THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Refining the RNA Force Field with Small-Angle X-ray Scattering of Helix–Junction–Helix RNA

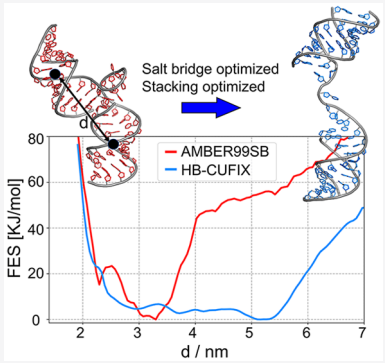
Weiwei He, Nawavi Naleem, Diego Kleiman, and Serdal Kirmizialtin*

Cite This: *J. Phys. Chem. Lett.* 2022, 13, 3400–3408

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The growing recognition of the functional and therapeutic roles played by RNA and the difficulties in gaining atomic-level insights by experiments are paving the way for all-atom simulations of RNA. One of the main impediments to the use of all-atom simulations is the imbalance between the energy terms of the RNA force fields. Through exhaustive sampling of an RNA helix–junction–helix (HJH) model using enhanced sampling, we critically assessed the select Amber force fields against small-angle X-ray scattering (SAXS) experiments. The tested AMBER99SB, DES-AMBER, and CUFIX force fields show deviations from measured profiles. First, we identified parameters leading to inconsistencies. Then, as a way to balance the forces governing RNA folding, we adopted strategies to refine hydrogen bonding, backbone, and base-stacking parameters. We validated the modified force field (HB-CUFIX) against SAXS data of the HJH model in different ionic strengths. Moreover, we tested a set of independent RNA systems to cross-validate the force field. Overall, HB-CUFIX demonstrates improved performance in studying thermodynamics and structural properties of realistic RNA motifs.



Our understanding of RNA's role in cellular processes and curing diseases continues to grow, but it is desirable that this understanding be down to the atomic level. Because of the limitations in reaching atomic-level details by experiments, molecular dynamics (MD) simulations are especially attractive in studying RNA. However, the accuracy of MD to describe RNA and its interactions is limited by the theoretical models that represent the potential energy landscape of RNA atoms, called force fields. Among the empirical force fields, polarizable force fields[1,2] hold promise to accurately describe the underlying physics of the RNA's characteristic highly charged backbone with the added cost of computational expense. However, major efforts have been concentrated on the nonpolarizable models due to their computational efficiency. As a result, the point-charge models of CHARMM[3] and AMBER[4] have been developed and utilized over the past decade to describe the conformations of RNA.

As a criterion to assess the accuracy of force fields, some efforts utilize the structural stability of RNA motifs of noncanonical structures.[5−8] Others use short oligonucleotides where an exhaustive conformational search is accessible, and direct comparison with experimental data, which is NMR in this case, is available.[9−11] On the basis of these studies, a modestly improved description of nucleic acid structures has been obtained by refining torsional parameters. Recent efforts extend this approach to balance the van der Waals interactions.[12,13] A similar approach is used to refine the imbalance in Amber ff14 (DES-AMBER).[14] Together with the TIP4P-D water model,[15] the description of RNA showed improvements.[14] Recently, an interaction-specific refinement protocol has been employed to better account for h-bonding.[11] An independent line of refinement effort concerns the balance of the intramolecular interactions of nucleic acid motifs. Yoo et al. employed the NBFIX[16] strategy to refine the nonbonded interactions of nucleic acid backbone and cations.[17] This effort led to the description of inter-DNA forces that is consistent with experiments.[18−20] A detailed description of the approach can be found elsewhere.[21]

CUFIX has been extensively tested for proteins and DNA.[21] The similarity of DNA to RNA suggests the transferability of the CUFIX parameters for RNA; however, the critical assessment of CUFIX corrections to RNA still remains elusive. In addition, the validity of the refinements derived by short oligonucleotides demands studies against more realistic RNA systems that have secondary and tertiary interactions. The need for an exhaustive sampling of the folded and unfolded ensembles of such systems poses a challenge for current computational methods.

RNA global conformation is largely defined by the orientation of A-form helices topologically constrained by linkers. Among the known RNA structures, about 70% are linked by two-way junctions,[22,23] making the helix–junction–

Received: February 6, 2022
Accepted: March 11, 2022
Published: April 11, 2022

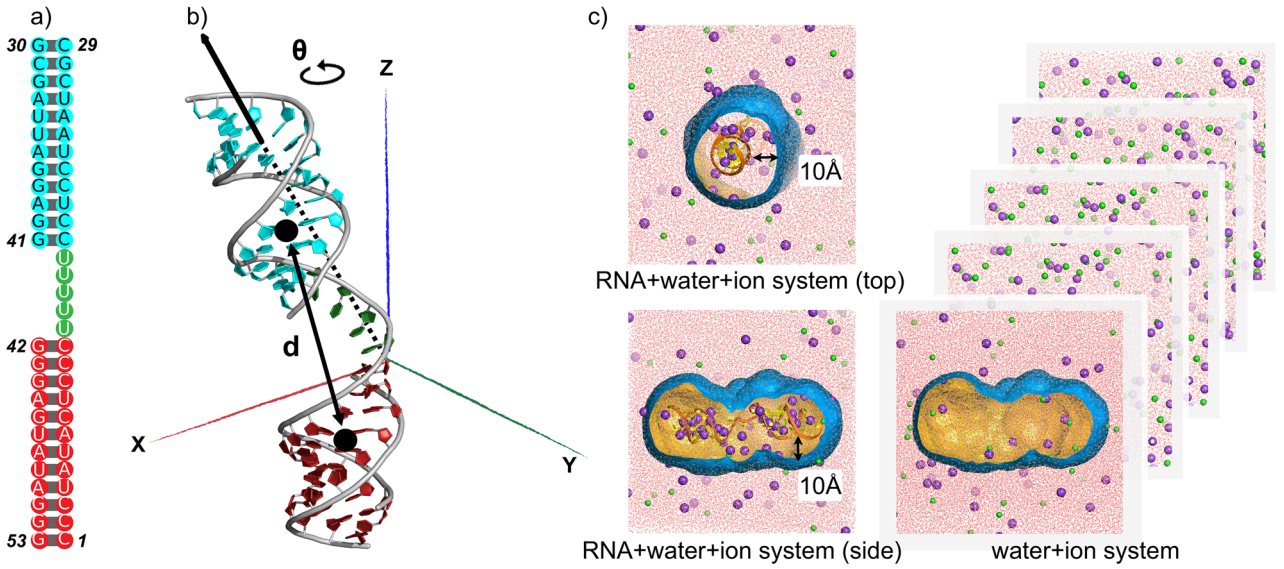


Figure 1. Helix–junction–helix (HJH) model used for force field optimization. (a) Sequence of the construct. (b) Collective variables (CV) used in well-tempered metadynamics simulations. Helix–helix distance, d, and customized azimuthal rotation, θ, serve to monitor the conformational ensemble. (c) Molecular setups used to compute small-angle X-ray scattering (SAXS) from simulations. The spatial envelope with a distance of 10 Å from the HJH surface serves to estimate the electron density of the bulk solvent and the excluded volumes (see section Computing Small-Angle X-ray Scattering Profiles of HJH from MD in Supporting Information for details).

