JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Insights into Stability and Folding of GNRA and UNCG Tetraloops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics

Susanta Haldar,[†,‡,#] Petra Kührová,[§,#] Pavel Banáš,[§] Vojtěch Spiwok,[∥] Jiří Šponer,[⊥] Pavel Hobza,*[,†,§] and Michal Otyepka*[,§]

[†]Institute of Organic Chemistry and Biochemistry and Gilead Science Research Center, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

[‡]Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University in Prague, Albertov 6, 128 43 Prague 2, Czech Republic

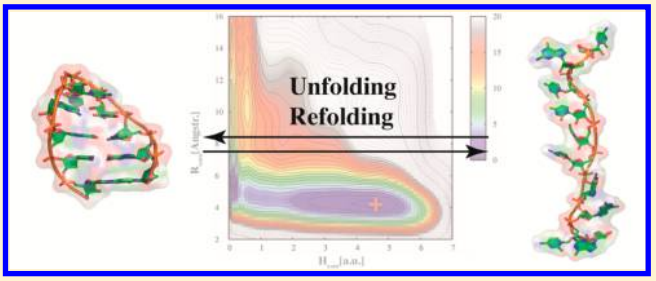
[§]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University Olomouc, tř. 17. Listopadu 12, 771 46 Olomouc, Czech Republic

[∥]Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague, Technická 3, 166 28 Prague 6, Czech Republic

[⊥]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

Supporting Information

ABSTRACT: RNA hairpins capped by 5′-GNRA-3′ or 5′-UNCG-3′ tetraloops (TLs) are prominent RNA structural motifs. Despite their small size, a wealth of experimental data, and recent progress in theoretical simulations of their structural dynamics and folding, our understanding of the folding and unfolding processes of these small RNA elements is still limited. Theoretical description of the folding and unfolding processes requires robust sampling, which can be achieved by either an exhaustive time scale in standard molecular dynamics simulations or sophisticated enhanced sampling methods, using temperature acceleration or biasing potentials. Here, we study structural dynamics of 5′-GNRA-3′ and 5′-UNCG-3′ TLs by 15-μs-long standard simulations and a series of well-tempered metadynamics, attempting to accelerate sampling by bias in a few chosen collective variables (CVs). Both methods provide useful insights. The unfolding and refolding mechanisms of the GNRA TL observed by well-tempered metadynamics agree with the (reverse) folding mechanism suggested by recent replica exchange molecular dynamics simulations. The orientation of the glycosidic bond of the G_{L4} nucleobase is critical for the UUCG TL folding pathway, and our data strongly support the hypothesis that G_{L4}-*anti* forms a kinetic trap along the folding pathway. Along with giving useful insight, our study also demonstrates that using only a few CVs apparently does not capture the full folding landscape of the RNA TLs. Despite using several sophisticated selections of the CVs, formation of the loop appears to remain a hidden variable, preventing a full convergence of the metadynamics. Finally, our data suggest that the unfolded state might be overstabilized by the force fields used.



INTRODUCTION

Correct structuring of RNA is critical for its biological function. However, our understanding of RNA folding and unfolding processes is still limited, even for small RNA molecules. GNRA (N stands for any nucleotide and R for purine) and UNCG tetraloops (TLs) are among the most prominent types of three-dimensional (3D) RNA structural motifs. These TLs enable a sharp bend of the RNA backbone[1] and are involved in many biologically important processes, including folding, stabilization of RNA helices, and interactions with other biomacromolecules, and they contribute to the structure and function of, e.g., ribozymes,[2−6] riboswitches,[7−9] and ribosome.[10−13] The highly thermostable UNCG TLs are considered centers where RNA folding can be initiated.[14,15] GNRA TLs are involved in recognition processes, which are realized via tertiary RNA−RNA, RNA−protein, and RNA−ligand interactions.[16] X-ray and NMR experiments have shown that GNRA and UNCG TLs have precisely defined 3D structures stabilized by a unique combination of rare noncanonical intermolecular interactions, determining their consensus sequences.[17−22] As single-strand RNAs containing TL motifs autonomously fold to form stable hairpins, they have attracted extensive attention in experimental and theoretical studies.[1,14,15,17−21,23−29] Recent theoretical

Received: January 6, 2015
Published: July 6, 2015

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

calculations, e.g., molecular dynamics (MD) or replica exchange MD (REMD) simulations, have aided our understanding of the structure, dynamics, and folding of TL motifs.[23−25,30]

MD simulations of RNA hairpins with GNRA and UNCG TLs—despite their small size—remained a nontrivial task for a long time. Early MD simulations failed to correctly describe the noncanonical but structurally strictly conserved structural features (known as signature interactions)[17] that maintain the GNRA and UNCG TLs' folds. Although often not admitted, these simulations resulted in gradual loss of the native structure of both TLs, despite achieving correct base-pairing of the stems. The lack of appropriate folding can be masked, for example, by using inadequate folding criteria, such as the root-mean-square deviation (RMSD) from the target structure.[31] This unsatisfactory simulation behavior was ultimately attributed to limitations of the empirical force fields (ff's) used in the classical MD simulations.[23] The dihedral parameters of the glycosidic bond (χ) were identified as important for the correct description of RNA and, in particular, the structural features of GNRA and UNCG TLs.[23,32,33] A recent reparameterization of the χ torsion of the Cornell et al. AMBER RNA force field[34] by Zgarbová et al.[35] was shown to well maintain all the key structural features of TLs.[23,25,35] Using this reparameterized force field, Kührová et al.[25] were able to correctly fold RNA sequences (5′-CGCGAGAGCG-3′) and (5′-CGCUUGGCG-3′) containing a GAGA and UUCG TL, respectively, in REMD simulations from the corresponding unfolded states. Later, Chen and Garcia[24] reported REMD folding of hairpins for shorter 8-mer sequences consisting of a two-base-pair canonical stem and GNRA or UNCG TLs with another variant of the force field. This included, besides similar χ dihedral tuning, reparameterized van der Waals parameters of nucleobases and modified combination rules. These successes may indicate that the new force fields are able to closely represent the native conformation of TLs, and therefore could potentially be used to explore (un)folding processes and the accompanying free energy changes.

Metadynamics (MetaD)[36] is a well-established simulation technique designed to overcome energy barriers and provide a more robust exploration of free energy surfaces (FESs).[37] This technique is able to search for different conformations of RNA on a reasonable time scale, which alleviates the enormous computational cost required for reconstruction of FESs of RNA molecules from unbiased MD simulations. It accelerates sampling through the introduction of a history-dependent biasing potential on chosen degrees of freedom, called the collective variables (CVs). Metadynamics basically helps the systems to escape from the trap of the free energy minima and accelerates sampling of rare events. In 2008, well-tempered metadynamics (WT-MetaD) was introduced to improve convergence of the free energy.[38] The main advantage of WT-MetaD is that the biasing potential converges within the simulation time and does not oscillate around the FES that has already been filled up. Thus, WT-MetaD avoids overfilling of the FES. Besides biasing of a few selected CVs, sampling of the orthogonal degrees of freedom might be simultaneously enhanced by the combination of metadynamics with parallel tempering (PT-MetaD),[39] giving the advantages of both of these mutually complementary methods. The efficiency of such a technique was subsequently improved even more by use of a well-tempered ensemble.[40] It should be noted, however, that the choice of CVs still remains the main issue in effective application of all CV-based sampling-enhancing methods. Choice of CVs is a nontrivial task that usually requires substantial insight into the studied system and some experience (i.e., several trial-and-error experiments).[37] In other words, metadynamics relies on the assumption that the dynamics of the simulated system can be effectively simplified using the CVs. As for all CV-based sampling-enhancing methods, correct results can be obtained only when the chosen CVs include all important slow dynamical motions that contribute to the studied process.[41] In addition, when evaluating the FESs, it should be taken into consideration that the (un)folding processes and unfolded (denaturated) state ensembles are always influenced by the physical method used for the denaturation. This applies to computational as well as experimental methods.[42]

