Supporting Information for "A Computational Study of RNA Tetraloop Thermodynamics, Including Misfolded States"

Gül H. Zerze,[†] Pablo M. Piaggi,[‡] and Pablo G. Debenedetti[*,†]

†*Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA*

‡*Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA*

E-mail: pdebene@princeton.edu

Supporting Text

Simulations with ff99bsc0$_{\chi_{OL3}}$ force field

In addition to the DESRES force field[1] (whose results are presented in the main text), we also tested the most recent Amber RNA force field variant, the modified Amber ff99bsc0$_{\chi_{OL3}}$[2–5] whose library files are provided by Kuhrova et al.[6] We combined this nucleic acid force field with OPC(4-site) water model[7] and ions were modeled with Joung - Cheatham parameters for TIP4PEw.[8] We used the same thermodynamic conditions (same simulation box type and initial volume, same ionic concentration, same pressure, and same range of temperatures with the same temperature and pressure coupling algorithms) as in the DESRES force field simulations. We also applied precisely the same advanced sampling methodology with the same parameters.

We observe a strong initial condition dependence and the lack of convergence after simulating the systems for long period of time with this force field (Figure S11 and S12), similar to previous reports of lack of convergence.[6,9] We calculated a much larger error between IC1 and IC2 (Figure S11A, right column), compared to the error calculated for the same tetraloop modeled with DESRES ff (Figure S2C). We found that some replicas are stuck in configurations that are affected from strong nonspecific intramolecular contacts (Figure S12A). This strong persistence of configurations with nonspecific contacts eventually leads to a severe initial condition dependence as other groups also reported before.[6,9] Despite the initial condition dependence, the misfolded state that we identified in tetraloops modeled by the DESRES ff also exists in the tetraloops modeled with ff99bsc0χ_{OL3} (Figure S11). A representative configuration of the most populated misfolded configuration is shown in Figure S11 for each tetraloop. We analyzed the alpha and zeta torsions as well (Figure S13) and verified that they are the same as in the M of the tetraloops modeled by DESRES ff. Only exception is in the misfolded configuration of GAGA tetraloop where the A_7 base is flipped out as reported for the "misfolded-bulge" configuration before.[9] We also note that these misfolded structures (except for that for GAGA) have been observed before, for example, "4-purine stack" configuration of Kuhrova et al.[6] as they reported for an 8mer tetraloop. Details of ff99bsc0χ_{OL3} simulations are presented in the subsection below.

Since we did not observe convergence in the ff99bsc0χ_{OL3} simulations (1.5 μs per replica simulation time was not enough to converge the simulations (Figure S12)), we presented the results for DESRES force field in the main text. We note that for correcting the biases in ff99bsc0χ_{OL3} force field, Kuhrova et al. have proposed a local fix by supporting selected (native) hydrogen bonds (HBfix).[6] They later extended this approach to "gHBfix" by weakening overstabilized nonspecific interactions and weakening/strengthening specific interactions.[10] They implemented this fix as a patch on AMBER simulation suite and they also made a GROMACS compatible version available through PLUMED. However, we did not use it in this work, as running this fix in GROMACS through PLUMED significantly slowed down

the simulations (we found a four-fold slow down). The authors reported no significant slow down for AMBER suite.[10] For the sake of consistency, we avoid changing the simulation suite. The authors also argued that the main force field that we used (DESRES force field) can also be improved with HBfix approach,[10] which may be worth investigating further in a future study.

Supporting Figures

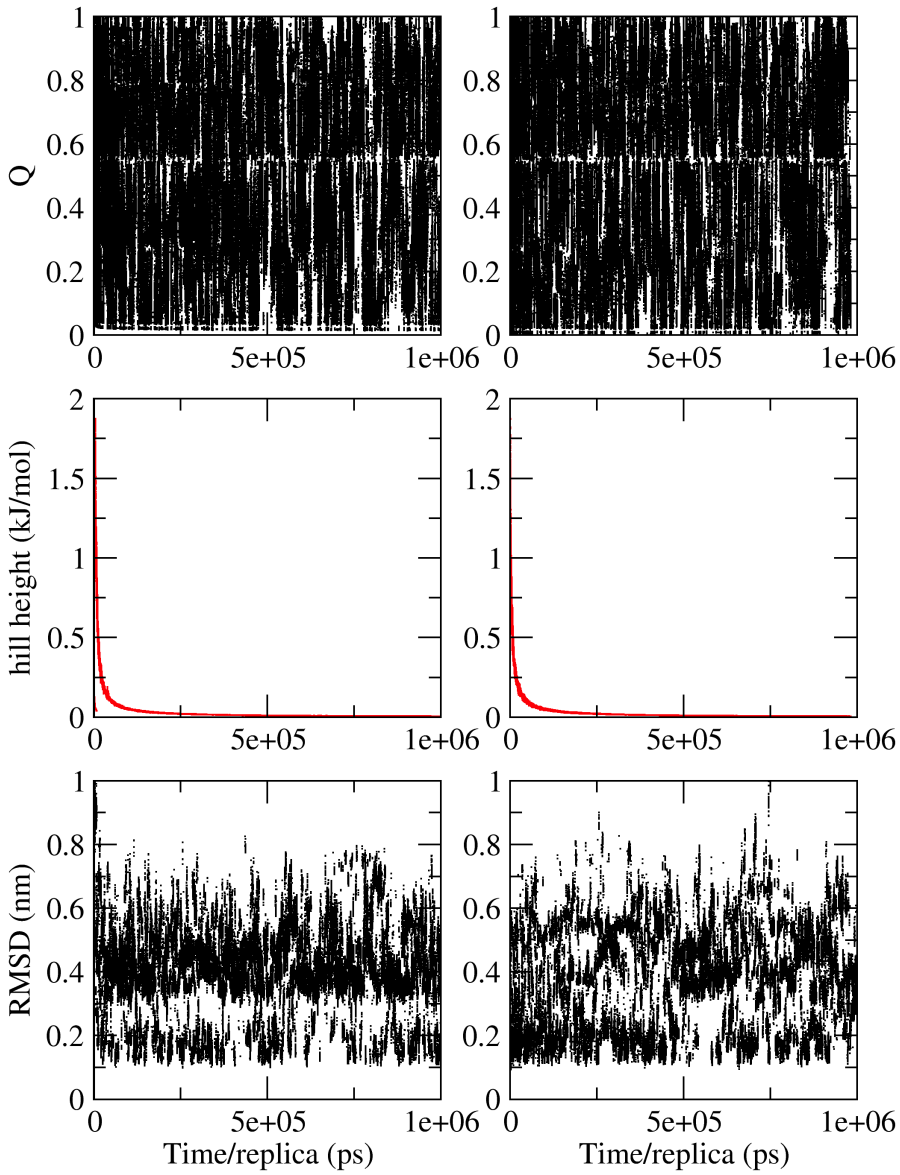

Figure S1: Sampling of Q (top), decay of hill height (middle), and RMSD (bottom) as a function of time for GAGA tetraloop starting from two independent initial conditions (left: IC1, right: IC2).

