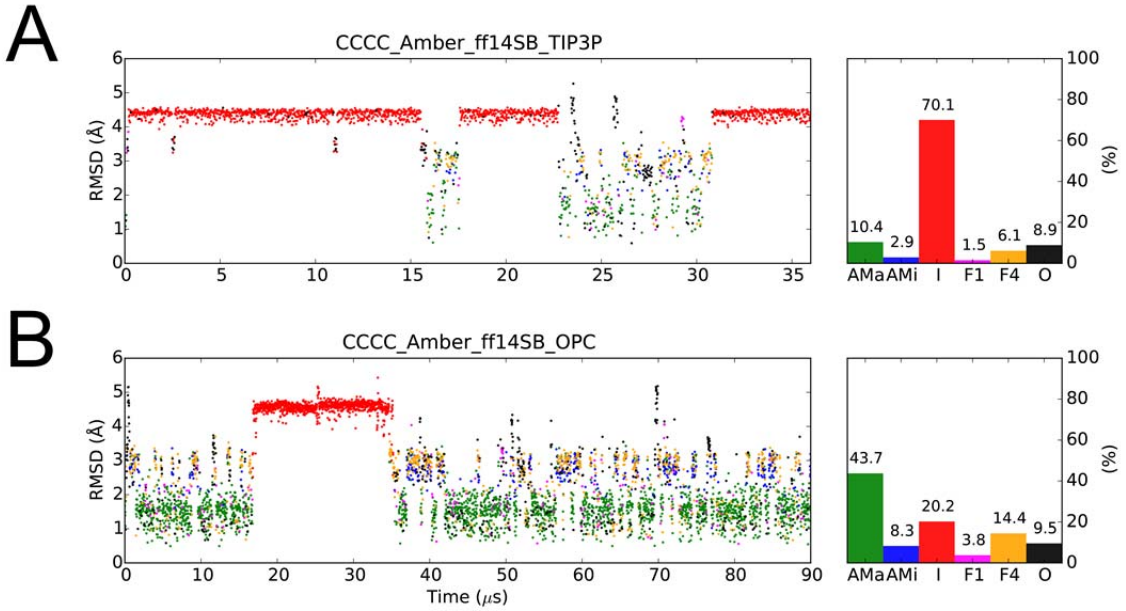


Figure S9. MD simulations of ssRNA tetramers using Amber ff14 and the TIP3P water model (A) or the OPC water model (B). The scatter plots on the left indicate RMSDs of the backbone (including the ribose) non-hydrogen atoms to the initial A-form structures. Based on the stacking patterns of the four nucleotides, 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.