In this work, we explored the structural dynamics of TLs using large-scale classical all-atom MD simulations and WT-MetaD in explicit water, which both complement contemporary insights from the latest REMD simulation studies.[24,25] The TLs are known to have (for a molecule of their size) a rather complex and rugged FES.[29,43] We aimed to reconstruct the FES of (un)folding of two small RNA (5′-CGCGAGAGCG-3′ and 5′-CGCUUCGGCG-3′) hairpins containing GNRA and UNCG TLs, respectively, using WT-MetaD simulations. In addition, we performed extensive 15-μs-long unbiased MD simulations of both TLs, which represent the most cutting-edge time scale currently available in explicit solvent MD. In all simulations, we used the latest ff99bsc0χ_{OL3} variant of the Cornell et al. AMBER RNA force field (sometimes also labeled as the ff10, ff12, or ff14 RNA force field).[34,35,44]

METHODS

System Settings. We studied small 10-mer RNA sequences 5′-CGCGAGAGCG-3′ (henceforth referred to as the GAGA hairpin) and 5′-CGCUUCGGCG-3′ (henceforth UUCG hairpin) containing the GAGA and UUCG TL motifs, respectively, capping an A-RNA stem comprising three Watson−Crick GC base pairs. The structure of the GAGA TL, including one GC base pair adjacent to the TL, was taken from the 1.04 Å resolution X-ray structure of the sarcin-ricin loop from *E. coli* 23S rRNA (PDB ID 1Q9A, residues 2658−2663)[45] and was extended by two additional GC base pairs. The structure of the UUCG TL, including one GC base pair adjacent to the TL, was taken from the high resolution NMR structure,[46] and, again, two GC base pairs were added to complete the hairpin loop (Figure 1).

Classical MD Simulations. We carried out long unbiased molecular dynamics (MD) simulations of both TLs at the cutting edge time scale of 15 μs for each TL. The simulations were executed using the Gromacs 4.5.1 package[49] with the all-atomic AMBER ff99bsc0χ_{OL3} force field. The ff99bsc0χ_{OL3} force field is based on the AMBER ff99[34] force field corrected by Barcelona α/γ[44] and Olomouc χ_{OL3}[35] reparameterizations. [ff99bsc0χ_{OL3} has been adopted as a standard AMBER RNA force field since 2010; i.e., it belongs to the ff10, ff12, and ff14 sets of AMBER default force fields.] The TLs were immersed in a rectangular simulation box and then solvated by TIP3P explicit water models and neutralized by adding 9 Na^+ counterions (r = 1.87 Å and ε = 0.0028 kcal/mol).[50] The simulations were carried out using periodic boundary conditions according to the Particle Mesh Ewald method with a 10 Å nonbonded cutoff and NpT conditions maintained through the use of a V-rescale thermostat[51] and isotropic Parrinello−Rahman (Andersen) barostat.[52] A 2.0 fs integration

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

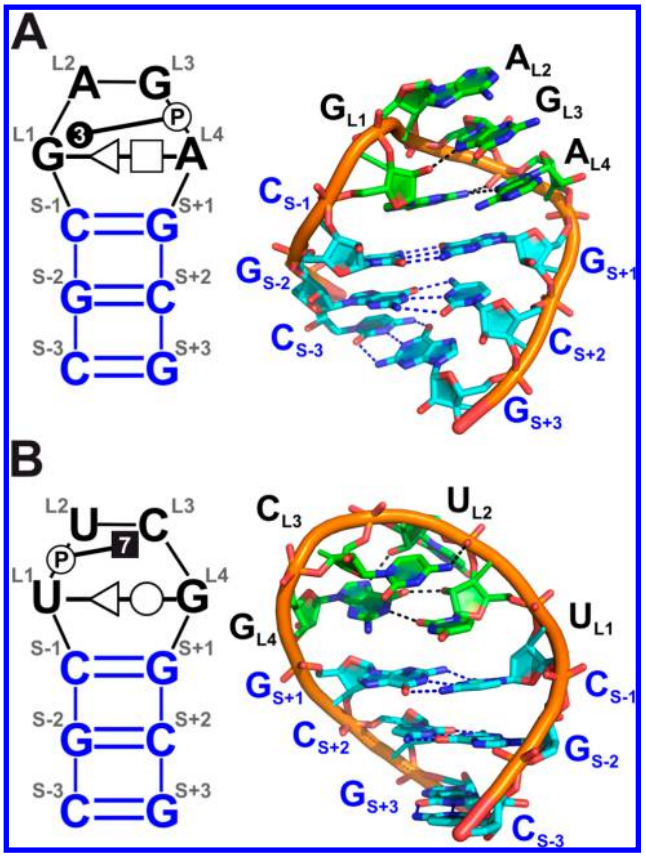


Figure 1. Secondary (left) and 3D (right) structures of the studied (A) GNRA and (B) UNCG hairpins, consisting of a short A-RNA stem (blue labels) and the tetraloop (TLs, black labels). The noncanonical base-pairing and base−phosphate interactions in the TL region are classified according to the standard Leontis−Westhof−Zirbel nomenclature;[47,48] the signature interactions of the TLs are shown in the structures by black dashed lines.

step and temperature of 300 K were used in the unbiased MD simulations.

Well-Tempered Metadynamics. Free energy surfaces of both TLs were explored by WT-MetaD simulations. The same system and simulation settings as described for the unbiased MD simulations were used for the WT-MetaD simulations. Prior to initiating the WT-MetaD, 5 ns equilibration MD runs (NpT) were performed with the Nosé−Hoover thermostat and isotropic Parrinello−Rahman (Andersen) barostat. The equilibration MD runs were followed by a 200-ns-long biased WT-MetaD simulation, performed with 1.0 fs time step at a temperature of 300 K. The PLUMED 1.3 plugin[53] was used to carry out the WT-MetaD simulations with the Gromacs code 4.5.5.[49] The biasing potential was calculated according to the WT-MetaD scheme using the following formula:

$$V(s,t) = \sum_{t'=0,\tau_G,2\tau_G,\ldots}^{t'<t} \omega\tau_G \, e^{-V(s(q(t'),t'))/\Delta T} \, e^{-\sum_{i=1}^{2}[(s_i(q)-s_i(q(t')))^2/2\sigma_i^2]} \quad (1)$$

where the deposition rate, ω, and deposition stripe, τ_G, of the Gaussian hills were set to 0.478 kcal/mol·ps (2.0 kJ/mol·ps) and 1 ps, respectively. The bias factor $(T + \Delta T)/T$ was set to 15, and the final FES was calculated as follows:

$$F(s,t) = -\frac{T+\Delta T}{\Delta T}(V(s,t) - C(t)) \quad (2)$$

Thus, the CVs were sampled at a fictitious temperature $T + \Delta T$ of 3000 K.