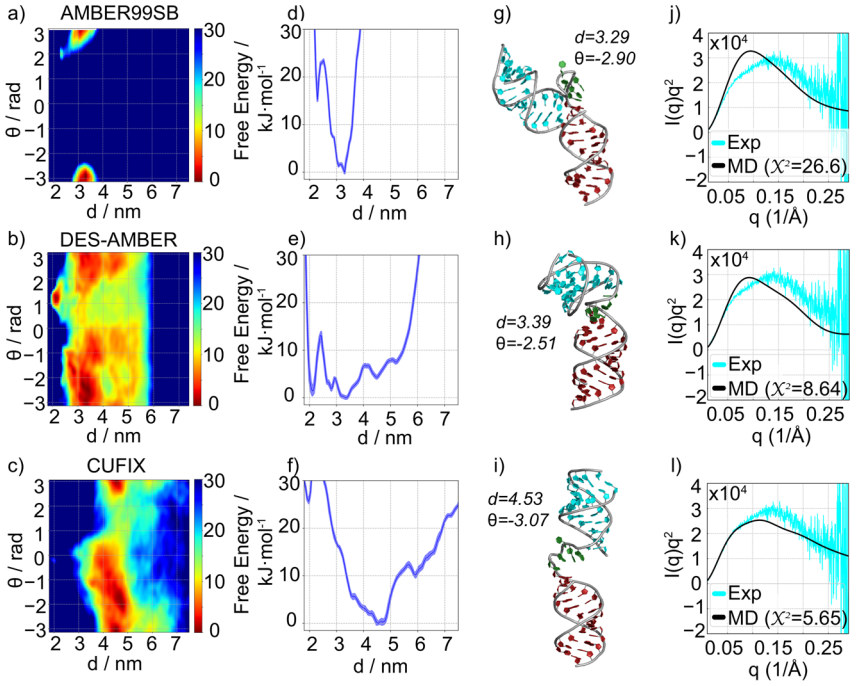


Figure 2. Comparison of the conformational ensembles sampled by well-tempered metadynamics simulations for AMBER99SB, DES-AMBER, and CUFIX force fields. The free energy surface (FES) projected onto the two collective variables (d, θ) described in Figure 1. (a–c) Two-dimensional projection of the (FES). (d–f) 1D projection of the FES along with the interhelix distance, d. (g–i) The representative conformations from the lowest energy states. (j–l) Comparison of experimental small-angle X-ray scattering (SAXS) data (100 mM KCl,[35] cyan) with computed SAXS from simulations. The similarity between experiment and simulation measured by χ^2 (see eq 1 in Supporting Information for details) where a smaller value represents better agreement.

helix (HJH) a ubiquitous RNA motif.[24−27] HJH is comprised of rigid and flexible parts where long-range electrostatic forces,[28] short-range stacking, volume exclusion, and hydrophobic forces[29] contribute to the overall conformations, making it an ideal model to benchmark the RNA empirical potentials.

The sequence and the structure of the HJH studied here are described in Figure 1a,b. Here, an explicit treatment of RNA, water, and ions allows a direct comparison with experiments (details are in the section Helix Junction Helix Simulations in Supporting Information). As an experimental benchmark, we used small-angle X-ray scattering (SAXS) as it serves as an ideal method to study the short- and long-range distance

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408

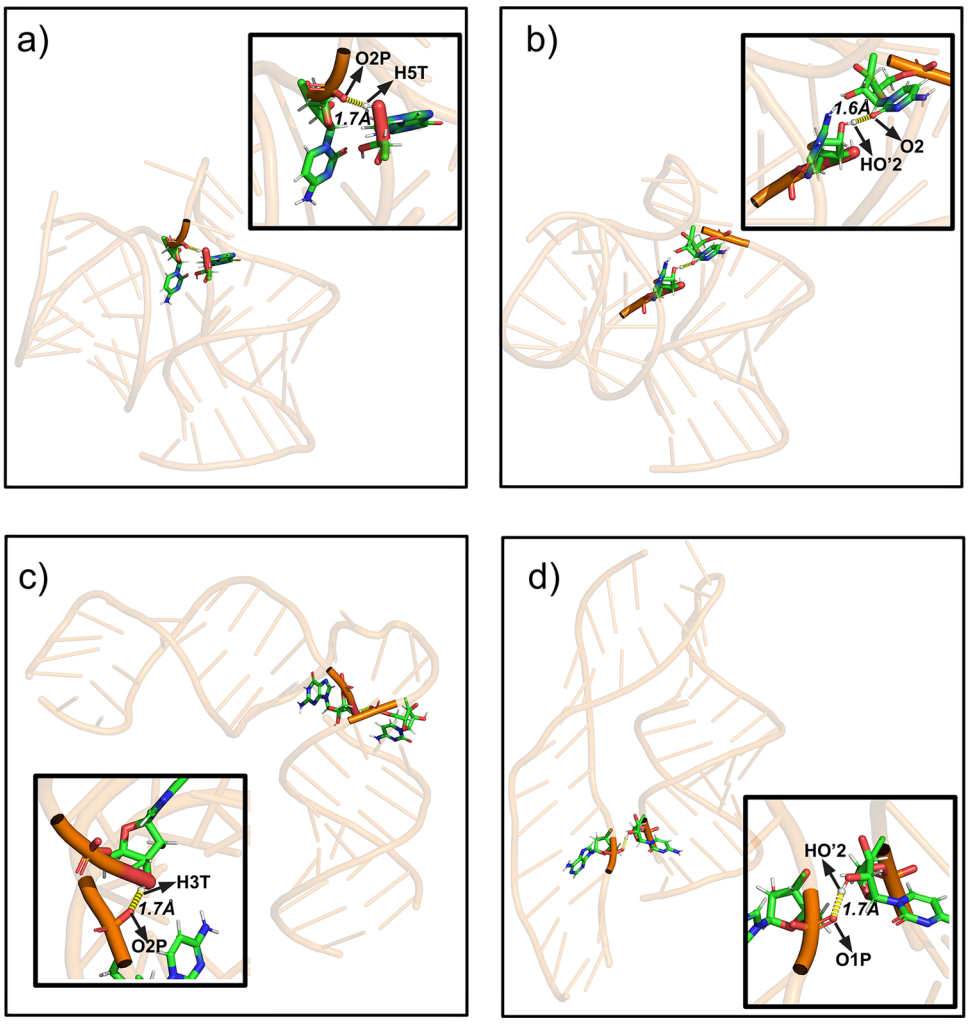


Figure 3. Demonstration of the hyperstable hydrogen bonding leading to collapsed conformation of HJH. (a–b) Conformations in AMBER99SB that lead to long-lived collapsed RNA conformations. (c–d) Conformations leading to collapsed states in CUFIX where the hydrogen bonding is observed between terminal and nonterminal hydrogen atoms (HO′2) and phosphate oxygens that lead to partial fraying. Insets in each figure are a zoom-in image detailing the atoms involved in the hydrogen bonding.