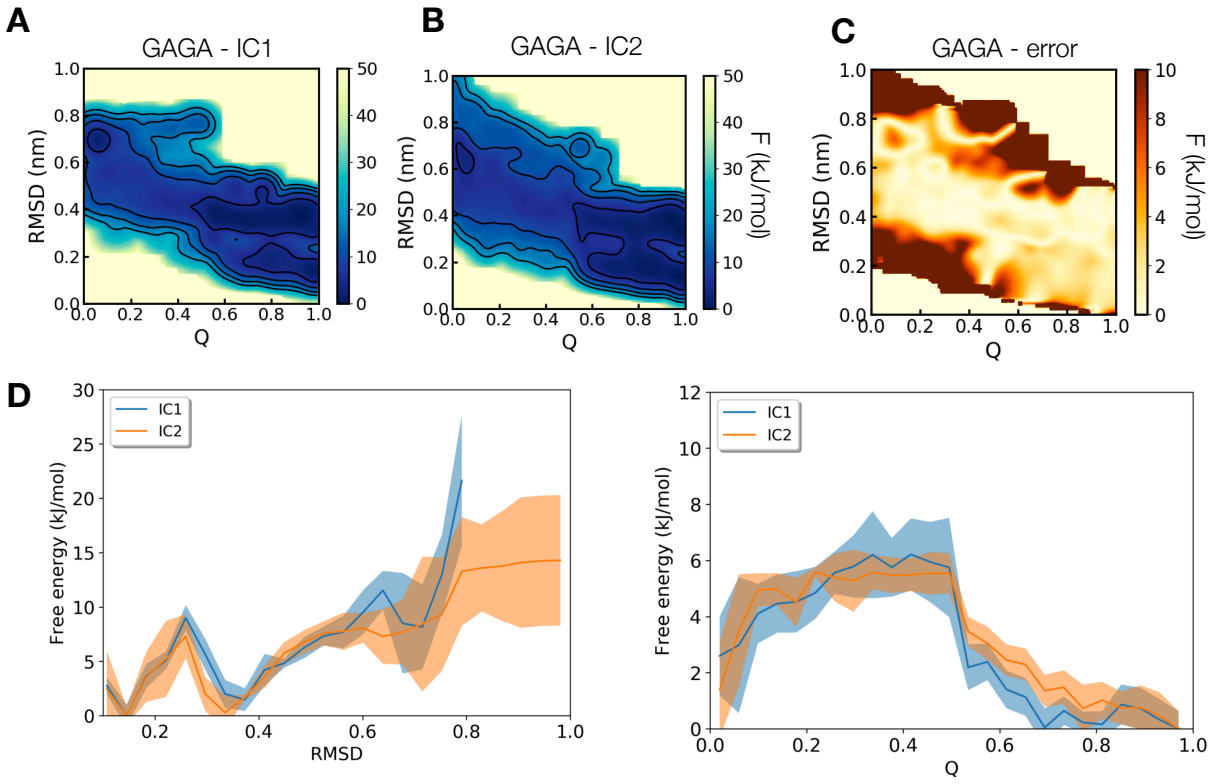

Figure S2: FES of GAGA tetraloop as a function of Q and RMSD for different initial conditions at 300 K (A. IC1) and B. IC2). The absolute deviation between IC1 and IC2 is reported as error. (C) We note that the error does not exceed ± kT within the regions of low free energy (< 20 kJ/mol). D. The one-dimensional projections of the free energy on RMSD and Q for IC1 and IC2. The shaded regions are blocked standard errors, dividing the equilibrated production data (last 500 ns/replica) into 4 equal, non-overlapping blocks.