We tested the performance of several types of CVs, s_i. Each WT-MetaD simulation was biased by combination of two or three CVs; the particular choice of CVs is reported in the Result section for each WT-MetaD simulation. Namely, we employed the following types of CVs:

(i) H_{core}, which measured the extent of native hydrogen bonding in three base pairs of the stem, the L1−L4 noncanonical base pair, and the other signature interactions of the TL. This CV was chosen to guide the making/breaking of the native hydrogen bonds of the entire hairpin loop, i.e., both the loop and the stem.

(ii) R_{core} (radius of gyration), which accounted for the global shape of the RNA strand. This CV aimed to describe global bending from straight ssRNA to the bent hairpin fold; see, e.g., ref 54 using similar combination of H_{core} and R_{core} CVs.

(iii) RMSD of the first, third, and fourth nucleotides of the loop (those nucleotides are involved in the signature interactions within the loop) from the corresponding native structure. This CV should sample unfolding/refolding of the loop region.

(iv) In the case of the UUCG TL, also the glycosidic torsion of the G_{L4} nucleotide (G_{L4}-χ) was used as CV. This nucleotide is *syn*-oriented in the native structure.

The Gaussian widths were set to 0.2 arbitrary unit, 2 Å, 0.5 Å, and 0.2 rad for H_{core}, R_{core}, RMSD, and G_{L4}-χ, respectively.

The H_{core} CVs were calculated using the switching function as follows:

$$H_{core} = \sum_i \frac{1-(r_i/r_0)^n}{1-(r_i/r_0)^m} \quad (3)$$

where r_0 was set to either 2.0 or 2.5 Å, henceforth labeled as $H_{core,2.0}$ and $H_{core,2.5}$, respectively. The n and m parameters were set to 6 and 12, respectively. In the case of $H_{core,2.0}$, the index i corresponded in both TLs to the 11 hydrogen bonds involved in base-pairing of the three GC base pairs of the stem region and additional two hydrogen bonds of the L1−L4 noncanonical base pair of the TL. The $H_{core,2.5}$ included in addition one or two signature hydrogen bonds outside the L1−L4 base pair. The r_i was the distance between the hydrogen acceptor and hydrogen atom bound to the hydrogen donor of the above-mentioned hydrogen bonds (see Table 1 for the list of atoms involved in the definition of $H_{core,2.0}$ and $H_{core,2.5}$).

The R_{core} CVs were calculated as follows:

$$R_{core} = \frac{\sum_i^n (r_i - r_{com})^2 m_i}{\sum_i^n m_i} \quad (4)$$

where the sums extended either over six nitrogens of the stem nucleobases, one from each nucleobase, henceforth $R_{core,base}$, or over all nine phosphorus atoms, henceforth $R_{core,P}$ (see Table 1 for atom list), and the center of mass r_{com} was defined by

$$r_{com} = \frac{\sum_i^n r_i m_i}{\sum_i^n m_i} \quad (5)$$

The RMSD CV was defined as mass-weighted RMSD of L1, L3, and L4 nucleotides (i.e., G_{L1}, G_{L3} and A_{L4}, and U_{L1}, C_{L3} and

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

Table 1. Definition of the R_{core} and H_{core} Types of Coordinates, Showing the Atoms Used in the Calculation of the Radii of Gyration in the $R_{core,base}$ and $R_{core,P}$ Collective Variables, and Pairs of Hydrogen and Hydrogen Bond Acceptors Used in the Definition of the $H_{core,2.0}$ and $H_{core,2.5}$ Collective Variables

	Atoms Used in the Calculation of the R_{core} CVs	
	GAGA	UUCG
$R_{core,base}$	C_{S-3}(N3); G_{S-2}(N1); C_{S-1}(N3); G_{S+1}(N1); C_{S+2}(N3); G_{S+3}(N1)	C_{S-3}(N3); G_{S-2}(N1); C_{S-1}(N3); G_{S+1}(N1); C_{S+2}(N3); G_{S+3}(N1)
$R_{core,P}$	G_{S-2}(P); C_{S-1}(P); G_{L1}(P); A_{L2}(P); G_{L3}(P); A_{L4}(P); G_{S+1}(P); C_{S+2}(P); G_{S+3}(P)	G_{S-2}(P); C_{S-1}(P); U_{L1}(P); U_{L2}(P); C_{L3}(P); G_{L4}(P); G_{S+1}(P); C_{S+2}(P); G_{S+3}(P)

Pairs of Hydrogen and Hydrogen-Bond Acceptors Used in the Definition of the H_{core} CVs			
GAGA		UUCG	
H-bond donor	H-bond acceptor	H-bond donor	H-bond acceptor
Shared by $H_{core,2.0}$ and $H_{core,2.5}$			
C_{S-3}(H41)	G_{S+3}(O6)	C_{S-3}(H41)	G_{S+3}(O6)
G_{S+3}(H1)	C_{S-3}(N3)	G_{S+3}(H1)	C_{S-3}(N3)
G_{S+3}(H21)	C_{S-3}(O2)	G_{S+3}(H21)	C_{S-3}(O2)
C_{S+2}(H41)	G_{S-2}(O6)	C_{S+2}(H41)	G_{S-2}(O6)
G_{S-2}(H1)	C_{S+2}(N3)	G_{S-2}(H1)	C_{S+2}(N3)
G_{S-2}(H21)	C_{S+2}(O2)	G_{S-2}(H21)	C_{S+2}(O2)
C_{S-1}(H41)	G_{S+1}(O6)	C_{S-1}(H41)	G_{S+1}(O6)
G_{S+1}(H1)	C_{S-1}(N3)	G_{S+1}(H1)	C_{S-1}(N3)
G_{S+1}(H21)	C_{S-1}(O2)	G_{S+1}(H21)	C_{S-1}(O2)
G_{L1}(H21)	A_{L4}(*pro*-R_P)	G_{L4}(H1)	U_{L1}(O2)
G_{L1}(H22)	A_{L4}(N7)	U_{L1}(HO2′)	G_{L4}(N3)
Specific for $H_{core,2.5}$			
G_{L1}(HO2′)	G_{L3}(N7)	C_{L3}(HO2′)	G_{L4}(N7)
		C_{L3}(N4)	U_{L2}(*pro*-R_P)

G_{L4} nucleotides in case of GAGA and UUCG TLs, respectively) from their native conformation represented by the corresponding X-ray or NMR structure in case of GAGA and UUCG TL, respectively. Finally, the G_{L4}-χ CV was defined as the G_{L4}(O4′)−G_{L4}(C1′)−G_{L4}(N9)−G_{L4}(C4) dihedral angle.