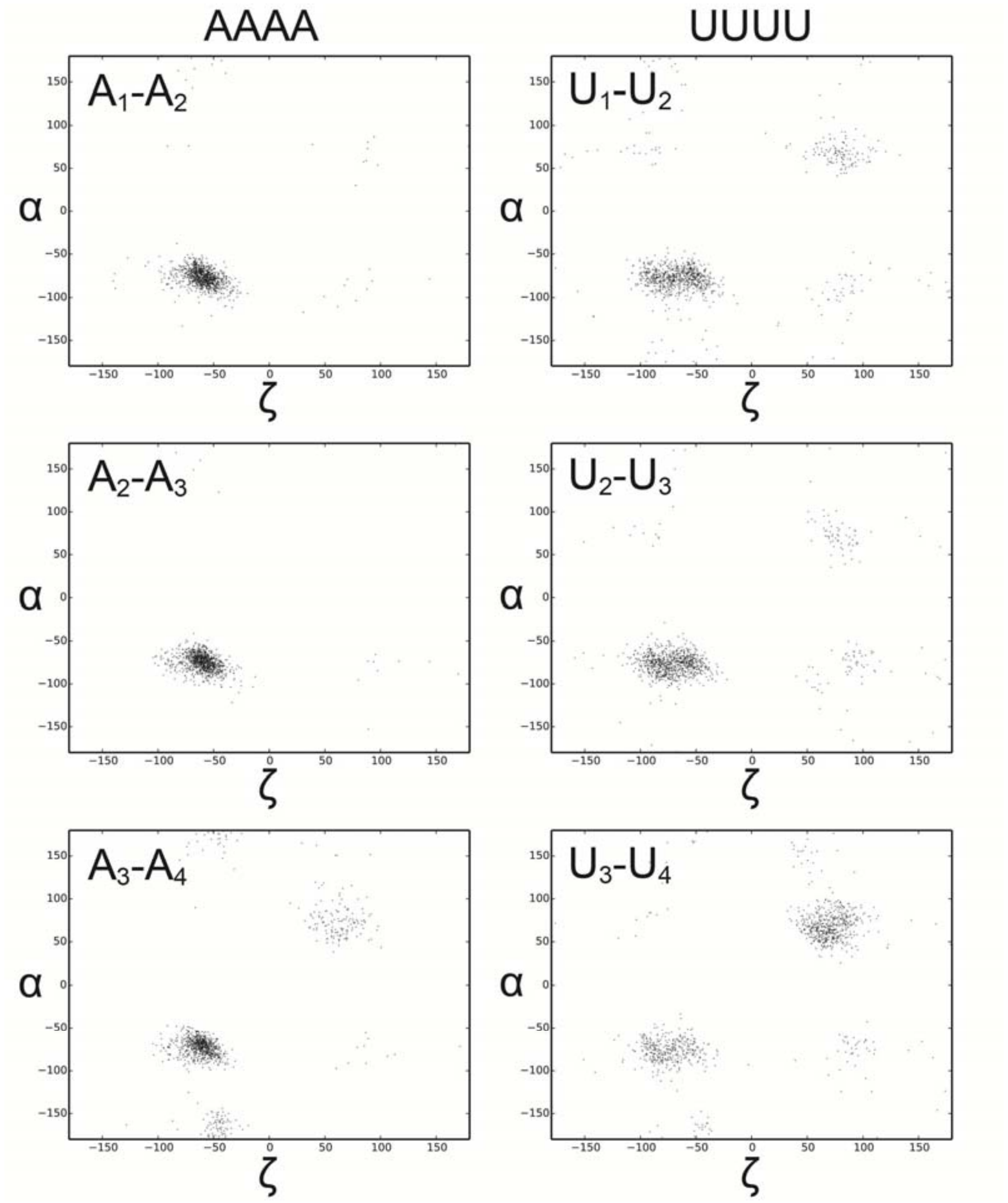


Figure S10. MD simulations of ssRNA AAAA and UUUU tetramers using the revised force field. Scatter plots of the α and ζ angles observed during the 90-μs simulations are shown.

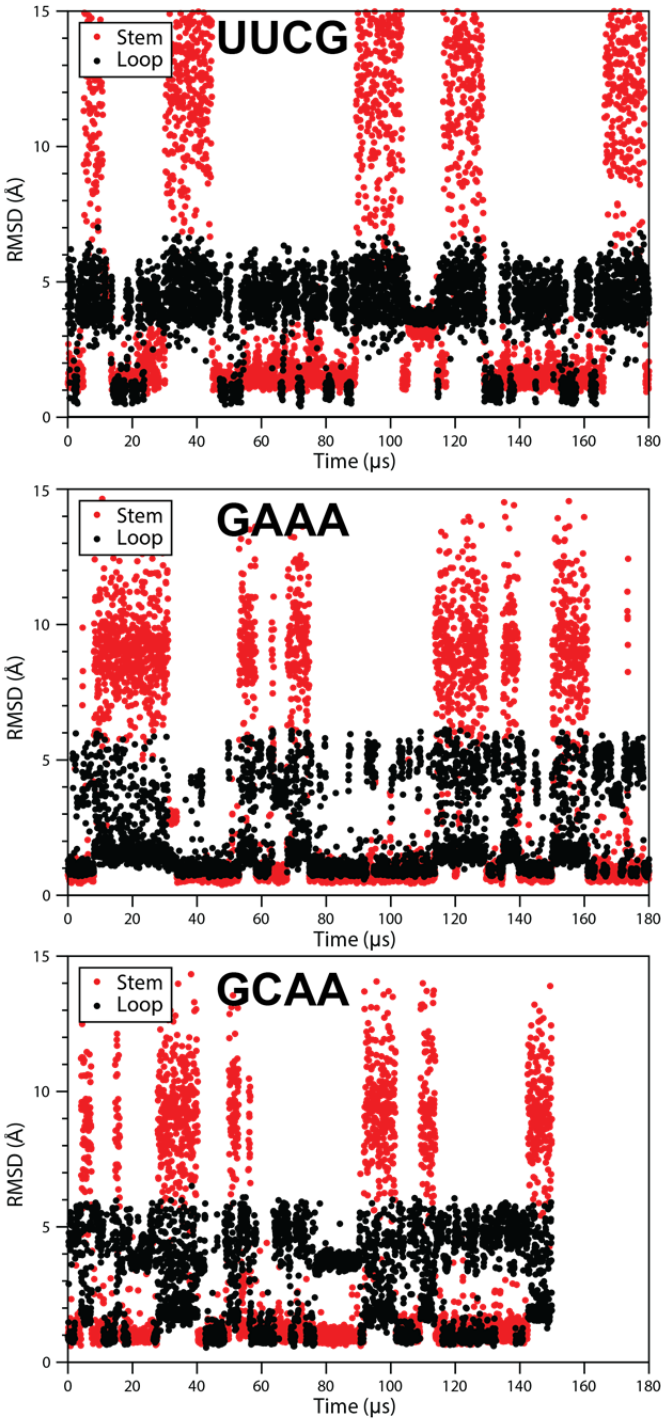


Figure S11. Simulated tempering simulations of three tetraloops using the revised force field. Plots of the heavy-atom RMSDs for the stem (red) and loop (black) regions are shown.