correlations and their ensembles measured at the physiological conditions.[30] We investigated the commonly used force fields from the AMBER family: amber99sb_parmbsc0 (denoted as AMBER99SB[31]), DES-AMBER,[14] and CUFIX.[18,21] Using well-tempered metadynamics (WTMD),[32−34] we exhaustively sampled the conformational space with collective variables (CV) inherent to HJH (namely, the interhelix distance, d, and azimuthal angle θ, as described in Figure 1b). The dynamics of the variables and the convergence of the potential energy surface are assessed for each system (Figure S1a−c). Later, the conformations sampled are used to compute the small-angle X-ray scattering (SAXS) profiles where scattering from RNA, explicit water, and ions are taken into account (see Supporting Information Methods, section Computing Small-Angle X-ray Scattering profiles of HJH from MD, for details) (Figure 1c). The computed SAXS profiles are weighted by the Boltzmann factor computed from the free energy mapped to CVs. The accurate computation of SAXS profiles and statistically converged ensembles of RNA conformations allowed providing a stringent test of the force fields against available experimental data.[35]

Surprisingly, the three force fields studied show notable differences in conformations sampled at identical conditions. The free energy surface (FES) projected onto the two collective variables (d, θ) reveals these differences (Figure 2a−f). AMBER99SB exhibited a state located at ($d \approx 3.29$ nm, $\theta \approx -2.90$ rad) that corresponds to a partially collapsed conformation (Figure 2g), whereas DES-AMBER showed a more diverse ensemble even though the energetically favorable state is still a collapsed form of HJH reminiscent of AMBER99SB. In contrast, the RNA adopts more open conformations and shows structural heterogeneity in the case of CUFIX (Figure 2c,i). To benchmark the findings, we computed the SAXS profile from each ensemble and displayed our results as Kratky plots ($I(q)q^2$ vs q) together with the experimental data at the same salt condition (Figure 2j−l).[35] The fitness of force fields is measured by the weighted sum of the squared error (χ^2) between the experimental and computed intensity, $I_{exp}(q)$ and $I_{comp}(q)$, where a high fitness has a low value of error (see Data Analysis for Helix Junction Helix in Supporting Information and eq 1 for details). On the basis of this, we observed that AMBER99SB gave rise to a poor ($\chi^2 \approx 26.6$) agreement with the experiment; DES-AMBER and CUFIX on the other hand showed marked improvements, $\chi^2 \approx$ 8.6, and 5.6, respectively. Despite the relatively good performance, all the aforementioned force fields failed to

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408

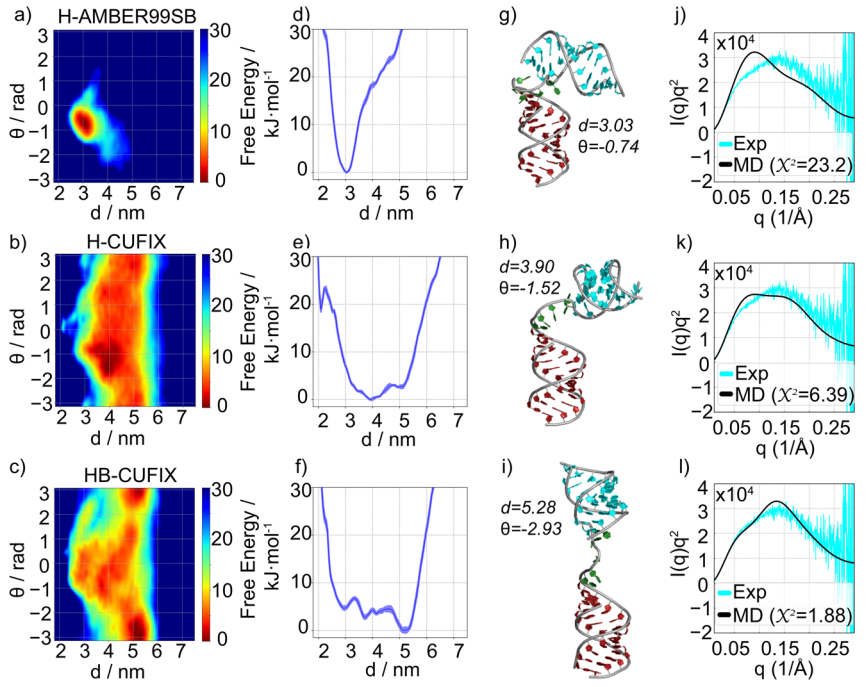


Figure 4. Comparison of the conformational ensembles sampled by well-tempered metadynamics simulations using the modified force fields H-AMBER99, H-CUFIX, and HB-CUFIX. (a−c) Free energy surface (FES) projected on the two CVs (d, θ). (d−f) 1D projection of the FES along with the helix-to-helix distance, d. (g−i) Representative conformations selected from the lowest energy states. (j−l) Comparison of experimental SAXS data (100 mM KCl, cyan) with computed SAXS profiles with the χ^2 of each force field reported.

capture the major peak located at 0.15, which corresponds to distance correlations around 4 nm (Figure 2j−l). The discrepancy of the SAXS profile between experiment and simulations suggests problems in describing HJH conformations with these force fields.

To modify the force fields to generate conformational ensembles consistent with the SAXS experiment of HJH, we inspect the trajectories generated by WTMD at atomic detail. Remarkably, whenever a collapsed conformation appeared, we observed the formation of hyperstable hydrogen-bond formation connecting adjacent helices (Figure 3, Figure S1e−f). The insets of Figure 3a−d capture a few instances. The donors of these specific interactions were mostly from the sugar hydroxyl group (−HO), which mainly lies on H5T/HO′2 atoms on the 5′ end and H3T/HO′2 atoms on the 3′ end, while the acceptors are among the phosphate oxygens of the backbone and sometimes from the negatively charged atoms forming the base. These contacts create long-lived kinetic traps of microseconds or longer evident in brute force MD. Unbiased simulations started from the collapsed state stayed at the collapsed state during the course of the simulation (>1.5 μs) (Figure S2). We note that Mlynsky et al. reports a similar problem for terminal nucleotides while studying RNA tetranucleotides (TNs).[11] In our study, in addition, we observe that strong hydrogen bonding extends to interhelix interactions. These contacts likely lock the RNA to collapsed states in the examined force fields, while in CUFIX in addition they resulted in the partial fraying of helices during WTMD simulations (Figure 3d).