We present altogether 11 biased simulations. First, we analyzed four WT-MetaD simulations, each biasing two CVs: $H_{core,2.0}$ + $R_{core,base}$ and $H_{core,2.5}$ + $R_{core,P}$ CVs in case of GAGA TL, and two independent simulations of UUCG with $H_{core,2.0}$+$R_{core,base}$ CVs. Furthermore, we performed seven additional WT-MetaD simulations, namely two independent GAGA simulations using $H_{core,2.5}$ + $R_{core,P}$; a GAGA simulation combining $H_{core,2.5}$ + $R_{core,P}$ + RMSD; two independent UUCG simulations with $H_{core,2.5}$ + $R_{core,P}$; an UUCG simulation with $H_{core,2.5}$ + RMSD; an UUCG simulation with $H_{core,2.5}$ + RMSD + G_{L4}-χ and, finally, one well-tempered ensemble parallel tempering metadynamics of the GAGA TL biasing $H_{core,2.0}$ + $R_{core,base}$ CVs (see Supporting Information for further details).[39,40]

RESULTS AND DISCUSSION

The aim of this work was to study the stability and free energy changes accompanying folding and unfolding of two short RNA hairpin loops containing either GNRA and UNCG TLs, namely, the GAGA and UUCG hairpins, using extensive unbiased MD and WT-MetaD simulations. Another aim was to assess limits of the CV-based approaches to study folding of small nucleic acids. As demonstrated below, metadynamics provides very useful insights into the studied systems, complementing other computational approaches. However, our study also clearly shows challenges when using CV-based methods to describe folding of RNA molecules, even small ones such as the hairpins. Although part of the difficulties reported below could be caused by the approximate nature of the force fields, our study indicates that the folding pathways of nucleic acids seem to be significantly more complex compared to the peptides of the same size (see, e.g., ref 39). Thus, a few CVs do not appear to be sufficient to achieve a complete description of the folding landscape of RNA TLs. This challenge might be solved by using biased-exchange metadynamics[55] (BEMD); see, e.g., ref 56 for a study of the folding of G-DNA quadruplex using BEMD. Nonetheless, we suggest that even the choice of CVs appears to be more challenging compared to the corresponding simulations of proteins. This should be taken into considerations in assessment of future metadynamics (or CV-based in general) studies of other nucleic acids systems.

Overview of the WT-MetaD Simulations. The WT-MetaD simulations should be able to provide sufficiently robust sampling of the unfolding−refolding events once a set of few CVs well representing such events is used. The choice of CVs is thus the crucial point for the applicability of WT-MetaD. We explored the efficiency of the unfolding−refolding sampling in the WT-MetaD simulations of both TLs using various choices of CVs and analyzed their performance.

The basic technical convergence of the FESs was monitored by calculating differences between two consecutive FESs constructed at times $i - 10$ and i ns, according to the scheme introduced by Branduardi et al.[57] This approach uses the former FES as a reference and examines regions in the CV space defined by points lying up to 5 kcal/mol above the global minimum of the reference FES (see Figure S1 in the Supporting Information). It should be noted that a progressive convergence between consecutive FESs still does not guarantee that the WT-MetaD simulation is converged. A false convergence can be, e.g., indicated when the system is trapped in some conformation corresponding to the local minima of a hidden variable (orthogonal to the selected CVs) that does not allow proper sampling of all states across the FES. Thus, besides that, we monitored unfolding−refolding events in the WT-MetaD simulations and considered the simulations to be fully converged WT-MetaD only when observing several repetitive unfolding−refolding events and diffusive behavior of the simulations in selected CVs.

GAGA TL. Initially, we have attempted a 200-ns-long WT-MetaD simulation of GAGA TL using two CVs, namely $H_{core,2.0}$ and $R_{core,base}$ (see Methods). Similar choice of CVs was successfully used, e.g., for folding of small β-hairpin peptide,[39] G-DNA triplex,[54] or G-DNA quadruplex.[56] In addition, this choice seems to be logical as the unfolded states of TL hairpins are characterized by an unpaired A-RNA stem region, loss of the TL signature interactions and straightening of the RNA strand.[58] Thus, one CV involving native hydrogen bonding contacts of the base pairs of the A-RNA stem and signature interactions of the noncanonical L1−L4 base pairs of the TL and the second CV accounting for the radius of gyration of the stem; i.e., the shape of the system should be enough to effectively describe the unfolding−refolding transitions. During the initial phase of both WT-MetaD simulations, we observed the expected rapid flooding of the native state region ($R_{core,base}$ ≈ 4−6, $H_{core,2.0}$ ≈ 4−7; the initial X-ray structure corresponded to $R_{core,base}$ ≈ 3.7, $H_{core,2.0}$ ≈ 6.5). However, at later times, this

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

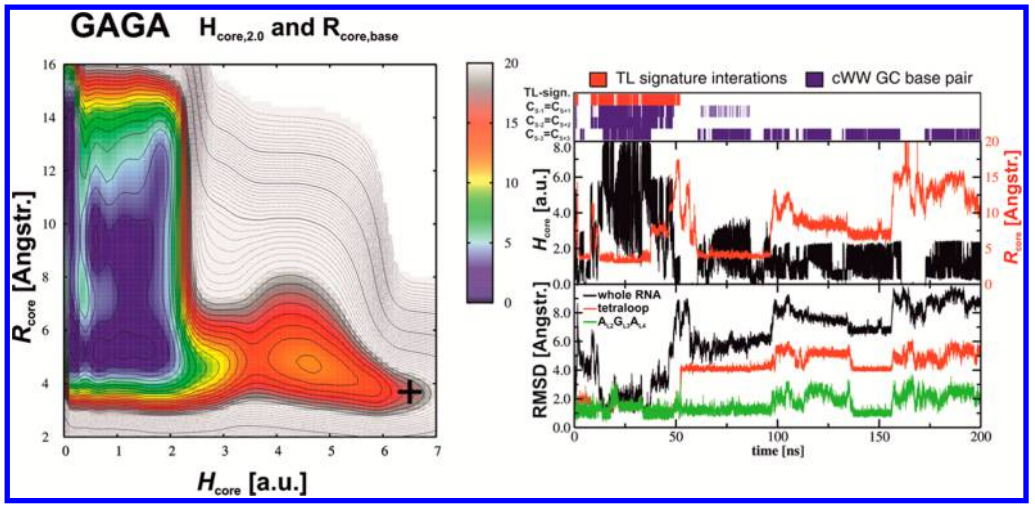


Figure 2. FES plot (left) and the development of key structural features (right) obtained from 200-ns-long WT-MetaD simulation of GAGA TL using $H_{core,2.0}$ and $R_{core,base}$ CVs. The FES is contoured by 2 kcal/mol (thin lines) and 10 kcal/mol (thick lines). Energies lower than 20 kcal/mol relevant for structural dynamics are colored. An X-ray structure corresponding to $H_{core,2.0}/R_{core,base}$ of 6.5/3.7 Å, respectively, is denoted by black cross. In the top of the right panel, the red and blue stripes indicate the presence of TL signature interactions and GC base pairs of the stem, respectively, both defined on the basis of hydrogen bonding with 4.0 Å cutoff for heavy-atom distance. The middle left panel shows the time evolution of the $H_{core,2.0}$ (black) and $R_{core,base}$ (red) CVs. The lower left panel shows evolution of the root-mean-square deviation (RMSD) of the whole RNA hairpin (black), the TL (red), and the tripurine $A_{L2}G_{L3}A_{L4}$ stack (green).