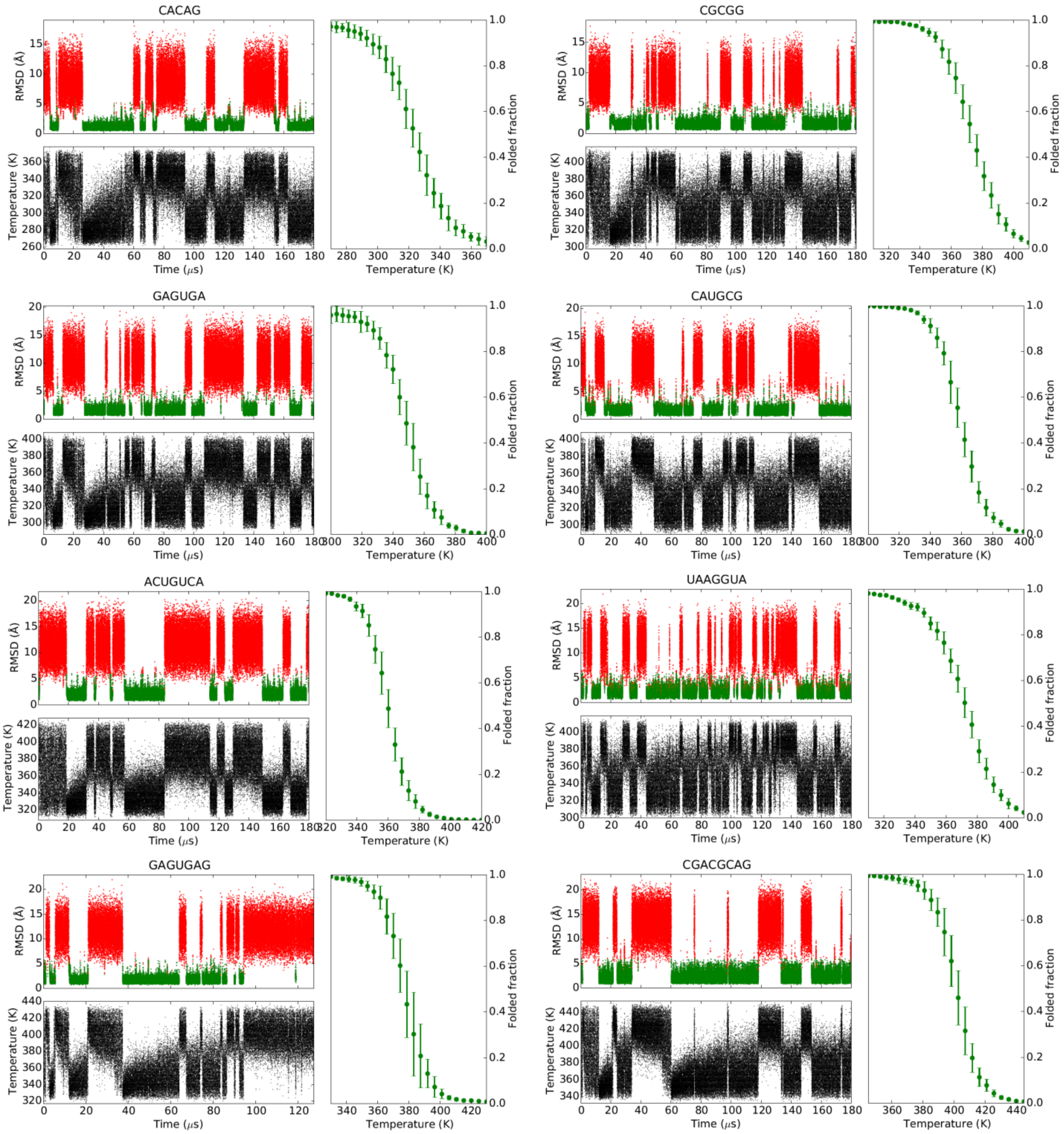


Figure S12. The non-hydrogen RMSD and temperature traces, as well as melting curves of the RNA duplexes.