To weaken the hydrogen-bonding strength, we introduce a finite size to the atom type −HO in AMBER, which is traditionally assigned a zero σ and ϵ. The correction is denoted as H-AMBER99SB and H-CUFIX, respectively. With this change, the frequent hydrogen bonding observed between the helices is suppressed (Figure S1d,f vs Figure S3d,e); however, H-AMBER99SB still displays compact forms giving rise to a poor agreement (Figure 4a,d,g,j). The H-CUFIX, on the other hand, results in expanding the accessible conformational space with a smoother FES where the conformational preferences get more delocalized (Figure 4b,e,h). Surprisingly, the hydrogen atom correction resolved the fraying issue of CUFIX. As a result of the −HO correction, however, the error in CUFIX is only slightly reduced ($\chi^2 \approx 6$) (Figure 4k), suggesting the need for further refinements.

We focus on the H-CUFIX due to its relative success in capturing the SAXS profile of HJH. The −HO correction resulted in shifting the governing force dictating the HJH conformation from h-bonding interactions to stacking forces rendering high orientational correlations (Figure S5). Typical conformations from each force field demonstrate these stacking tendencies in Figure S5c. To balance the base stacking, we adopt a refinement strategy proposed for AMBER99SB by Chen et al.[13] This correction is added to the CUFIX force field in addition to our modifications. Table 1 summarizes all the parameters modified in the new force field that we call HB-CUFIX henceforth.

HB-CUFIX allowed higher structural diversity (Figure 4c). In addition, the location of the energy minimum shifted to more extended conformations (d = 5.28 nm, θ = −2.93 rad). Surprisingly, the balanced forces between stacking and long-range electrostatics resulted in better agreement with the experiment (Figure 4l). The reported $\chi^2 \approx 1.88$ is the lowest among all force field trials.

The results reported here are when the HJH is in 100 mM KCl salt. To test the validity of the corrections over a wider range of ionic strength, we repeated our simulations at 50 and 200 mM where experimental data are available.[35] The convergence of the WTMD simulations was assessed (Figures

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408

Table 1. Optimized L−J Parameters[a]

atom type	σ (Å)	ϵ (kJ/mol)
HO	2.60000	0.656888
nucleobase carbon	3.22968	0.287859
nucleobase nitrogen	3.08750	0.569031
nucleobase oxygen	2.91192	0.702912
base C/TIP3P oxygen	3.27514	0.430672
base N/TIP3P oxygen	3.20030	0.605515
base O/TIP3P oxygen	3.05526	0.672989

[a]The details of stacking parameters as well as glycosidic torsion parameters are described in detail in Chen et al.[13]

S3−S4), and the Kratky plots comparing the two methods are shown in Figure S6. Remarkably, the χ^2 still remains low, 1.69 for 50 mM and 1.47 for 200 mM, respectively, suggesting the accuracy of the force field to a wider range of monovalent solvent conditions.

In addition to monovalent salt conditions, we tested the performance of HB-CUFIX against a divalent ion. Because of its importance for RNA and its abundance in cellular media, we studied Mg^{2+} ions. The SAXS experiment for HJH in the presence of Mg^{2+} ions is not available. To have a direct comparison, we compared our simulation results against Förster resonance energy transfer (FRET) experiments of the HJH conducted in the presence of $MgCl_2$.[36]

The WTMD methodology allowed constructing the energy landscape of HJH in the presence of Mg^{2+} ions. First, the convergence of the simulations was assessed (Figure S7). On the basis of the FES, we computed the average FRET efficiency (E_{FRET}) by considering the excluded volume and rotations of the fluorophore pairs explicitly using ref 37 (Figure 5e). We also report distance and (E_{FRET}) distributions to contrast each force field (Figure S8). Details of FRET computation can be found in the Computing FRET efficiency from MD simulations section of Supporting Information.

We show in Figure 5a−d the energy landscape for each force field under study. The insets show the lowest energy conformations both, suggesting major differences in the structures sampled. The FES for both AMBER99SB and DES-AMBER show collapsed states as the most favorable conformations. The CUFIX and HB-CUFIX, on the other hand, adopt more heterogeneous conformational ensembles (Figure 5c−d).

Among the force fields under study, we found that AMBER99SB and HB-CUFIX showed the best agreement with the experiment (Figure 5f). DES-AMBER exhibited a lower FRET value because the HJH conformations sampled by DES-AMBER resulted in a conformational state where the two labeling sites were positioned in opposite directions, leading to an increased dye−dye distance (see insets in Figure S8a,b). The CUFIX, on the other hand, showed fraying issues during WTMD, leading to a minimum at $d \approx 1.1$ nm (Figure 5c). Excluding the conformations showing fraying, the $E_{FRET} \approx 0.41$ estimated by CUFIX also deviates from the experimental result ($E_{FRET} \approx 0.58$).

The refinement protocol that balanced the forces governing HJH RNA conformations gave rise to a force field that is consistent with available experimental data near physiological monovalent and divalent salt conditions. The success of the force field gives us the confidence that HB-CUFIX can be used to study RNA structures comprised of rigid and flexible parts.

To test if HB-CUFIX is also successful in describing structural and dynamical properties of other RNA motifs, we first looked at the duplex RNA parameters. The helices of the HJH are used for the benchmark. We compute structural parameters that define an RNA duplex using the 3DNA program.[38] As an experimental benchmark, we identify duplexes of similar length derived by NMR experiments in salt conditions comparable to our study (~90−110 mM monovalent). The data are used to derive average parameters for an A-form RNA duplex. Table 2 provides the comparison of average duplex parameters from experiments and the force fields under study. Remarkably, the corrections introduced herein show no adverse effects in describing the structural parameters of an A-form duplex. Rather, the refinement improved the description of most of the parameters. Note that this comparison is rather indirect, as changes in duplex parameters likely show sequence dependence. Nevertheless,