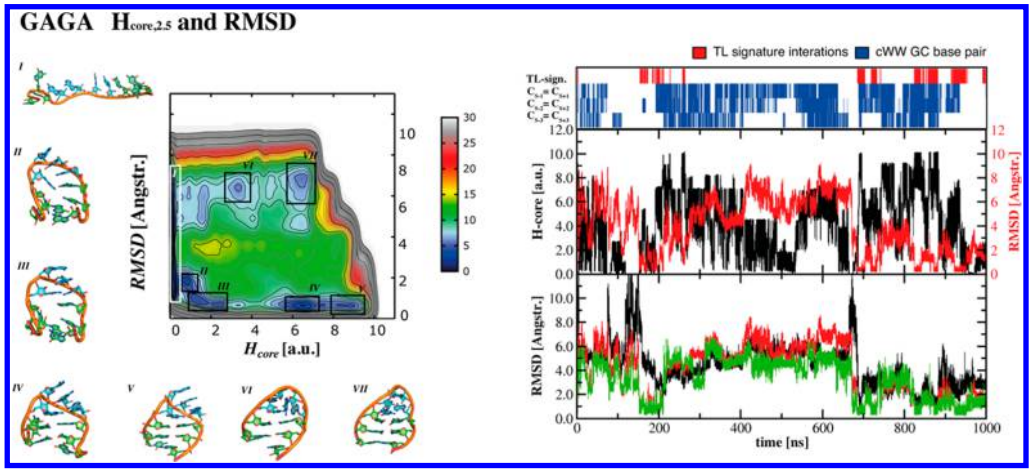


Figure 3. FES plot (left) and the development of key structural features (right) obtained from 1000-ns-long WT-MetaD simulations of GAGA WT-MetaD simulations using $H_{core,2.5}$ and RMSD of the loop region from the native conformation as CVs. The colors and contours correspond of the FES as well as the definition of all structural parameters of the right panel corresponds to those of Figure 2. The regions depicted on the FES by black boxes correspond to the conformational states populated during the simulation, the corresponding structures are shown in sticks and cartoons.

simulation never returned to the native state region and remained sampling only unfolded and misfolded states with $R_{core,base}$ and $H_{core,2.0}$ in the range 3−14 and 0−3, respectively (Figure 2). This might be explained either by the fact that selected CVs were not able to accelerate refolding event, or the possibility that GNRA TL is inherently not stable under the force field description and the above-mentioned extensively populated minimum on FES corresponded to some misfolded state artificially stabilized by the force field. The later possibility would mean that the native state is only kinetically stabilized on the microsecond time scale in classical MD simulations.[23] In order to test such hypothesis, we selected a snapshot corresponding to the R_{core} = 9 Å and H_{core} = 1.5 au (see the left panel of the Figure 2) and used it as a starting structure for a classical MD simulation. The GAGA TL however did not maintain these CV values and rapidly transformed to a straight ssRNA conformation. Thus, the free energy minima observed in the region of H_{core} = 1.5 au most likely corresponded to the fully unfolded RNA strand with a random formation of a few native hydrogen bonds which were, however, not able to form any inherently stable structure with a longer lifetime. In such a case, the refolding event could be mainly hindered by entropy (due to conformationally rich ssRNA state). Although we cannot entirely rule out the effect of the force field, the possibility that the utilized CV were not sufficiently robust to accelerate the refolding sampling should be considered as more likely explanation.

We then tested different choices of CVs aiming to improve the sampling of the refolding events. However, even an alternative definition of CVs based on native hydrogen bonds and radius of gyration did not improve the propensity of WT-MetaD simulations for sampling of the refolding events (see Supporting Information for two additional 400-ns-long WT-

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

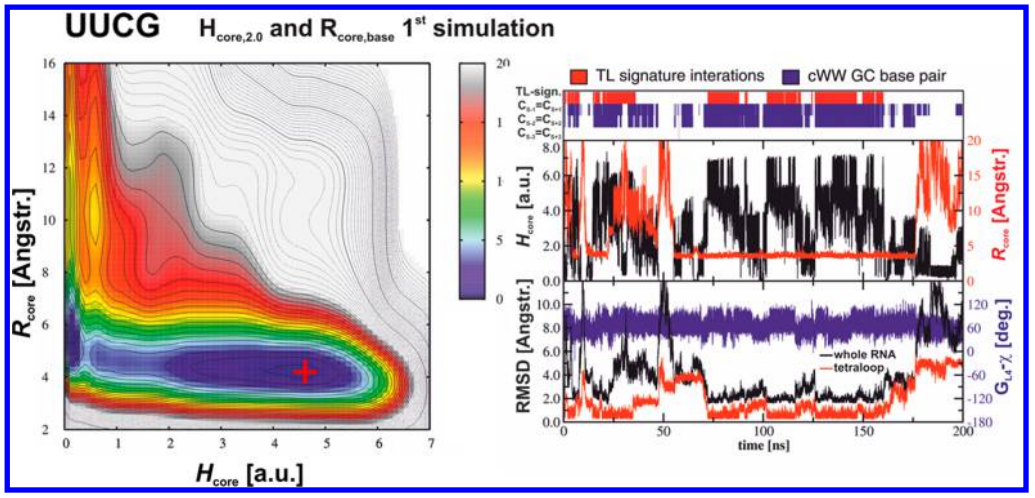


Figure 4. FES plot (left) and the development of key structural features (right) obtained from the first (G_{L4}-*syn* populating) 200-ns-long WT-MetaD simulation of UUCG TL using $H_{core,2.0}$ and $R_{core,base}$ CVs. The contouring and coloring scheme of the FES is same as in Figure 2. The MNR structure corresponding to $H_{core,2.0}/R_{core,base}$ of 4.6/4.2 Å, respectively, is denoted by red cross. The time development of the TL signature interactions and GC base pairs of the stem (top right) as well as the CVs (middle right) are presented as in Figure 2. The lower panel shows evolution of the root-mean-square deviation (RMSD) of the whole RNA hairpin (black), the TL (red), and glycosidic χ torsion of G_{L4} (blue).

MetaD simulations of GAGA TL using $H_{core,2.5}$ and $R_{core,P}$ CVs).