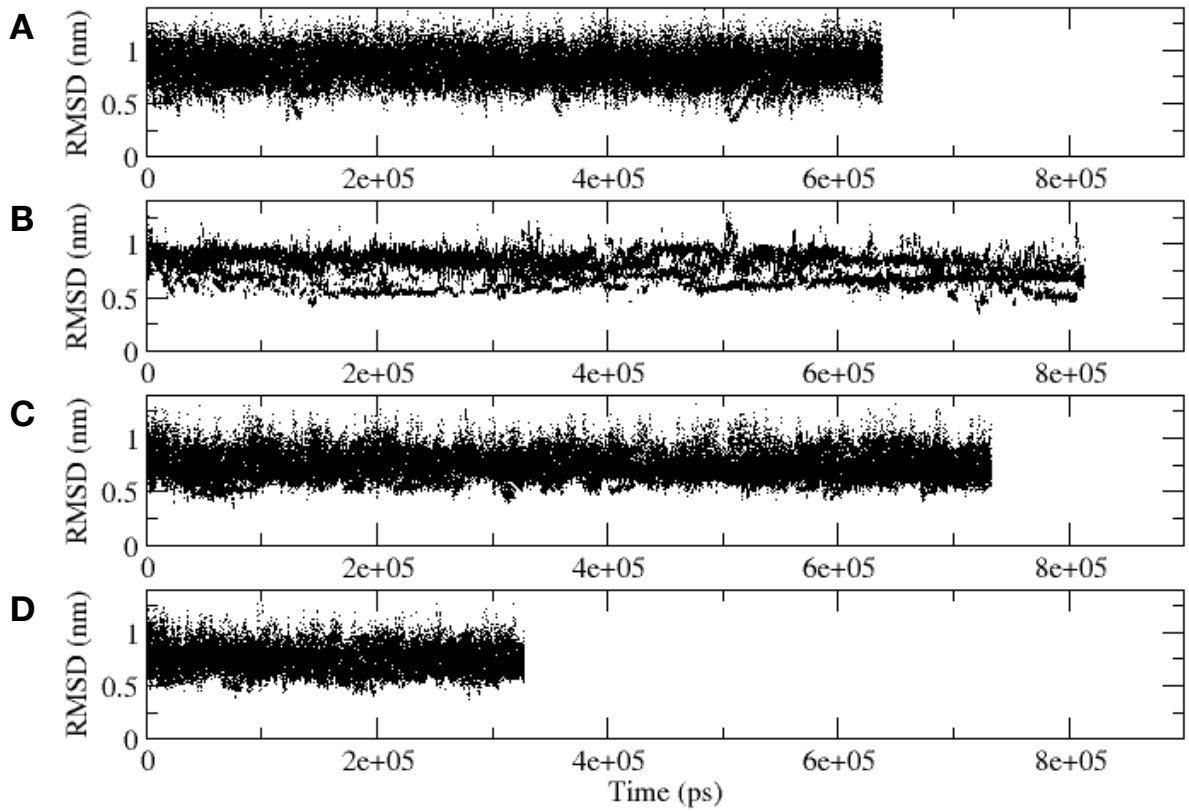

Figure S3: RMSD as a function of time (ns) per replica at 300 K from plain parallel-tempering simulations. A. GAGA (DESRES ff) B. GAGA (ff99bsc0χ_{OL3}) C. GAAA (ff99bsc0χ_{OL3}) D. GCAA (ff99bsc0χ_{OL3}) .

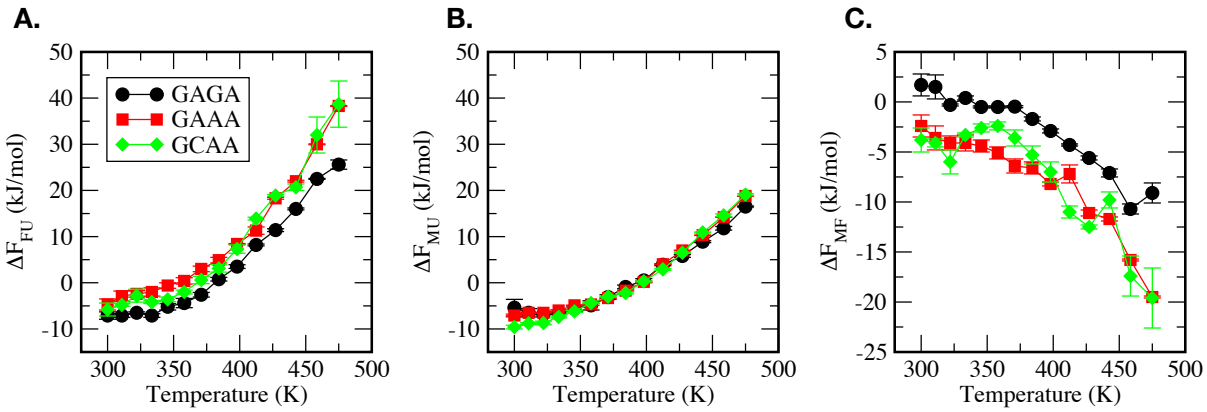

Figure S4: Temperature-dependent free energy differences between folded and unfolded (A); misfolded and unfolded (B); and folded and misfolded (C) states. For each temperature, we calculated the free energy differences using the equation $\Delta F_{AB} = -kT \ln P_A/P_B$, where P_A and P_B are the unbiased probabilities of finding the system in states A and B. A and B denote states F and U; M and U; and M and F for ΔF_{FU}; ΔF_{MU}; and ΔF_{MF}, respectively. The definition of the states F, M, and U in terms of the Q and RMSD variables is given in the first subsection of Results and Discussion and labeled on Figure 1. ΔF_{MF} decreases with temperature, which is consistent with basin M being structurally more heterogeneous (larger entropy), i.e., the M state encompasses a larger number of clusters compared to basin F, which contains only one cluster (see Figure 1, clusters).

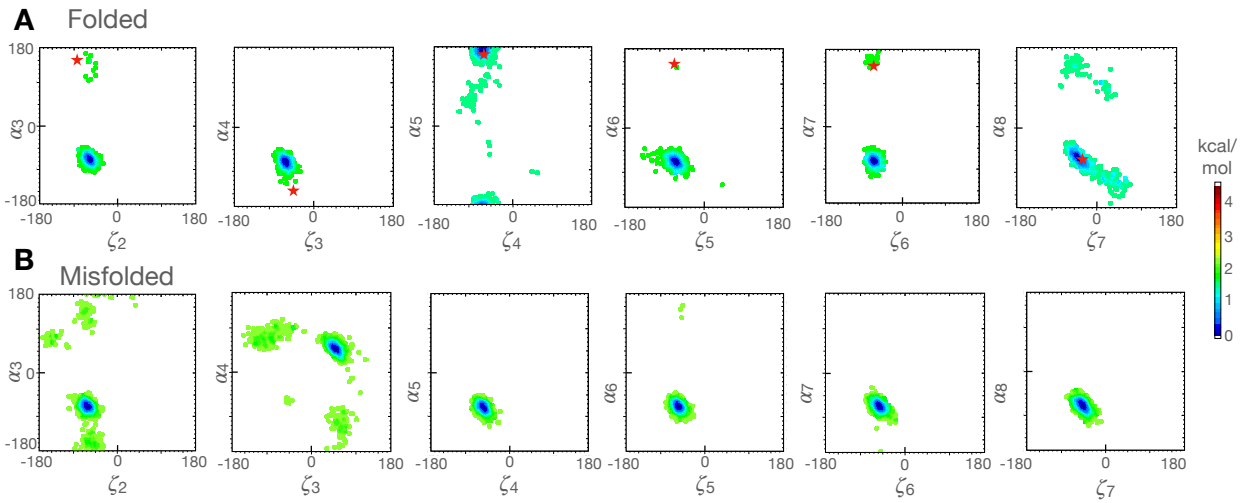

Figure S5: Free energies projected on alpha and zeta dihedrals of GAAA tetraloop, for the entire subpopulation in basin F (A) and for the most populating cluster in basin M (B). Dihedral angles in the native state, which are marked with a red star on each panel, belong to the first of the ten model structures deposited in the PDB entry 1ZIF.[11]