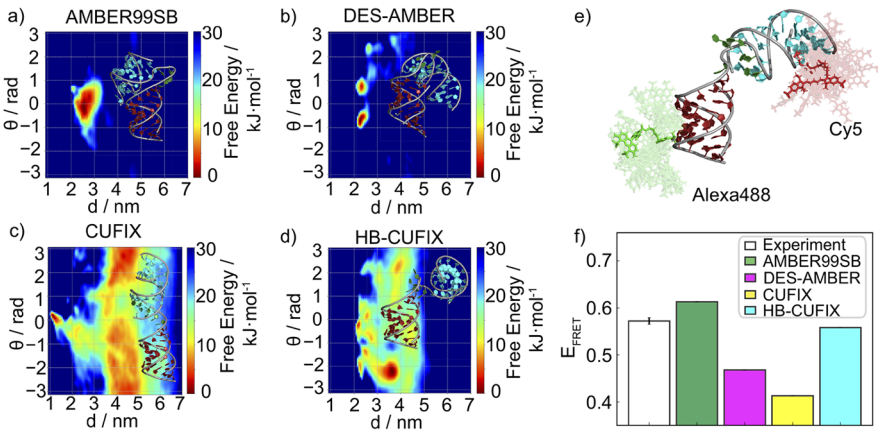


Figure 5. Comparison of the conformational ensembles sampled by well-tempered metadynamics simulations in the presence of Mg^{2+} for AMBER99SB, DES-AMBER, CUFIX, and HB-CUFIX force fields. (a−d) The free energy surface (FES) projected onto the two collective variables (d, θ). The insets provide representative conformations from the lowest energy states of each ensemble. The fraying conformations of CUFIX are not considered. (e−f) Comparison of experimental[36] and computational fluorescence resonance energy transfer efficiency (E_{FRET}) for each force field. Dye dynamics computed by ref 37. (Details of the computational approach can be found in the Computing FRET Efficiency from MD Simulations section of Supporting Information and Figures S8 and S9.)

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408

Table 2. Structural Parameters of RNA Duplexes from Experiments and Simulations[a]

	NMR	AMBER99SB	DES-AMBER	CUFIX	HB-CUFIX
twist [°]	32.0 ± 3.1	29.8 ± 0.8	29.1 ± 1.0	29.9 ± 0.9	30.5 ± 0.8
roll [°]	8.5 ± 0.7	9.8 ± 2.0	10.1 ± 1.4	10.3 ± 1.9	9.6 ± 1.5
slide [Å]	− 1.29 ± 0.50	− 1.48 ± 0.16	− 1.77 ± 0.18	− 1.40 ± 0.20	− 1.84 ± 0.15
inclination [°]	15.0 ± 2.2	17.8 ± 3.5	18.8 ± 2.7	18.7 ± 3.4	17.0 ± 2.7
h-rise [Å]	2.81 ± 0.14	2.61 ± 0.15	2.56 ± 0.15	2.66 ± 0.15	2.50 ± 0.13
propeller [°]	− 10.1 ± 6.3	− 12.8 ± 2.9	− 12.1 ± 2.4	− 14.6 ± 3.2	− 12.5 ± 2.8
pucker [°]	38.8 ± 2.0	38.3 ± 1.0	39.4 ± 1.0	38.1 ± 1.2	40.3 ± 1.0
χ [°]	− 157.5 ± 4.9	− 153.3 ± 6.9	− 156.6 ± 6.3	− 151.8 ± 7.1	− 153.6 ± 10.4
α [°]	− 68.2 ± 9.8	− 75.3 ± 3.8	− 74.1 ± 3.8	− 74.9 ± 4.3	− 75.5 ± 3.6
β [°]	112.7 ± 56.4	91.2 ± 26.3	100.7 ± 26.0	90.0 ± 25.1	92.5 ± 25.6
γ [°]	53.8 ± 12.5	60.2 ± 5.8	60.2 ± 5.3	63.0 ± 4.0	62.5 ± 3.7
δ [°]	81.9 ± 8.2	79.9 ± 2.1	76.2 ± 1.5	81.1 ± 2.3	77.3 ± 1.9
ϵ [°]	− 157.4 ± 3.3	− 155.0 ± 7.1	− 155.8 ± 5.9	− 153.7 ± 10.4	− 154.2 ± 4.7
ζ [°]	− 68.8 ± 4.9	− 68.4 ± 1.7	− 66.2 ± 1.5	− 68.1 ± 1.7	− 67.7 ± 1.8
P···P [Å]	10.0 ± 1.5	12.1 ± 0.7	13.3 ± 1.1	12.0 ± 0.9	12.6 ± 0.8

[a]Average helix parameters derived from NMR experiments compared with simulation results. To determine average duplex geometry, the parameters of the duplex sequences of similar salt conditions to ours (PDB ids: 1F5G,[46] 2L8F,[47] 2MVY,[48] 2MS5[49]) were selected. The simulation data were calculated from the duplex part of HJH. To obtain the average duplex parameters, 2000 conformations from the lowest energy basin of each force field were selected. For CUFIX, the fraying conformations were excluded.

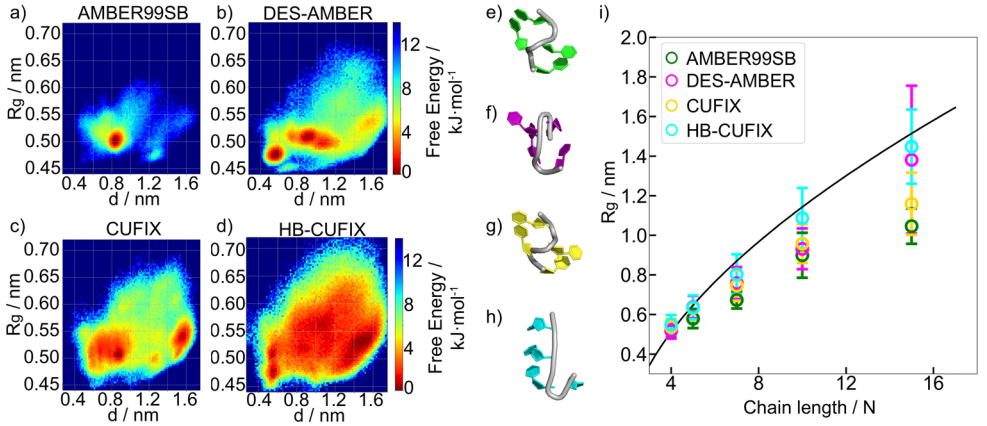


Figure 6. (a−d) The comparison of force fields in representing the free energy surface of the uracil tetraloop projected on the radius of gyration, R_g and end-to-end distance, R. The minimum energy conformations for the free energy plot were extracted at each force field and clustered to find the representative conformation shown for (e) AMBER99SB, (f) DES-AMBER, (g) CUFIX, and (h) HB-CUFIX. The ^3J coupling and stacking comparison with experiments for these conformations are shown in the supporting materials Table S5. (i) The change in the R_g of uracil oligomers as a function of chain length, N. AMBER99SB (green), DES-AMBER (magenta), CUFIX (yellow), and HB-CUFIX (cyan). The black solid line is the theoretical relationship $R_g = A_0(N - 1)^\nu$ fitted to experiments[44] with $A_0 = 0.342$ nm and $\nu = 0.58$. Error bars in the simulation results are estimated by dividing the time series data into three equal lengths.

this exercise demonstrates that the modifications in HB-CUFIX retain the A-form duplex geometry.