This motivated us to completely redefine the CVs and include the description of the native loop formation using RMSD of the loop region from the corresponding native conformation as a new CV (see Methods). Namely we performed two additional 400-ns-long WT-MetaD simulations of GAGA TL using RMSD and $H_{core,2.5}$, or RMSD, $H_{core,2.5}$, and $R_{core,P}$ CVs, i.e., with two or three CVs, respectively. Similarly to the WT-MetaD simulation described above (using $H_{core,2.0}$ and $R_{core,base}$), we observed the expected rapid loss of the native structure during the initial phase of both additional WT-MetaD simulations. Similarly to the first simulation, the simulation with three CVs (RMSD, $H_{core,2.5}$, and $R_{core,P}$) never returned to the native state and sampled the transitions between fully unfolded ssRNA and misfolded states with properly formed stem but non-native conformation of the loop. Thus, even this simulation failed to sample refolding events (see Figure S8 in the Supporting Information). However, the refolding events were finally sampled in the simulation with two CVs (RMSD and $H_{core,2.5}$) (see Figure 3). Again, the system rapidly flooded the basin of the native fold conformation (state VI in Figure 3) during initial phase of this simulation and sampled unfolded (state I in Figure 3) and misfolded states with formed stem region but unstructured loop (states VII and VIII in Figure 3). However, after that the system formed the $A_{L2}|G_{L3}|A_{L4}$ tripurine stack within the ssRNA conformation (state II in Figure 3) and finished formation of the loop by establishing the *t*HS $A_{L4}G_{L1}$ base pair (state III in Figure 3), followed by zipping the stem region (states IV and V in Figure 3, with one and two base pairs adjacent to the loop formed, respectively). After this refolding event, the system lost the native conformation of the loop (with the stem still firmly formed) and populated misfolded states (states VII and VIII in Figure 3). We thus decided to extent this simulation as much as possible up to 1000 ns, and we finally observed the several additional refolding evens after ∼700 ns sharing similar pathway with the first above-mentioned refolding event. The FES revealed several minima corresponding to above-mentioned folded (states IV and V in Figure 3), unfolded (state I in Figure 3), and misfolded states. It is worth noting that one of the unfolded states (state VI in Figure 3) was similar to the GL4-misfolded bulged structure described in our recent REMD study[25] resembling the X-ray structure of GNRA TL in complex with the restrictocin protein.[59] It seems that folded and misfolded states are similar in free energy, while the unfolded state appeared to be the most stable. However, we would like to reiterate that despite the relatively long simulation time scale and observation of several refolding events, the behavior of CVs is still not diffusive. Thus, even in this case, the simulation cannot be considered fully converged and most likely much longer simulation would be required to achieve a quantitative convergence. Therefore, also the interpretation of the energetics revealed by WT-MetaD simulation should be discussed with care.

UUCG TL. Similarly as in the case of GAGA TL, we initially tested the combination of $H_{core,2.0}$ and $R_{core,base}$ CVs (see Methods) in 200-ns-long WT-MetaD simulation of the UUCG TL. Again we observed the expected rapid flooding of the native state region ($R_{core,base} \approx 4-6$, $H_{core,2.0} \approx 4-7$; the initial NMR structure corresponded to $R_{core,base} \approx 4.2$, $H_{core,2.0} \approx 4.6$). However, in contrast to the GAGA, this UUCG WT-MetaD simulation reversibly fluctuated between unfolded states and near-native states (denoted as frayed or flipped-over states, sharing all features with the native folded state except that the terminal base pair was possibly broken;[60] see Figure S2 in the Supporting Information and Figure 4). Despite that the UUCG WT-MetaD in first sight readily sampled the unfolding−refolding events, we noticed one problem stemming from the basic principle of metadynamics and the particular choice of CVs used. The G_{L4} nucleotide retained its native *syn* conformation during the whole 200 ns WT-MetaD simulation (Figure 4). Its *syn* conformation is an important part of the UNCG signature, so the molecule remained poised to fold. This meant that the used WT-MetaD protocol did not sample conformational states along a genuine unfolding pathway with *anti* orientation of the G_{L4} nucleobase. This is because the *syn*−*anti* dynamics was not accelerated since it was not explicitly included in the chosen CVs.

In order to better understand the role of G_{L4} *syn*−*anti* dynamics during unfolding−refolding processes, we carried out

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

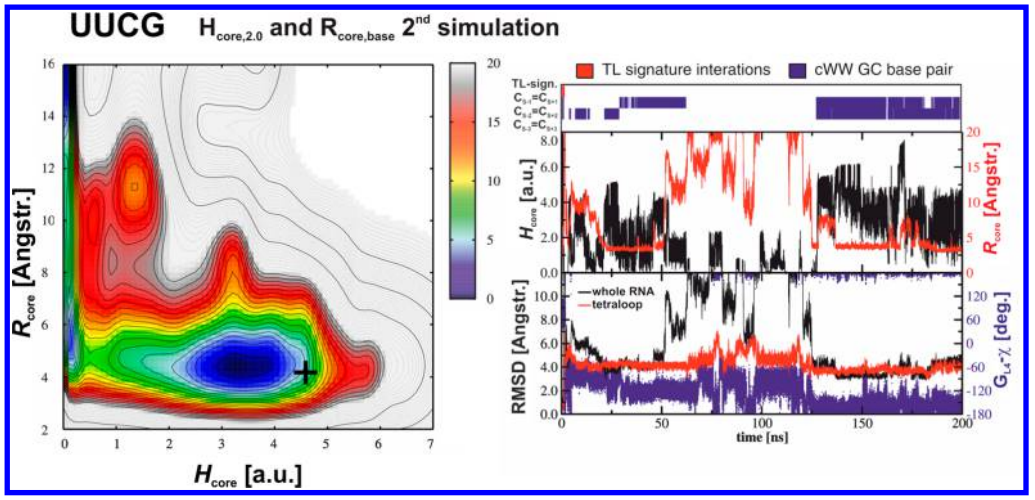


Figure 5. FES plot (left) and the development of key structural features (right) obtained from the second (G_{L4}-*anti* populating) 200-ns-long WT-MetaD simulation of UUCG TL using $H_{core,2.0}$ and $R_{core,base}$ CVs (see caption of Figure 4).

a second independent 200-ns-long UUCG WT-MetaD simulation, sharing the same initial structure and input settings as the first simulation including $H_{core,2.0}$ and $R_{core,base}$ CVs. In this run, the glycosidic bond of the G_{L4} nucleobase spontaneously flipped from the *syn* to *anti* state during the initial part of the simulation. The difference between these two WT-MetaD runs was solely due the coincidental occurrence of a rare event not accelerated by our CVs. However, once it had occurred, the G_{L4} nucleobase never flipped back to the *syn* orientation (see Figures 4 and 5). Similarly to the first UUCG WT-MetaD simulation, the second simulation with G_{L4} populating *anti*-orientation repeatedly visited the completely unfolded single-stranded states, as well as bent states with the formed A-RNA stem. However, the bent states structures could not be assigned to the native state owing to the highly distorted TL region, stemming largely from the G_{L4} in *anti*-orientation (Figure 5). Therefore, once the G_{L4} nucleobase had flipped from *syn* to *anti*, our WT-MetaD simulation sampled the unfolded-to-misfolded transitions and not the unfolding−refolding events.

Based on these results, we added the orientation of the glycosidic bond of the G_{L4} nucleobase into the set of CVs, specifically using the G_{L4} χ torsion angle as a third CV to accelerate the population of both G_{L4}-*syn* and G_{L4}-*anti* orientations (see the Supporting Information for two 400-ns-long WT-MetaD simulation using three CVs, $H_{core,2.5}$, $R_{core,base}$, and G_{L4}-χ). However, we found that these WT-MetaD simulations were even not able to sample the refolding events (Figure S3 in the Supporting Information), i.e., the addition of the G_{L4} glycosidic torsion as the third CV precluded the refolding events. This might be simply caused by the increased complexity of the CV space, i.e., due to its increased dimensionality; however, we consider this possibility as less likely. It seems that the observed repetitive unfolding/refolding events in the first G_{L4}-*syn* populating WT-MetaD simulations were allowed because the sugar−phosphate backbone of the loop region retained its structural memory, including the *syn*-orientation of the G_{L4} nucleobase, and the observed refolding events were facilitated by relatively simple conformational space of the loop within the framework of the G_{L4}-*syn* orientation. In other words, once the sampling over all orientations of G_{L4} was achieved, the conformational space of the loop became significantly more complex and the method was no more capable to sample unfolding−refolding events. Some more sophisticated CV guiding unfolding−refolding events would be required. The unfolding−refolding dynamics evidenced in the first WT-metaD run should thus rather be considered as a false-positive result. Although this conclusion might seem at first sight unsatisfactory, our data still provided a valuable piece of information. They indicated that formation of the loop is a crucial step in the folding of the entire UUCG hairpin and most likely represented a hidden variable in our WT-MetaD simulations. The exact nature of this hidden variable is, however, not known and its description by a limited set of straightforward CVs appears difficult. We thus conclude that the formation of the loop in case of UUCG TL must be more complicated than we initially assumed and involves additional important slow dynamics processes that are orthogonal to our CVs (even those extended by the G_{L4}-χ torsion). This should be also taken into consideration when planning and interpreting CV-based (un)folding studies of larger nucleic acids, namely those, where *syn*/*anti* dynamics might be important.