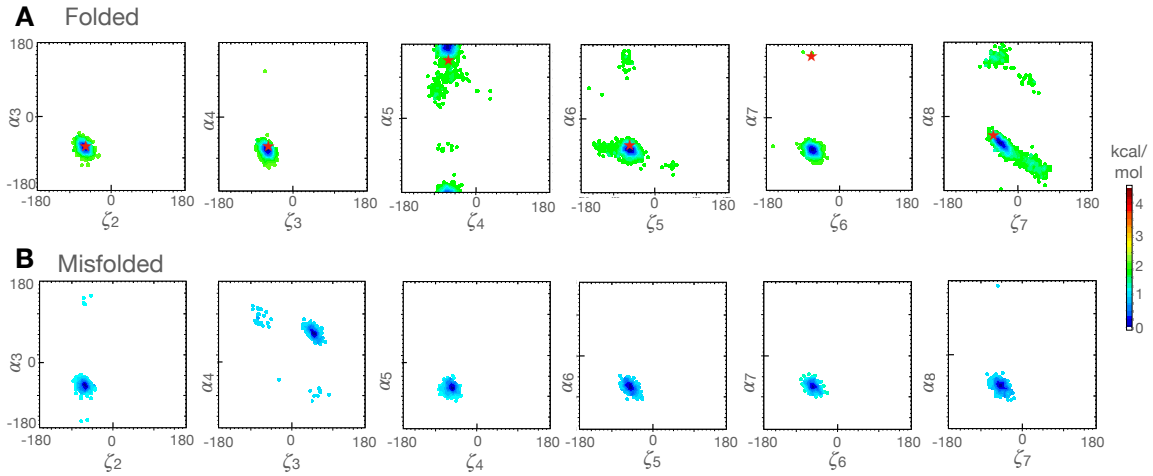

Figure S6: Free energies projected on alpha and zeta dihedrals of GCAA tetraloop, for the entire subpopulation in basin F (A) and for the most populating cluster in basin M (B). Dihedral angles in the native state, which are marked with a red star on each panel, belong to the first of the ten model structures deposited in the PDB entry 1ZIH.[11]

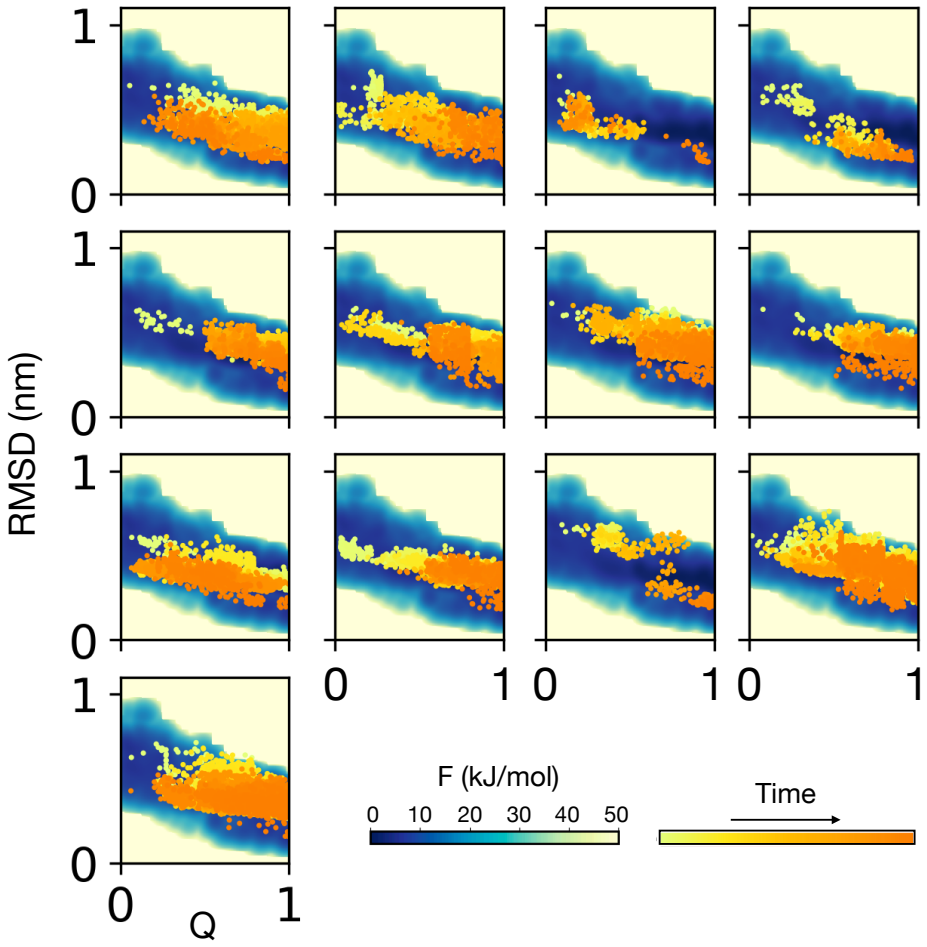

Figure S7: The paths of each reactive trajectory (starting from U and landing in F) found for GAAA tetraloop are illustrated on its two dimensional (Q vs RMSD) free energy surface (evaluated at 300K). Only 2 trajectories achieved such a transition without visiting the state M (first row-third column, first row-fourth column) out of 13 total transitions.