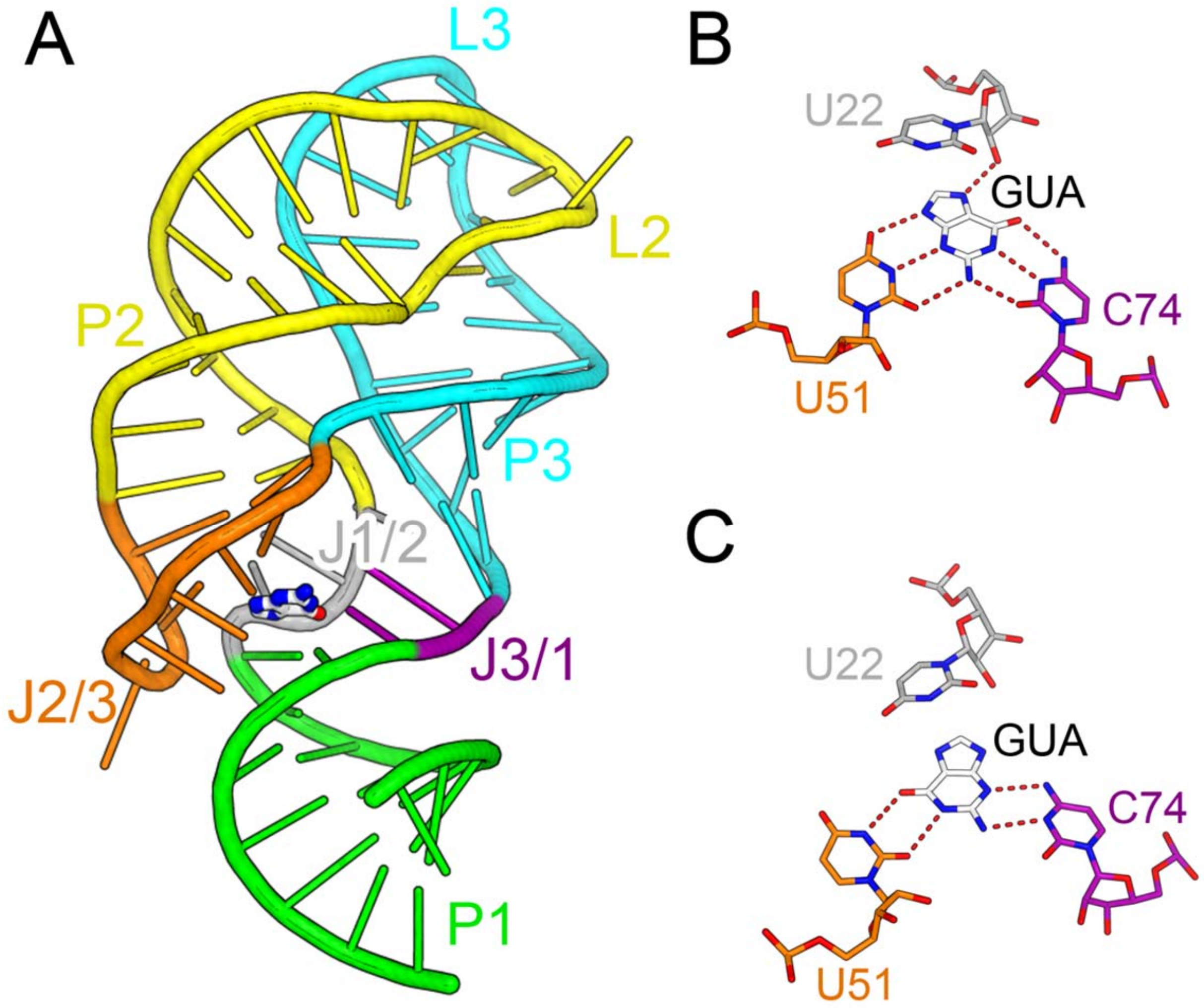


Figure S13. (A) The crystal structure of guanine-bound guanine riboswitch aptamer domain (PDB entry 4FE5). (B) The native guanine binding pose in the crystal structure. Hydrogen bonds are indicated as red dashes. (C) The non-native guanine binding pose observed in the simulation.