We tested an alternative selection of CVs including RMSD of the loop region following the scheme used for GAGA TL. We performed two additional 400-ns-long WT-MetaD simulations of UUCG TL using RMSD and $H_{core,2.5}$, or RMSD, $H_{core,2.5}$, and G_{L4}-χ CVs, i.e., with two or three CVs, respectively. We observed a rapid loss of the native structure during the initial phases of both these WT-MetaD simulations, however, in both cases the simulations again failed to sample the refolding events (see Figure S8 in the Supporting Information). Similarly as in the case of the two above-mentioned simulations (with $H_{core,2.5}$, $R_{core,base}$, and G_{L4}-χ CVs), the failure of refolding events sampling in the simulations with RMSD CV might be attributed to the fact that RMSD most likely accelerated sampling of the G_{L4} nucleotide's conformations including *syn*−*anti* transitions, but was not robust enough to keep the conformational variability of the UUCG loop fully under control. Taken together, the observed importance of the G_{L4} *syn*−*anti* dynamics in the present WT-MetaD simulations further supports the hypothesis that the orientation of the glycosidic bond of the G_{L4} nucleobase is crucial for folding of the UUCG TL. In particular, the G_{L4}-*anti* state may act as a kinetic trap along the folding pathway of the UUCG TL, which

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

is consistent with experimental kinetic measurements of G_{L4} 8-bromoguanisine[26,28] as well as our recent REMD simulations.[25]

As noted above, our WT-metaD simulations sampled several characteristic G_{L4}-misfolded states, namely G_{L4}-*anti* and G_{L4}-bulged states, which were observed also in our recent REMD simulations.[25] In order to test their stability, we performed unbiased simulations starting from these misfolded states. We found that both these misfolded states are stable (similarly to the native G_{L4}-*syn* state) on a microsecond time scale (see Supporting Information) and thus represent long-living substates.

In conclusion, proper sampling of *syn*−*anti* states represents a significant challenge for the theoretical description of the structural dynamics of nucleic acids due to its variational complexity. Thus, particular attention should be paid to this issue in metadynamics studies of nucleic acids where *syn*−*anti* flips are involved in the folding process, as exemplified for guanine quadruplex DNAs.[61] Lack of *syn*−*anti* sampling may easily produce false unfolding−refolding pathways in CV-based simulations with entirely unrealistic thermodynamics.

Insights into Mechanism of Unfolding of the TLs. As explained above, the CVs used in the WT-MetaD simulations did not achieve sufficiently repetitive unfolding and refolding events, and exhaustive sampling of unfolded states. Nevertheless, the WT-MetaD simulations were still able to capture some attributes of the structural changes along the unfolding and eventual refolding events. Therefore, despite the limitations of CV choice, WT-MetaD simulations might provide a valuable insight into the mechanism of folding. However, we would like to stress that the mechanisms of unfolding and refolding events observed by extended sampling methods should always be interpreted with care. Particularly, the refolding events observed by all CV-based methods might be highly dependent on the specific choice of CVs and thus might not share the same mechanism with the native process. The same most likely stands also for REMD simulations due to temperature effects. Nevertheless, in the subsequent paragraphs we show a qualitative agreement between these two fundamentally different methods, indicating that the computations provide some insight into the mechanism of folding.

GAGA TL. The mechanism of GAGA TL unfolding in the WT-MetaD comprised several steps, including opening of the terminal CG pairs via base pair fraying initiated by the 5′-terminal nucleotide, A-RNA stem opening, i.e., rupture of the remaining two base pairs of the stem, and, finally, breakage of the TL signature interactions. At the beginning of the first (200-ns-long) WT-metaD simulation (see above), the system rapidly lost the base pairing and part of the signature interactions within the TL region, but the hairpin kept the bent conformation. In particular, the $A_{L2}|G_{L3}|A_{L4}$ tripurine stack (Figure 2) and sharp bend between G_{L1} and A_{L2} remained in their native conformation. After ∼15 ns, both the base-pairing and signature interactions of the TL were reestablished to the native form and the simulations started to sample native, and near-native frayed or flipped-over states. Note that this was the only refolding event observed in this particular WT-MetaD simulation of GAGA, which, as mentioned above, was most likely facilitated by the near-native conformation of the loop backbone. At this stage, we observed fraying of the 5′-terminal cytosine and/or its flip-over, followed by stacking of this base to the stem by the opposite stacking face. This type of fluctuation was not surprising because 5′-end nucleobases are known to exhibit weaker stacking with the adjacent base pair than the 3′-end nucleobases. Such fraying of terminal nucleobases has also been observed in NMR experiments and MD simulations.[27,60,62] During this process, $R_{core,base}$ fluctuated around 5 Å and $H_{core,2.0}$ within 2−8 (Figure 2). Immediately after unstacking of the 3′-terminal base (besides the unstacking of the 5′-terminal base during the above-mentioned base fraying), the second and third base pairs (G_{S-2}=C_{S+2} and C_{S-1}=G_{S+1}) were destabilized, which resulted in loss of the helical stem. The TL region remained temporarily folded for few additional ns. Notably, the simulation sampled the state with native conformation of the TL region, including the *trans*-Hoogsteen sugar edge (*t*HS) A_{L4}-G_{L1} interaction containing G4(N2)···A7(*pro*-R_P) and G4(N2)···A7(N7) signature interactions, but with an already unfolded stem. After destabilization of the *t*HS A_{L4}-G_{L1} base pair, the system was completely unfolded. More interestingly, in the second WT-MetaD simulations, all observed refolding events were initiated from the straightened unfolded ssRNA by the formation of the loop region followed by zipping of the stem base pairs; i.e., the refolding path corresponded to the time-reversed version of the above-mentioned unfolding mechanism (Figure 3). It is interesting to note that the observed unfolding and refolding mechanisms resembled (reversal) of also the folding mechanism suggested by our recent REMD simulations (see Figure 5b in ref 25). Such agreement between two conceptually different enhanced sampling methods is encouraging and indicates that they capture at least some essential features of the (un)folding process of the GNRA TL.