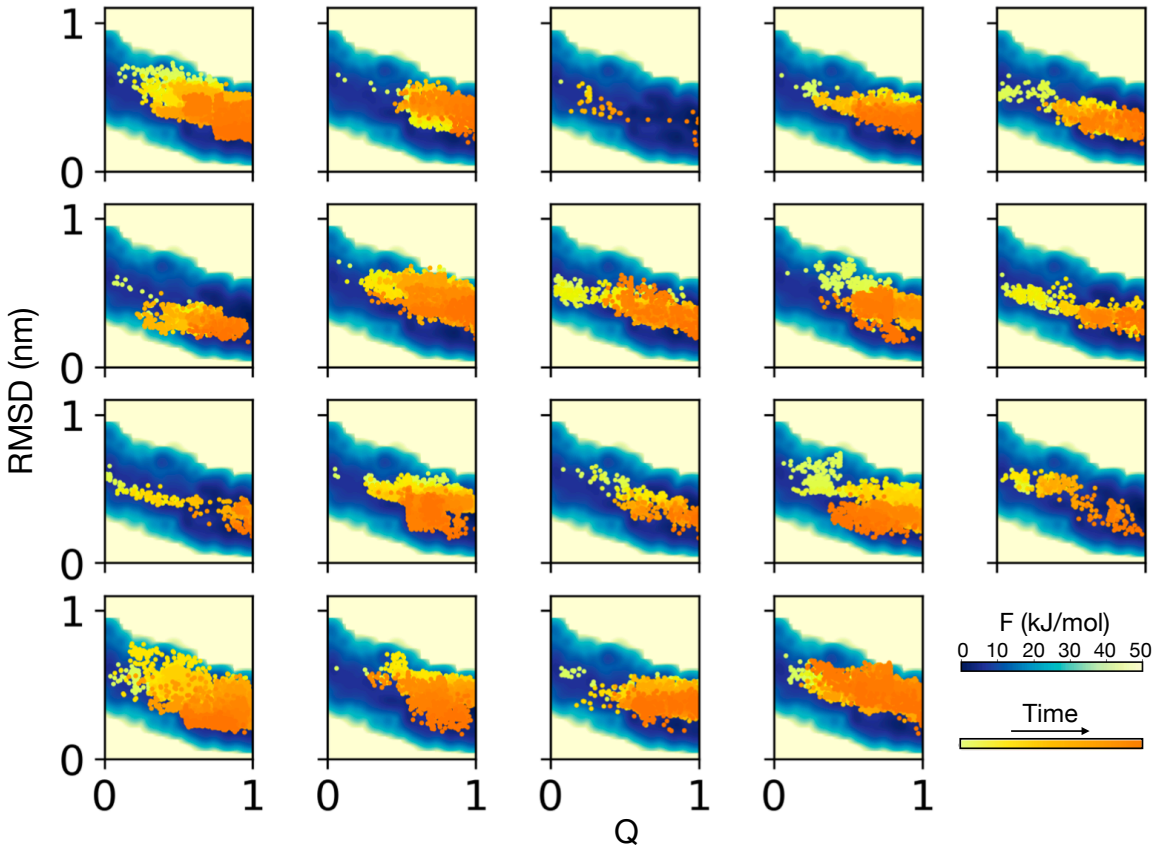

Figure S8: The paths of each reactive trajectory (starting from U and landing in F) found for GCAA tetraloop are illustrated on its two dimensional (Q vs RMSD) free energy surface (evaluated at 300K). Only 1 trajectory achieved such a transition without visiting the state M (third row-fifth column) out of 19 total transitions.

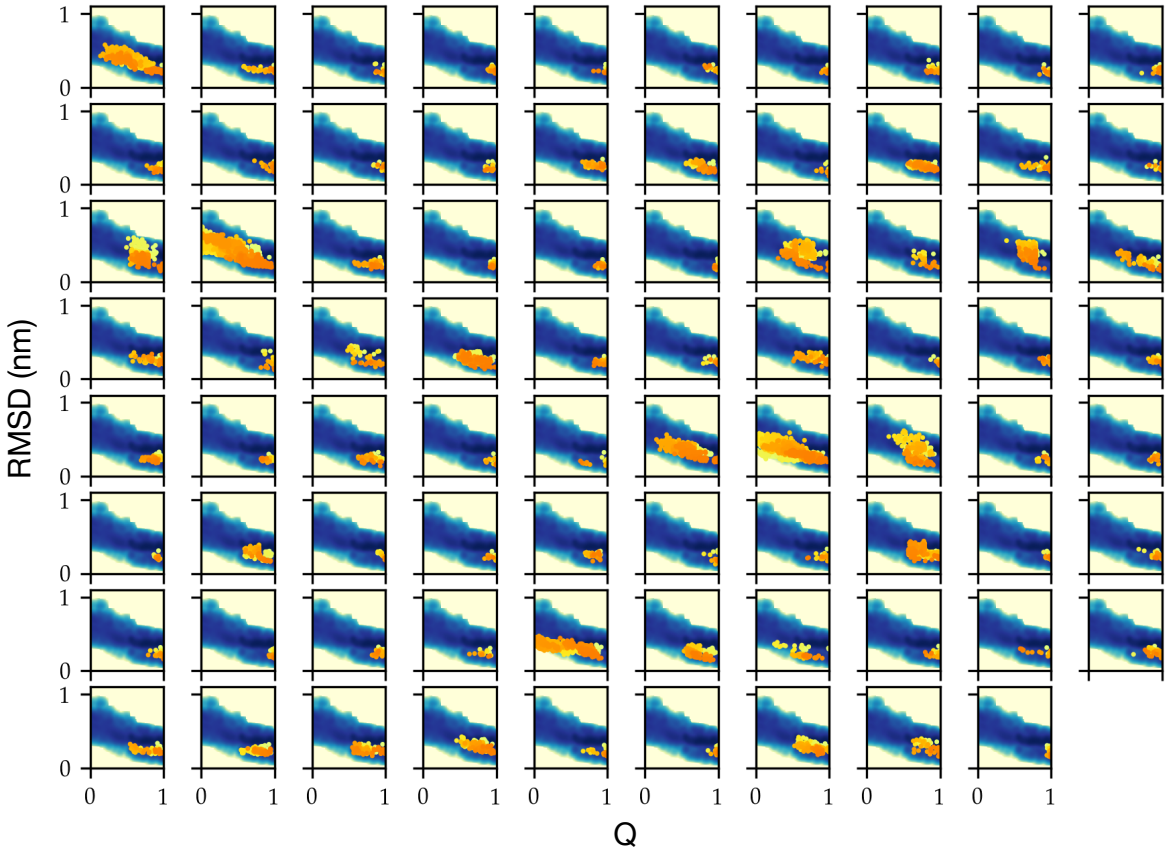

Figure S9: The paths of each reactive trajectory (starting from M and landing in F) found for GAAA tetraloop are illustrated on its two dimensional (Q vs RMSD) FES (evaluated at 300K). Only 9 trajectories went through extensive unfolding (visiting Q < 0.5) to achieve M to F transition out of total of 79 M to F transitions.