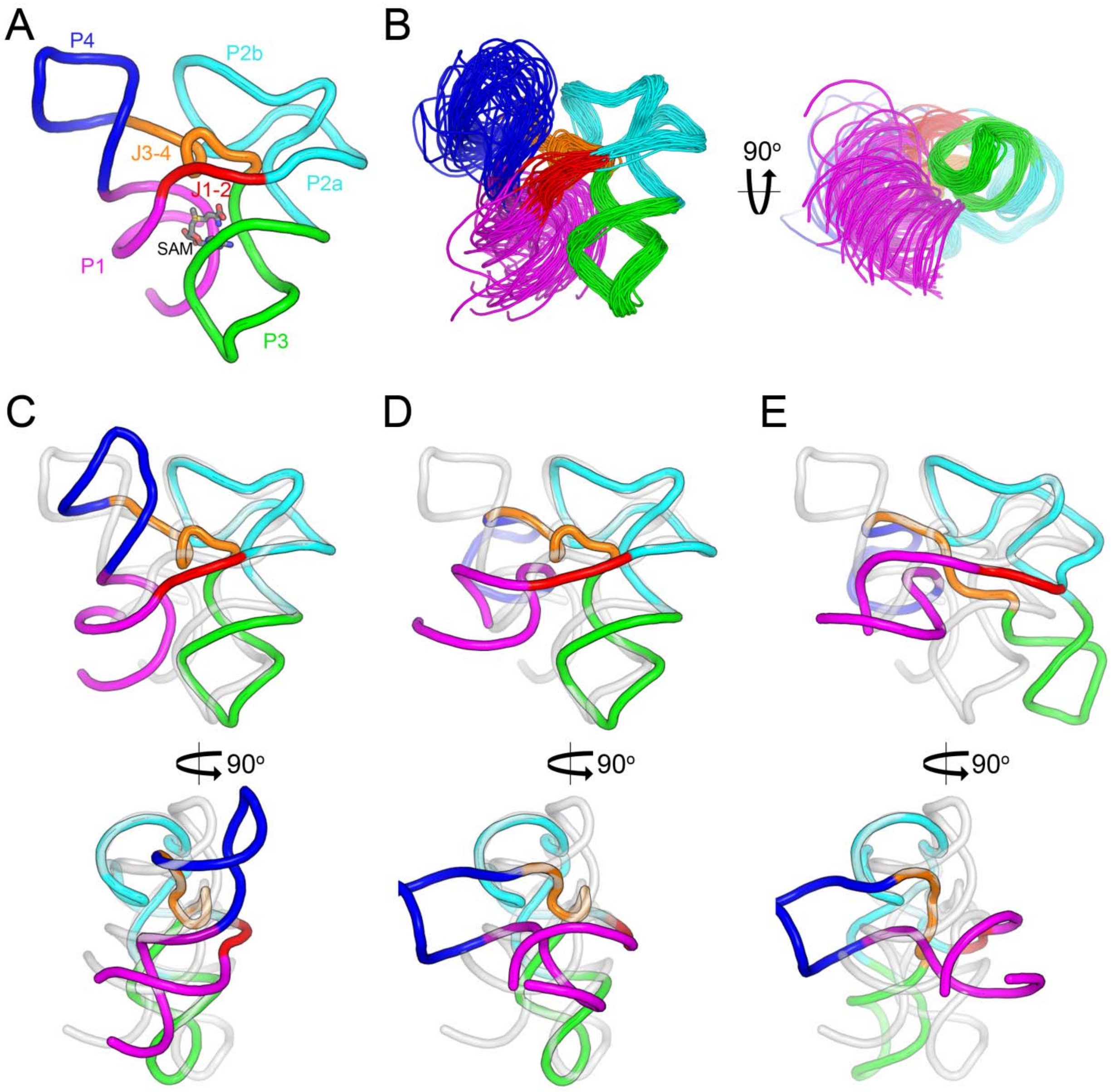


Figure S14. (A) The crystal structure of SAM-bound SAM-I riboswitch (PDB entry 3IQR). (B) Two views of the overlay of 40 frames from the Mg-C ensemble. The P2-P3-J3/4 core is the reference for the structural alignment. (C–E) Two views of the overlay between the average structures of the Mg-C (C), K-C (D), and K-E (E) ensembles and the crystal structure (in gray).