As a next test, we investigated the performance of the force fields with single-stranded RNAs. The rU_4 has been a challenging sequence for simulation studies,[14,39,40] thus serving as an excellent system for our study. Three microseconds of brute force MD simulations were used to sample conformations. To elucidate the conformational differences, we projected the free energy surface (FES) onto the radius of gyration, R_g, and end-to-end distance, R, as shown in Figure 6a−d. The time evolution of R_g is shown in Figure S10 for comparison. A representative structure from the lowest energy state of each force field details structural differences (Figure 6). To obtain the representative conformation, we extracted the minimum energy conformations based on FES and clustered them with a cutoff value of 0.2 nm using the gromos method.[41] The center conformation of the most populated cluster served as the representative structure and is shown in Figure 6e−h. From the FES, a dominant conformation representing the collapsed states ($R_g \approx 0.5$nm, $R \approx 0.8$ nm) is observed by all force fields. In addition, DES-AMBER, CUFIX, and HB-CUFIX show an additional prevalent extended state ensemble with a higher conformational heterogeneity during the time scale of the simulation.

To benchmark our findings with experiments, we computed ^3J scalar couplings from the data. The force fields are tested against available NMR experiments[39] (see Data Analysis of ssRNA in Supporting Information for details). The root mean squared error between simulation and experiment was used to quantify the performances, where a lower χ value suggests a better agreement between experiment and simulation. Table 3 summarizes the results of ^3J couplings for each force field. Overall, the HB-CUFIX performs well in capturing the couplings constants better than CUFIX and AMBER99SB, and DES-AMBER shows the best performance. The cumulative χ error of HB-CUFIX is similar to AMBER99SB.

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408

Table 3. Comparison of Experiment and Simulations in ^3J Coupling Parameters for rU_4[a]

	residue	AMBER99SB	DES-AMBER	CUFIX	HB-CUFIX
H5′ − P	2	2.24	1.72	1.68	1.87
(β)	3	1.81	2.30	1.76	1.96
	4	1.46	1.78	1.38	2.06
H5″ − P	2	1.30	1.41	1.76	1.20
(β)	3	2.95	1.79	1.54	1.43
	4	1.61	1.61	1.50	1.42
H4′ − H5′	1	1.77	3.08	1.85	2.83
(γ)	2	1.91	2.21	1.18	2.04
	3	2.26	3.04	1.06	1.97
H4′ − H5″	1	1.25	1.28	2.55	1.31
(γ)	2	1.03	0.89	1.48	1.00
	3	0.94	1.14	1.34	1.23
H3′ − P	1	1.94	2.46	1.72	1.96
(ϵ)	2	1.76	2.35	1.58	1.85
	3	1.47	2.16	1.59	1.92
H1′ − H2′	1	4.01	3.13	3.48	2.81
(ν1)	2	3.52	3.15	3.47	3.11
	3	3.48	2.94	3.40	2.79
	4	3.84	4.01	4.23	3.76
H2′ − H3′	1	0.81	0.84	0.82	0.79
(ν2)	2	0.81	0.79	0.73	0.75
	3	0.83	0.83	0.82	0.79
H3′ − H4′	1	2.79	3.25	3.08	3.53
(ν3)	2	3.08	3.19	3.01	3.39
	3	2.86	3.22	2.94	3.48
average backbone		1.71	1.95	1.60	1.74
average ribose ring		2.60	2.54	2.60	2.52
average total		2.07	2.18	2.00	2.05

[a]The difference between experiment and simulation is measured by χ described in Supporting Information (eq 5). The ^3J coupling data for experiments were taken from refs 14 and 39, and an experimental error of 1.5 Hz was employed as in refs 39 and 50.

Table 4. Comparison of Experiment and Simulations of Base Stacking and Intercalation for rU_4 Full Trajectory[a]

	NMR	AMBER99SB (%)	DES-AMBER (%)	CUFIX (%)	HB-CUFIX (%)
1−2 stack	none	5.8	15.0	19.2	32.3
2−3 stack	none	3.9	16.6	18.2	26.7
3−4 stack	none	3.9	9.4	11.3	11.8
1−3 stack	none	50.9	32.7	28.6	15.6
1−4 stack	none	2.9	16.1	3.5	24.8
2−4 stack	none	32.6	10.0	19.3	10.9
1−3−2 intercalation	none	13.1	7.1	1.6	0.2
3−1−4 intercalation	none	8.4	6.3	1.2	0.4

[a]Comparison of structural parameters of rU_4. The nucleotides are numbered 1 to 4 starting with the 5′ end. The population of stacking and intercalation from the simulation is compared with derived parameters from NMR experiments.[39] To compute stacking, we used Barnaba[51] software, and for intercalation we monitored the shortest nonbonded distance between bases. The distances that do not change more than 0.5 Å within 1 ns were considered intercalation events.

As far as the average error is concerned, CUFIX showed slightly better performance than HB-CUFIX. We also benchmark ^3J from the most populated ensemble (see Table S4). To compare structural parameters, we computed base stacking and intercalation ratios (Table 4). Overall, HB-CUFIX demonstrates good performance in describing ^3J couplings and nonbonded interactions of rU_4 even though the data were not used in our training.

The difference between HB-CUFIX describing the overall size of rU_4 demands more investigation. The short sequence of rU_4 likely assesses the short-range (van der Waals) interactions that influence the excluded-volume; however, to account for a correct balance between short-range van der Waals and long-range electrostatics, longer RNA chains are needed. For that purpose, we studied different lengths of rU_N where $N = 4-15$, and we compared our results with available data.[42]

The excluded volume of single-strand RNA (ssRNA) can be successfully described by Flory's scaling law: $R_g = A_0(N-1)^{\nu}$ where N is the number of residues, ν is the scaling exponent, and A_0 is the scaling prefactor. On the basis of measurements, these parameters are determined for ssRNAs. Using parameters derived from SAXS experiments of oligo-Us[42,43] at 100 mM monovalent salt, we extrapolated the chain dimensions using Flory's law for the chain size under study. Shown in Figure 6l is the radius of gyration computed directly from simulation compared with the theoretical curve describing the experiment. HB-CUFIX and DES-AMBER show more expanded conformations with a better agreement with experiments, while AMBER99SB and CUFIX adopt more collapsed states that show increased deviation from the scaling law.