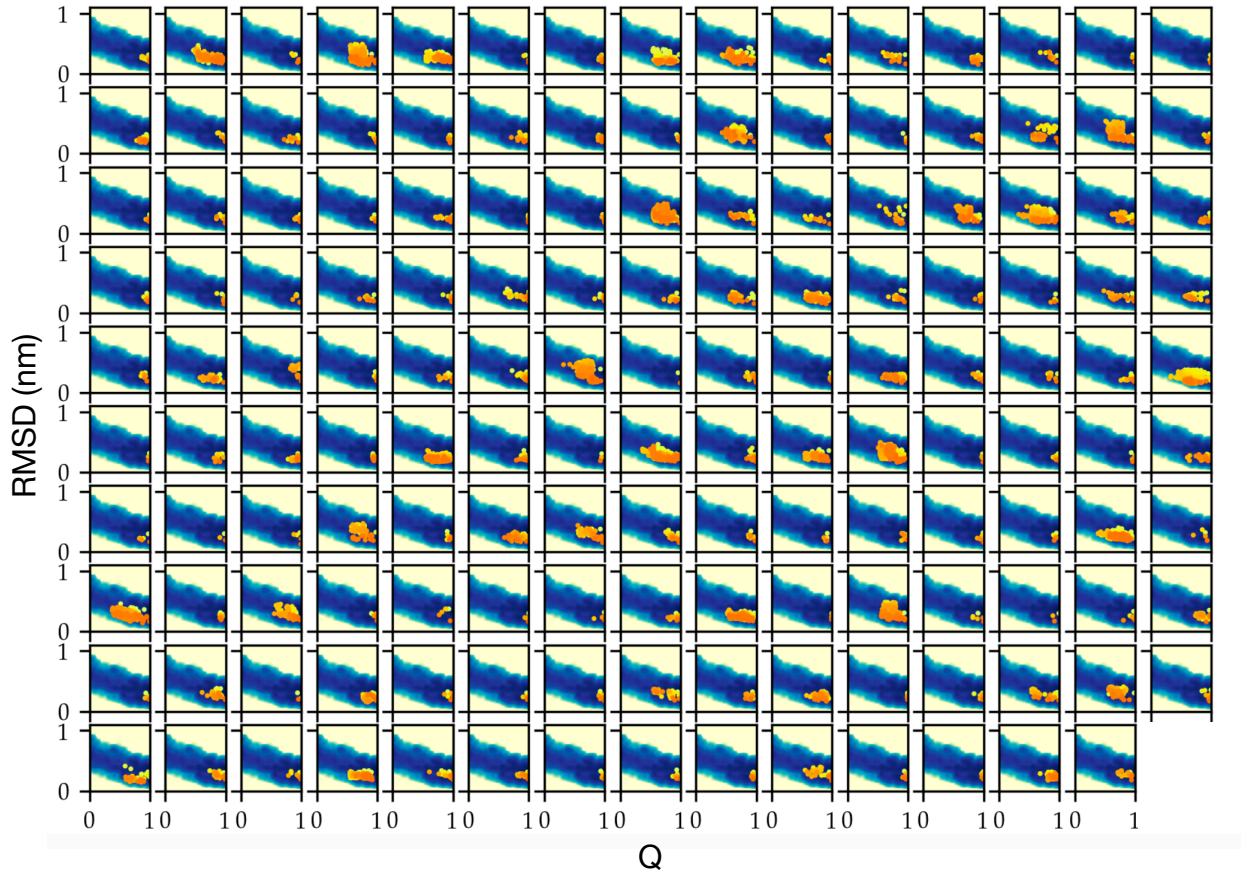

Figure S10: The paths of each reactive trajectory (starting from M and landing in F) found for GCAA tetraloop are illustrated on its two dimensional (Q vs RMSD) FES (evaluated at 300K). Only 8 trajectories went through extensive unfolding (visiting Q < 0.5) to achieve M to F transition out of total of 149 M to F transitions.