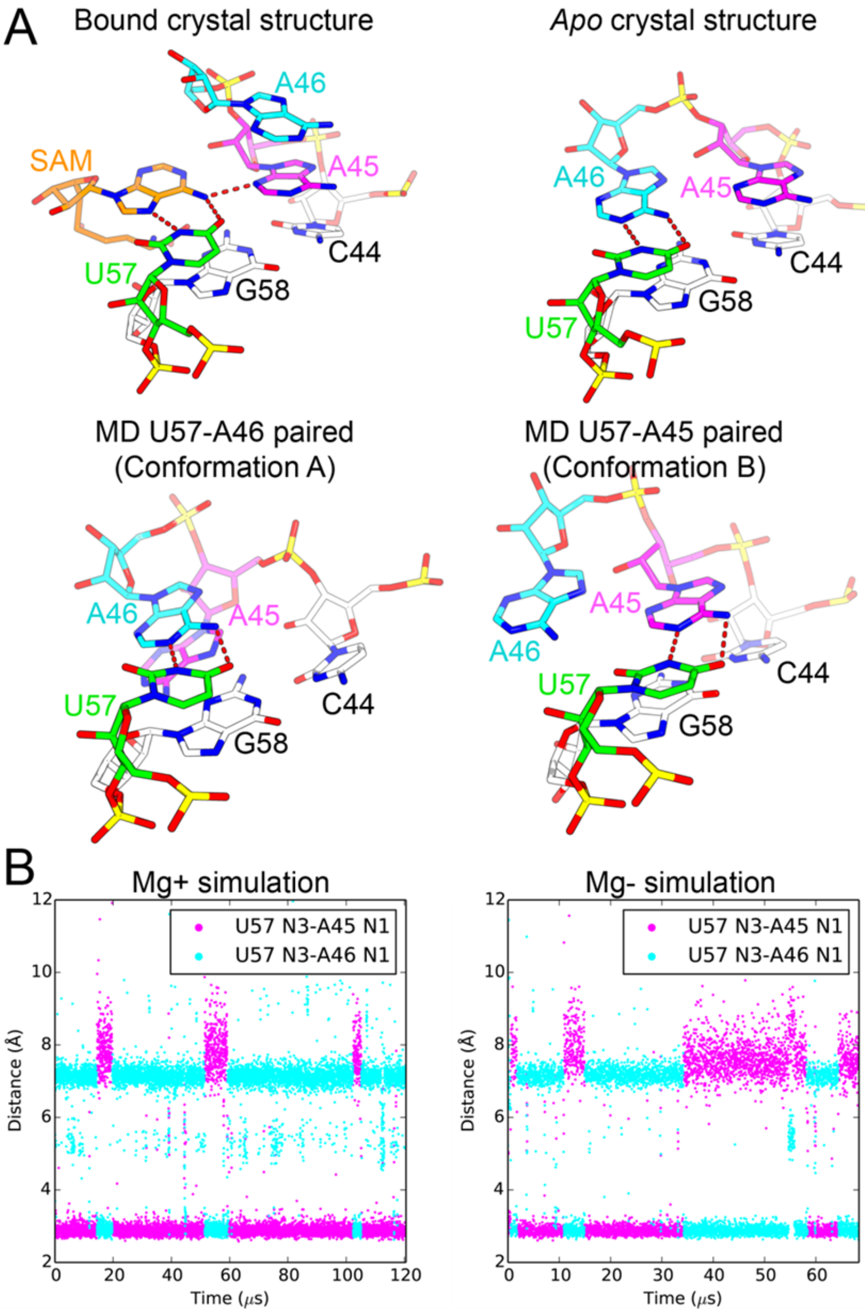


Figure S15. (A) Different experimental and simulated conformations of the SAM-binding pocket. (B) The distances between the N3 atom of U57 and the N1 atom of A45 (magenta) or A46 (cyan) in the simulations.

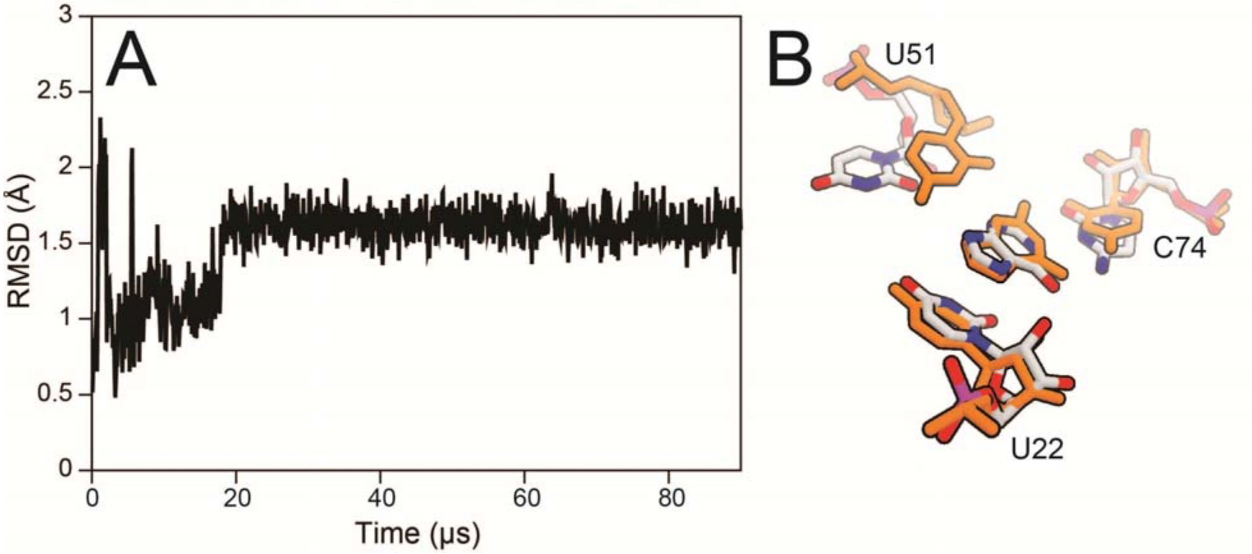


Figure S16. Simulation of the guanine riboswitch performed with the Amber ff14SB force field. (A) RMSD of the binding pocket (guanine and residues 22, 51, and 74) as a function of time. (B) Comparison of the initial (orange) and final structures of the binding pocket.