Lastly, for a charged polymer like ssRNA, the balance between short-range forces and surrounding salt concentration due to electrostatic screening can be measured using persistence length, L_P (details of calculations in Supporting Information, Data Analysis of ssRNA). To assess the force fields, we compare this property for oligo-Us. The experiment[44] reports $L_P \approx 1.9$ nm for U_{30} at 100 mM KCl. Computed L_P from rU_{15} on the other hand shows notable differences among force fields. HB-CUFIX provides the closest value to the experiment ($L_P^{\mathrm{HB-CUFIX}} \approx 1.31$ nm,) while DES-AMBER, AMBER99SB, and CUFIX show larger deviations ($L_P^{\mathrm{AMBER99SB}} \approx 0.45$ nm, $L_P^{\mathrm{DES-AMBER}} \approx 0.91$ nm, and $L_P^{\mathrm{CUFIX}} \approx 0.56$ nm).

To develop a balanced RNA force field, realistic RNA models with ample experimental data that are directly computable from simulations are necessary. Small-angle X-ray scattering (SAXS) that provides local and global information about the conformational ensemble of RNA provides benchmark systems for force field refinement. Helix−junction−helix (HJH) RNA that has flexible and rigid parts, long- and short-range forces possessing secondary and tertiary folding, and solvent effects serve as a minimal model to represent functional RNA and ideal models to benchmark force fields. Using the HJH RNA model, we observe that AMBER99SB, DES-AMBER, and CUFIX force fields do not capture the conformational ensembles consistent with SAXS data, likely due to the imbalances among the energy terms. Combining the orthogonal refinement strategies,[13,21] we developed a new force field that balances hydrogen bonding and base stacking (HB-CUFIX) that successfully describes RNA conformational ensembles consistent with solution

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408

studies. We tested our approach in monovalent and divalent salt conditions for HJH. The parameters that show excellent agreement against SAXS data also gave rise to a good agreement in describing the double- and single-stranded RNAs. Future research and refinement efforts will extend the study to the conformational ensemble of other RNA motifs and salt conditions. We expect that HB-CUFIX will enable the study of structural and thermodynamic properties of many functional RNA molecules in the future. We anticipate that using SAXS to test and refine force fields can be a viable strategy for other molecular systems as well.

In addition to the force field optimization, we also investigate the energy landscape of a two-way junction in this study. The conformational ensemble of HJH in monovalent salt is an open state where helices are separated due to electrostatic repulsion. In this condition, RNA is flexible and samples diverse structures dictated by its junction (Figure 4c–f). The presence of Mg cations on the other hand modifies the FES and leads to the formation of a well -defined tertiary structure (Figure 5d). These findings align well with current understanding[45] of RNA folding and highlight the unique role of Mg^{2+} ions in the folding and assembly of RNA structures.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.2c00359.

> The force field parameters, molecular simulation setups for the HJH RNAs under study, brute force MD and well-tempered metadynamics simulation details, details of computing small-angle X-ray and FRET from MD simulations, molecular simulation setups for single-strand RNA simulations, details of analysis of single-strand RNAs and computing NMR J-couplings from simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Serdal Kirmizialtin − *Chemistry Program, Science Division, New York University, Abu Dhabi, United Arab Emirates;* orcid.org/0000-0001-8380-5725; Email: serdal@nyu.edu

Authors

Weiwei He − *Chemistry Program, Science Division, New York University, Abu Dhabi, United Arab Emirates; Department of Chemistry, New York University, New York, New York 10003, United States*

Nawavi Naleem − *Chemistry Program, Science Division, New York University, Abu Dhabi, United Arab Emirates*

Diego Kleiman − *Chemistry Program, Science Division, New York University, Abu Dhabi, United Arab Emirates;* Present Address: Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.2c00359

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was carried out on the High Performance Computing resources at New York University Abu Dhabi and supported by an AD181 faculty research grant. The authors thank Lois Pollack and Yen-Lin Chen for valuable discussions and for kindly providing the SAXS data.

■ REFERENCES

(1) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *Journal of Chemical Theory and Computation* **2018**, *14*, 2084−2108.

(2) Lemkul, J. A.; MacKerell, A. D. Polarizable force field for RNA based on the classical drude oscillator. *J. Comput. Chem.* **2018**, *39*, 2624−2646.

(3) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.

(4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(5) Wheatley, E. G.; Pieniazek, S. N.; Mukerji, I.; Beveridge, D. Molecular Dynamics of a DNA Holliday Junction: The Inverted Repeat Sequence d(CCGGTACCGG)4. *Biophys. J.* **2012**, *102*, 552−560.

(6) Wolski, P.; Nieszporek, K.; Panczyk, T. G-Quadruplex and I-Motif Structures within the Telomeric DNA Duplex. A Molecular Dynamics Analysis of Protonation States as Factors Affecting Their Stability. *The Journal of Physical Chemistry B* **2019**, *123*, 468−479.

(7) Suddala, K. C.; Price, I. R.; Dandpat, S. S.; Janeček, M.; Kührová, P.; Šponer, J.; Banáš, P.; Ke, A.; Walter, N. G. Local-to-global signal transduction at the core of a Mn^{2+} sensing riboswitch. *Nat. Commun.* **2019**, *10*. DOI: 10.1038/s41467-019-12230-5

(8) Salsbury, A. M.; Dean, T. J.; Lemkul, J. A. Polarizable Molecular Dynamics Simulations of Two c-kit Oncogene Promoter G-Quadruplexes: Effect of Primary and Secondary Structure on Loop and Ion Sampling. *Journal of Chemical Theory and Computation* **2020**, *16*, 3430−3444.

(9) Condon, D. E.; Yildirim, I.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Turner, D. H. Optimization of an AMBER Force Field for the Artificial Nucleic Acid, LNA, and Benchmarking with NMR of L(CAAU). *The Journal of Physical Chemistry B* **2014**, *118*, 1216−1228.

(10) Bergonzo, C.; Cheatham, T. E. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *Journal of Chemical Theory and Computation* **2015**, *11*, 3969−3972.

(11) Mlynsky, V.; Kuhrova, P.; Kuhr, T.; Otyepka, M.; Bussi, G.; Banas, P.; Sponer, J. Fine-tuning of the AMBER RNA force field with a new term adjusting interactions of terminal nucleotides. *Journal of chemical theory and computation* **2020**, *16*, 3936−3946.

(12) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405−4412.

(13) Chen, A. A.; García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2013**, *110*, 16820−16825.

(14) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E1346−E1355.

(15) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B* **2015**, *119*, 5113−5123.