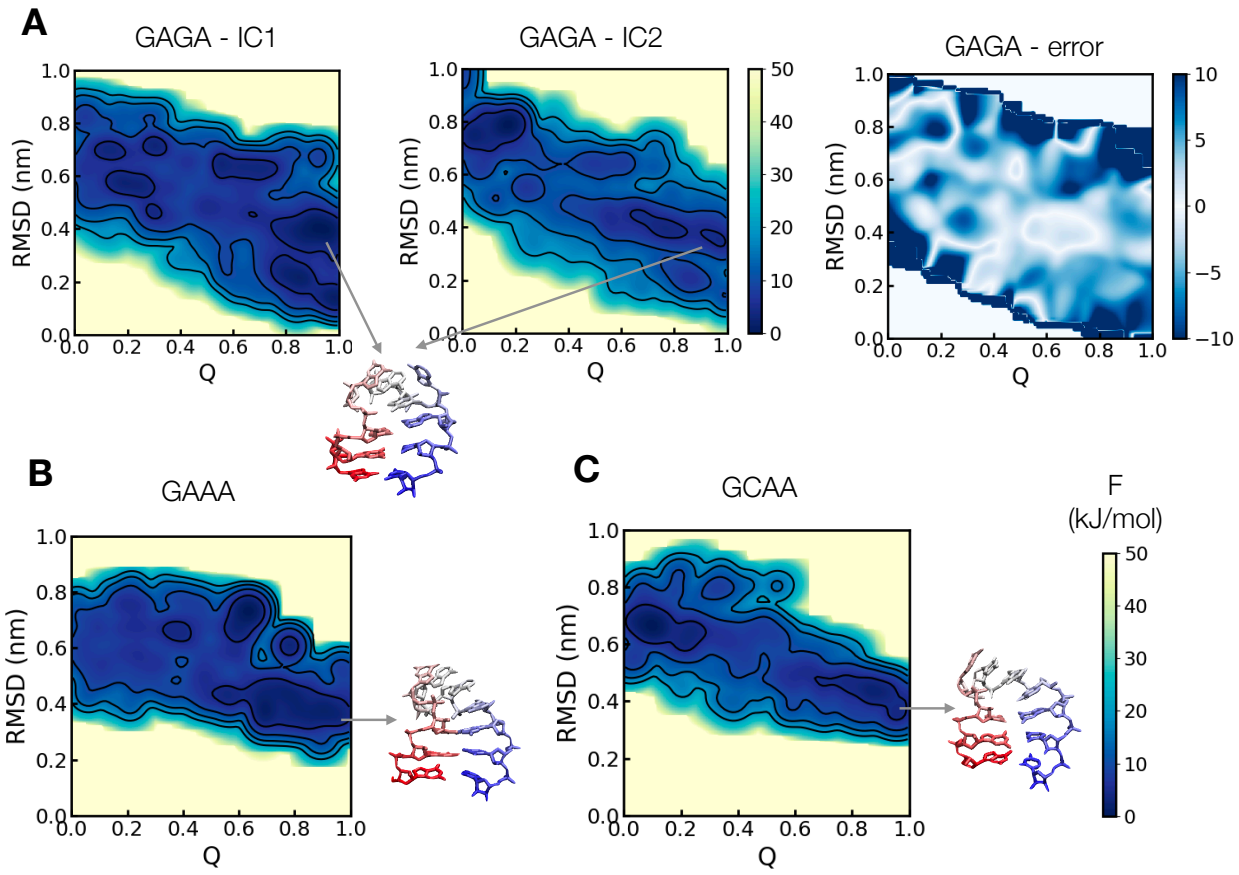

Figure S11: FES of tetraloops modeled with a modified ff99bsc0χ_{OL3} (see the Methods in the main text) at 300 K. A. GAGA tetraloop for IC1 (left) and IC2 (middle). The error calculated as the difference between the FES for IC1 and FES for IC2 is reported in the right panel. B. GAAA tetraloop c. GCAA tetraloop. Representative structures of the most populating clusters from the misfolded basin is shown for each tetraloop. Percentages of these clusters are 26, 28, and 38 for GAGA, GAAA, and GCAA tetraloops, respectively.

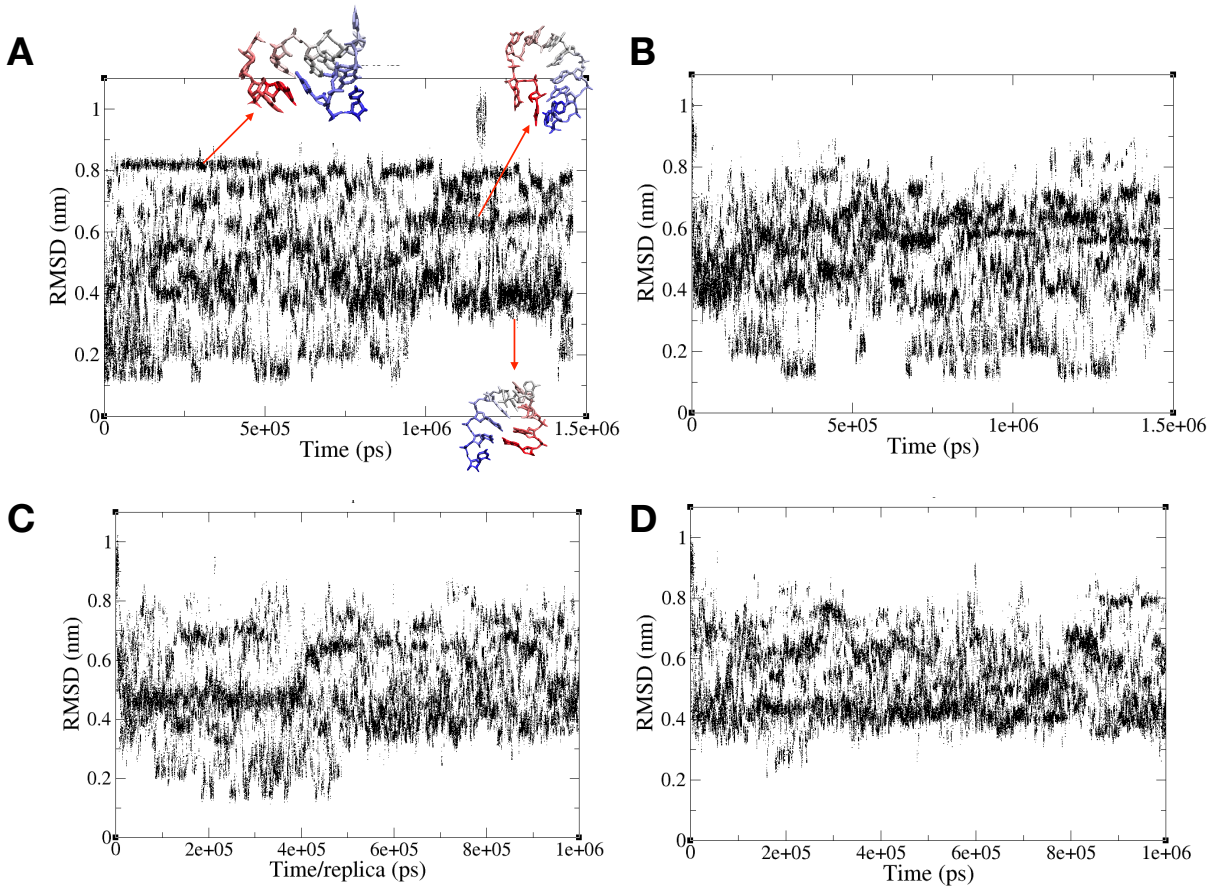

Figure S12: RMSD as a function of time/replica (ns) for the tetraloops modeled with the modified ff99bsc0χ_{OL3}. A. GAGA IC1. Configurations that persisted for a long while during the course of the simulation are shown with arrows. B. GAGA IC2 C. GAAA D. GCAA