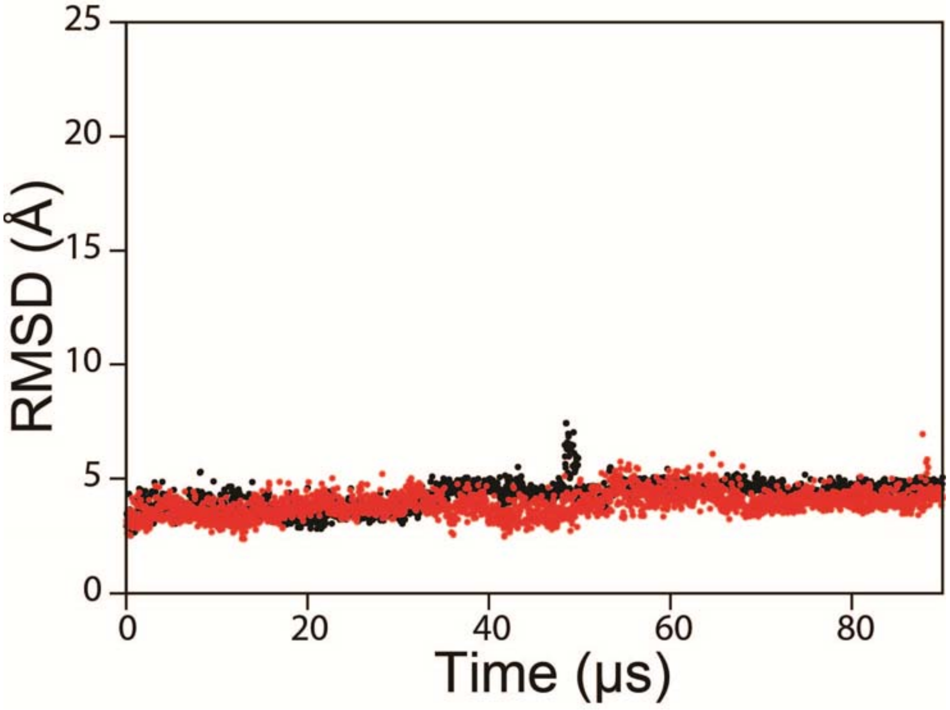


Figure S17. Simulation of the apo-SAM riboswitch performed with the Amber ff14SB force field. Heavy-atom RMSD as a function of time in simulations performed with (black) and without (red) 0.1 M $MgCl_2$.

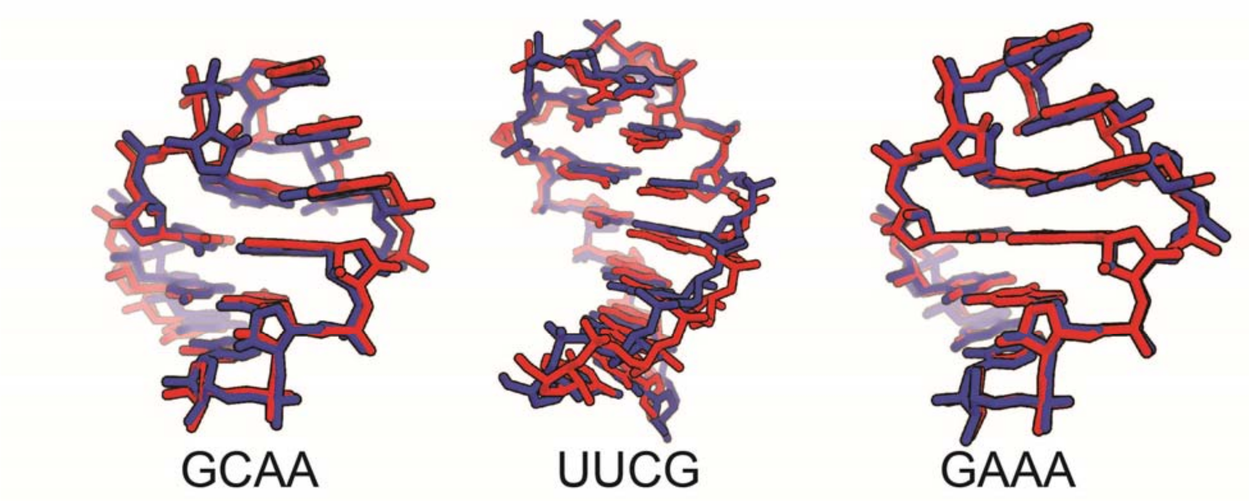


Figure S18. Comparison between the experimental structure (blue) and the average structure of the most populated cluster in the lowest temperature rung (red) of the three simulated tempering simulations of reversible folding of RNA tetraloops.