(16) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *The Journal of Physical Chemistry Letters* **2010**, *1*, 183−189.
(17) Yoo, J.; Aksimentiev, A. Improved parameterization of amine−carboxylate and amine−phosphate interactions for molecular dynamics simulations using the CHARMM and AMBER force fields. *Journal of chemical theory and computation* **2016**, *12*, 430−443.
(18) Yoo, J.; Aksimentiev, A. Refined parameterization of non-bonded interactions improves conformational sampling and kinetics of protein folding simulations. *The journal of physical chemistry letters* **2016**, *7*, 3812−3818.
(19) Yoo, J.; Kim, H.; Aksimentiev, A.; Ha, T. Direct evidence for sequence-dependent attraction between double-stranded DNA controlled by methylation. *Nat. Commun.* **2016**, *7*, 11045.
(20) Srivastava, A.; Timsina, R.; Heo, S.; Dewage, S. W.; Kirmizialtin, S.; Qiu, X. Structure-guided DNA-DNA attraction mediated by divalent cations. *Nucleic Acids Research* **2020**, *48*, 7018.
(21) Yoo, J.; Aksimentiev, A. New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* **2018**, *20*, 8432−8449.
(22) Bindewald, E.; Hayes, R.; Yingling, Y. G.; Kasprzak, W.; Shapiro, B. A. RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. *Nucleic Acids Res.* **2008**, *36*, D392−D397.
(23) Bailor, M. H.; Sun, X.; Al-Hashimi, H. M. Topology Links RNA Secondary Structure with Global Conformation, Dynamics, and Adaptation. *Science* **2010**, *327*, 202−206.
(24) Bailor, M. H.; Mustoe, A. M.; Brooks, C. L.; Al-Hashimi, H. M. 3D maps of RNA interhelical junctions. *Nature Protocols* **2011**, *6*, 1536−1545.
(25) Yesselman, J. D.; Denny, S. K.; Bisaria, N.; Herschlag, D.; Greenleaf, W. J.; Das, R. Sequence-dependent RNA helix conformational preferences predictably impact tertiary structure formation. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 16847−16855.
(26) Shi, H.; Rangadurai, A.; Assi, H. A.; Roy, R.; Case, D. A.; Herschlag, D.; Yesselman, J. D.; Al-Hashimi, H. M. Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. *Nat. Commun.* **2020**, *11*, 5531.
(27) Chen, Y.-L.; Sutton, J. L.; Pollack, L. How the Conformations of an Internal Junction Contribute to Fold an RNA Domain. *The Journal of Physical Chemistry B* **2018**, *122*, 11363−11372.
(28) Bai, Y.; Chu, V. B.; Lipfert, J.; Pande, V. S.; Herschlag, D.; Doniach, S. Critical Assessment of Nucleic Acid Electrostatics via Experimental and Computational Investigation of an Unfolded State Ensemble. *J. Am. Chem. Soc.* **2008**, *130*, 12334−12341.
(29) Templeton, C.; Elber, R. Why Does RNA Collapse? The Importance of Water in a Simulation Study of Helix−Junction−Helix Systems. *J. Am. Chem. Soc.* **2018**, *140*, 16948−16951.
(30) He, W.; Chen, Y.-L.; Pollack, L.; Kirmizialtin, S. The structural plasticity of nucleic acid duplexes revealed by WAXS and MD. *Sci. Adv.* **2021**, *7*, No. eabf6106.
(31) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophysical journal* **2007**, *92*, 3817−3829.
(32) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*. DOI: 10.1103/PhysRevLett.100.020603
(33) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(34) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(35) Chen, Y.-L.; Lee, T.; Elber, R.; Pollack, L. Conformations of an RNA helix-junction-helix construct revealed by SAXS refinement of MD simulations. *Biophysical journal* **2019**, *116*, 19−30.
(36) Sutton, J. L.; Pollack, L. Tuning RNA flexibility with helix length and junction sequence. *Biophysical journal* **2015**, *109*, 2644−2653.
(37) Walczewska-Szewc, K.; Corry, B. Accounting for dye diffusion and orientation when relating FRET measurements to distances: three simple computational methods. *Phys. Chem. Chem. Phys.* **2014**, *16*, 12317−12326.
(38) Lu, X.-J.; Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature protocols* **2008**, *3*, 1213.
(39) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2729−2742.
(40) Bottaro, S.; Gil-Ley, A.; Bussi, G. RNA folding pathways in stop motion. *Nucleic Acids Res.* **2016**, *44*, 5883−5891.
(41) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angewandte Chemie International Edition* **1999**, *38*, 236−240.
(42) Plumridge, A.; Meisburger, S. P.; Pollack, L. Visualizing single-stranded nucleic acids in solution. *Nucleic Acids Res.* **2017**, *45*, No. e66.
(43) Doose, S.; Barsch, H.; Sauer, M. Polymer properties of polythymine as revealed by translational diffusion. *Biophys. J.* **2007**, *93*, 1224−1234.
(44) Plumridge, A.; Andresen, K.; Pollack, L. Visualizing disordered single-stranded RNA: connecting sequence, structure, and electrostatics. *J. Am. Chem. Soc.* **2020**, *142*, 109−119.
(45) Woodson, S. A. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Current Opinion in Chemical Biology* **2005**, *9*, 104−109.
(46) Burkard, M. E.; Turner, D. H. NMR Structures of r (GCA G GC G UGC) 2 and Determinants of Stability for Single Guanosine-Guanosine Base Pairs. *Biochemistry* **2000**, *39*, 11748−11762.
(47) Lerman, Y. V.; Kennedy, S. D.; Shankar, N.; Parisien, M.; Major, F.; Turner, D. H. NMR structure of a 4× 4 nucleotide RNA internal loop from an R2 retrotransposon: Identification of a three purine−purine sheared pair motif and comparison to MC-SYM predictions. *RNA* **2011**, *17*, 1664−1677.
(48) Roost, C.; Lynch, S. R.; Batista, P. J.; Qu, K.; Chang, H. Y.; Kool, E. T. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* **2015**, *137*, 2107−2115.
(49) Tawani, A.; Kumar, A. Structural insights reveal the dynamics of the repeating r (CAG) transcript found in Huntington's disease (HD) and spinocerebellar ataxias (SCAs). *PloS one* **2015**, *10*, No. e0131788.
(50) Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Sci. Adv.* **2018**, *4*, No. eaar8521.
(51) Bottaro, S.; Bussi, G.; Pinamonti, G.; Reißer, S.; Boomsma, W.; Lindorff-Larsen, K. Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* **2019**, *25*, 219−231.

https://doi.org/10.1021/acs.jpclett.2c00359
J. Phys. Chem. Lett. 2022, 13, 3400−3408