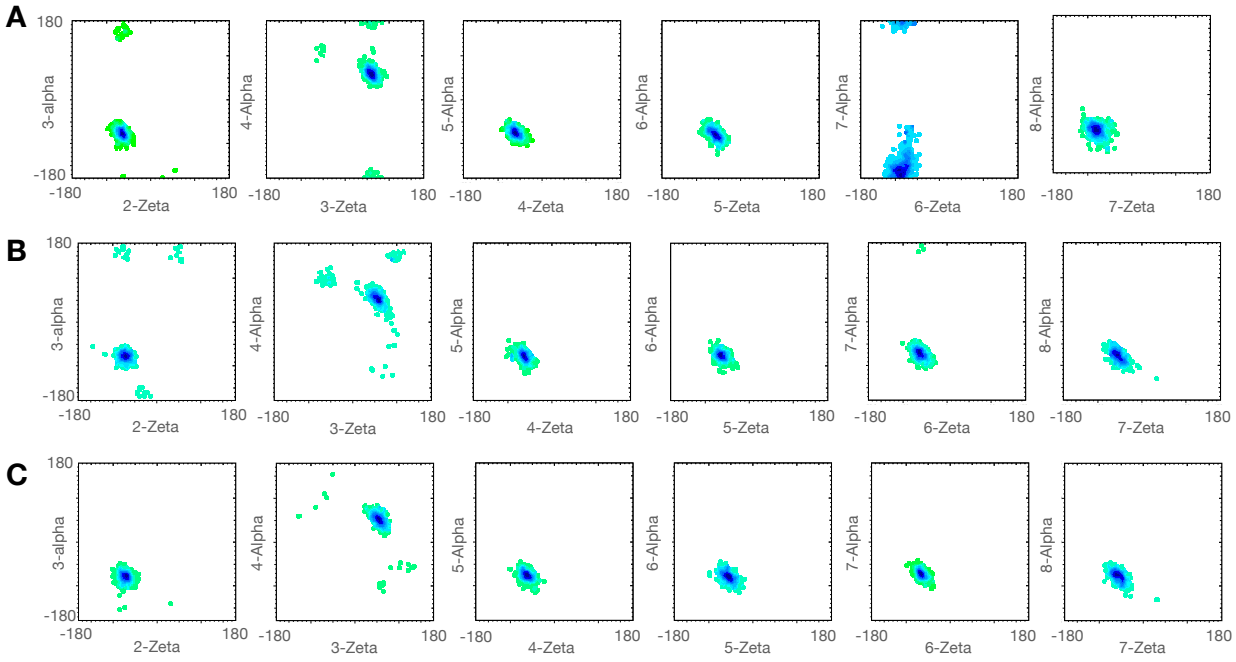

Figure S13: Free energies projected on alpha and zeta dihedrals for their most populated misfolded cluster (a representative structure is this cluster is shown in Figure S12) of the tetraloops modeled with ff99bsc0χ_{OL3} A. GAGA B. GAAA C. GCAA.

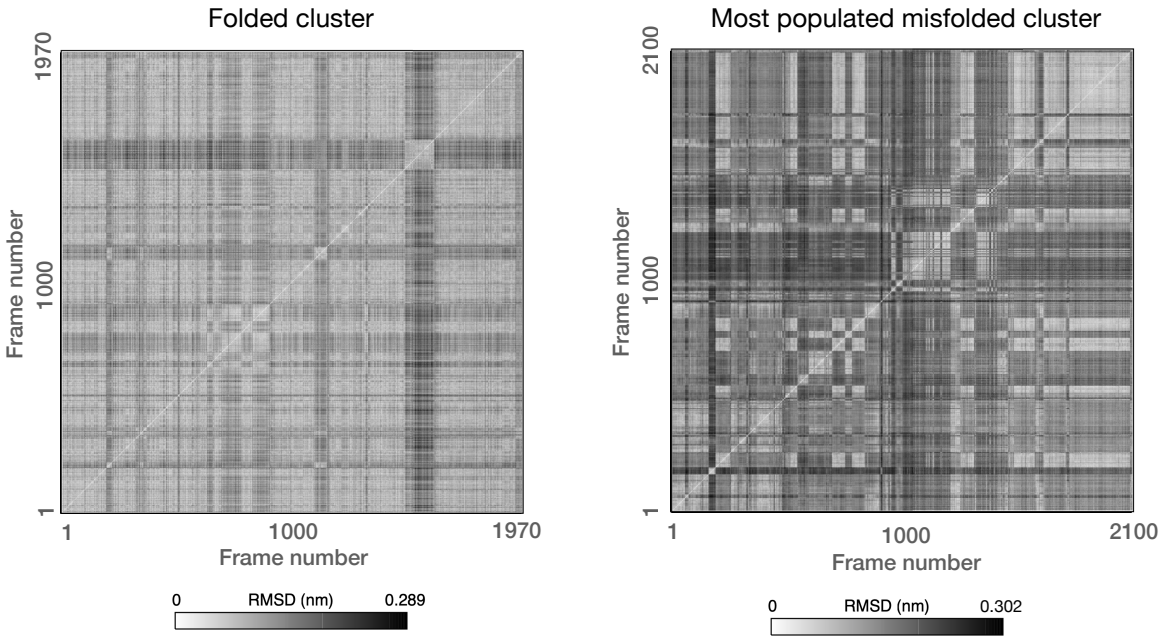

Figure S14: RMSD between individual configurations (indicated by frame numbers) that make up the folded cluster (left) and the most populated misfolded cluster (right) of the GAGA tetraloop. The folded cluster is composed of 1970 structures whereas the most populated misfolded cluster contains 2100 structures.

References

(1) 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. Nat. Acad. Sci.* **2018**, *115*, E1346–E1355.

(2) 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.

(3) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

(4) Zgarbová, M.; Otyepka, M.; Spoer, J.; Mladek, A.; Banas, P.; Cheatham III, T. E.; Jurecka, P. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

(5) Steinbrecher, T.; Latzer, J.; Case, D. Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405–4412.

(6) Kuhrova, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banas, P. Computer folding of RNA tetraloops: identification of key force field deficiencies. *J. Chem. Theory Comput.* **2016**, *12*, 4534–4548.

(7) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building water models: a different approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.

(8) Joung, I. S.; Cheatham III, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

(9) Kuhrova, P.; Banas, P.; Best, R.; Sponer, J.; Otyepka, M.; and, Computer folding of RNA tetraloops? Are we there yet? *J. Chem. Theory Comput.* **2013**, *9*, 2115–2125.

(10) Kuhrova, P.; Mlynsky, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Sponer, J.; Banas, P. Improving the performance of the amber RNA force field by tuning the hydrogen-bonding interactions. *J. Chem. Theory Comput.* **2019**, *15*, 3288–3305.

(11) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H.; Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **1996**, *264*, 968–980.