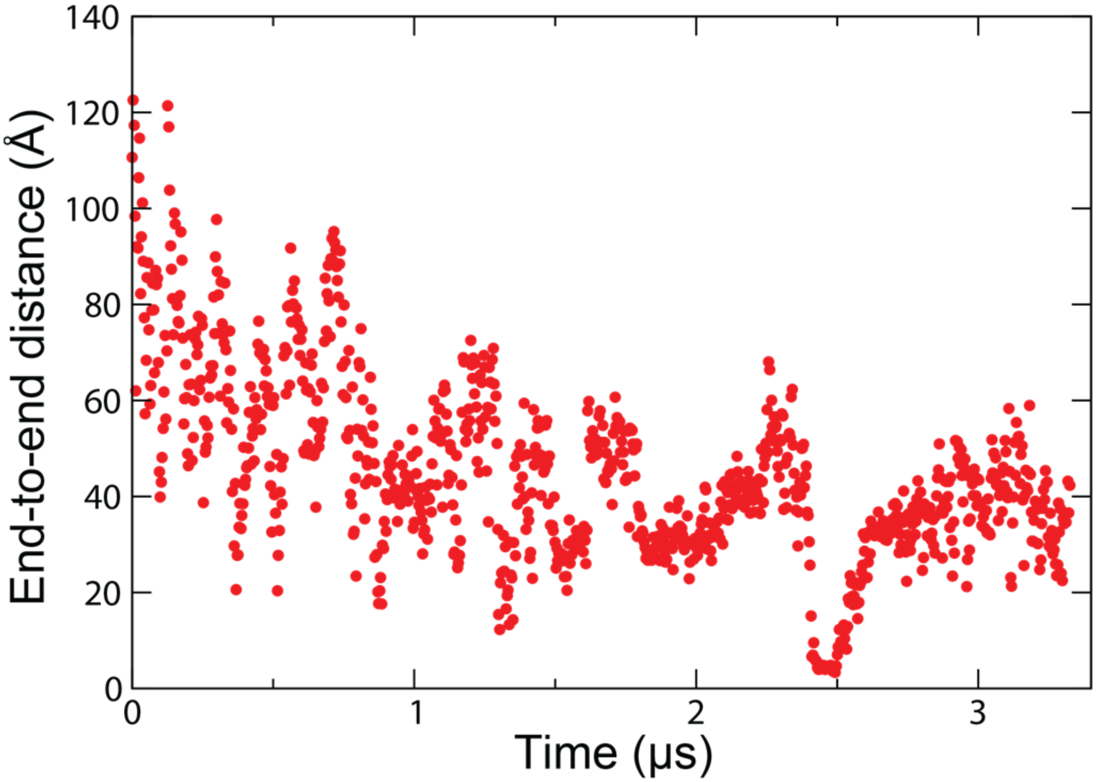


Figure S19. The end-to-end distance of the rU_{40} long ssRNA simulated with Amber ff14 and TIP4P-D.

References

1. Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT (2010) Free state conformational sampling of the SAM-I riboswitch aptamer domain. *Structure* 18(7):787–797.

2. Montange RK, Batey RT (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* 441(7097):1172–1175.

3. Izadi S, Anandakrishnan R, Onufriev AV (2014) Building water models: A different approach. *J Phys Chem Lett* 5(21):3863–3871.

4. Condon DE, Kennedy SD, Mort BC, Kierzek R, Yildirim I, Turner DH (2015) Stacking in RNA: NMR of four tetramers benchmark molecular dynamics. *J Chem Theory Comput* 11(6):2729–2742.

5. Tubbs JD, Condon DE, Kennedy SD, Hauser M, Bevilacqua PC, Turner DH (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(6):996–1010.

6. Nozinovic S, Fürtig B, Jonker HRA, Richter C, Schwalbe H (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res* 38(2):683–694.

7. Sheehy JP, Davis AR, Znosko BM (2010) Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* 16(2):417–429.

8. Lu X-J, Bussemaker HJ, Olson WK (2015) DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res*:gkv716.