UUCG TL. In contrast to GAGA, in the first WT-metaD UUCG TL simulation we observed reversible unfolding and refolding events with repeated full restoration of the native UUCG TL conformation. Again, the system rapidly lost the base-pairing and some of the signature interactions at the very beginning of the simulation but remained in a bent conformation. After ∼2 ns, both the base-pairing and signature interactions were re-established, except the base-pairing of the terminal base pair. Notably, the terminal base pair did not reestablish its canonical *c*WW base-pairing throughout the entire simulation. However, after ∼40 ns, the 3′-end guanine flipped over, changed the orientation of its glycosidic torsion from *anti* to *syn*, and, after ∼55 ns, stacked back to the stem with the opposite stacking face. It established a base pair with the 5′-end cytosine in *t*WH GC base-pairing family. This is consistent with recent simulation studies of base fraying of canonical helices.[60] In contrast to the GAGA TL, we also observed distortions of the TL region even in the states with a base-paired stem, which could be considered an alternative TL conformation under the force field description. This conformation was characterized by structural flipping of the U_{L1} nucleotide accompanied by repuckering from the native C3′-endo to C2′-endo form, flipping of the ε/ζ torsions of U_{L1} nucleotide, and consequent loss of signature interactions. The hairpin was repeatedly fully unfolded during the simulation. However, as noted above, despite the at first sight full unfolding and straightening of the hairpin to ssRNA lacking any bend, the glycosidic torsion of the G_{L4} guanine remained in its native *syn* orientation throughout the whole simulation (Figure 4). As explained above, this likely means that the system unfolded only to a state that was still preorganized for formation of the UUCG TL. In contrast, the second equivalent UUCG WT-MetaD simulation achieved flipping of the G_{L4} glycosidic bond at the beginning of the simulation to *anti*. However, this then prevented reestablishment of the native TL's conformation.

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

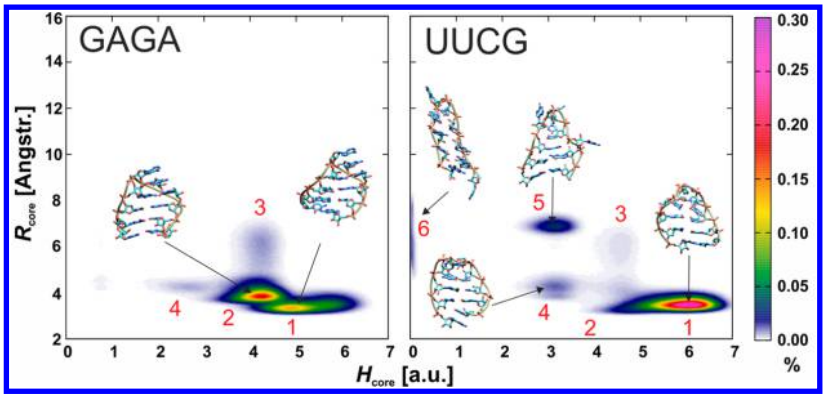


Figure 6. Density maps showing the population of structures sampled via $R_{core,base}$ vs $H_{core,2.0}$ over 15-μs-long unbiased MD trajectories of the GAGA and UUCG TLs. Red numbers indicate different regions of the sampled structures: structure 1, native state; structure 2, flipped-over state with different base-pairing of the terminal base pair; structure 3, fraying of terminal base pair; structure 4, stacking of the terminal base pair; structure 5, unfolded loop; and structure 6, fully unfolded structure.

Consistently with this finding, in a recent REMD study we found that *syn*−*anti* flipping of the G_{L4} nucleotide is the main kinetic hindrance to the formation of the UUCG TL.[25] This also agrees with the hypothesis that the *anti* orientation of G_{L4} makes a kinetic trap for the folding of UNCG TLs.[26,28] The relatively high tendency of the UUCG TL bearing *syn* G_{L4} to refold observed in the WT-MetaD simulations and, in turn, lack of refolding in the other UUCG WT-MetaD simulations with *anti* orientation of G_{L4} or with G_{L4} χ in the set of CVs, further support this hypothesis and illustrate some of the limitations of using WT-MetaD in studies of nucleic acid folding.

15-Microseconds-Long Standard MD Simulations. Although enhanced sampling methods, such as REMD or WT-MetaD simulations, can provide valuable insights into the conformational variability of RNA systems, classical unbiased MD simulations are still important for description of the stability and structural flexibility of these systems. We thus studied the structural dynamics of both hairpins using unbiased MD simulations on microseconds simulation time scales, namely we performed 15-μs-long MD simulations of the GAGA and UUCG TLs at 300 K with the all-atomic AMBER ff99bsc0χ_{OL3} force field. To the best of our knowledge, these simulations currently represent the longest time scale achievable in a single unbiased all-atom MD simulation of these RNA hairpins in explicit water. Like in our recent (shorter) simulations,[23] both TLs were stable on a time scale of a few microseconds. Namely both TLs fluctuated around their native states for more than 8 μs, suggesting that the native state corresponded to a relatively stable minimum in the utilized force field. During this part of the simulations, the behavior of both TLs was rather uniform: the loop and central section of the stem were stable, and the systems fluctuated around the starting structures. The terminal base-pair was the most flexible part of both hairpins. This base-pair passed through the native *c*WW structure (structure 1 in Figure 6), a frayed structure with unstacked 5′-terminal nucleobase (structure 3 in Figure 6), a flipped-over structure, where the 5′-terminal nucleotide was flipped over with *syn*-orientation and stacked back to the stem by the opposite stacking face (structure 2 in Figure 6), and a stacked structure, where the 3′- and 5′-terminal nucleotides were stacked to each other rather than base-paired (structure 4 in Figure 6). See ref 60 for a discussion about the force field description of base-pair fraying. The stable parts of the simulations were followed by parts where structural distortions of the stem-loop structures occurred in both systems (Figures S7 and S8 in the Supporting Information).

GAGA TL. The TL region of GAGA was distorted at ∼8.2 μs (structure 5 in Figure 6). After ∼8.5 μs, the structure of the loop temporarily returned back near to the native structure. However, after this transient formation of the native structure, the loop became distorted and flexible until ∼10.3 μs. Afterward, the hairpin started to fold back to the native state and beyond ∼12.1 μs fluctuated around the native conformation with a well-defined stem and all signature interactions of the TL region (Figure S4 in Supporting Information). In other words, we observed spontaneous (partial) unfolding and refolding events in classical unbiased MD simulations on a few microsecond time scale. Such relatively rapid refolding and reestablishment of the stable native conformation in the remaining part of the simulation suggests that the observed distortion of the TL after ∼8 μs may represent the native breathing of the RNA hairpin.

UUCG TL. In contrast, the UUCG TL showed complete melting of both the loop and stem regions after ∼8.3 μs, where base-pairing and all signature interactions were completely lost until the end of the simulation (structure 6 in Figure 6). Although after ∼10.2 μs two terminal base pairs of the stem were reestablished, the loop and base pair adjacent to the loop remained unfolded (Figure S5 in Supporting Information). It should be noted that the melting at ∼8.3 μs was accompanied by *syn*/*anti* flipping of the G_{L4} nucleobase (Figure S6 in Supporting Information). Taking into account the hypothesis that the G_{L4}-*anti* state forms a kinetic trap (supported by folding measurements on a G_{L4} 8-bromoguanisine mutant[26,28] and recent REMD simulations[25]), it is not surprising that the unfolded UUCG TL was not able to refold back in the remaining few microseconds. It thus seems that full description of the coupling between unfolding/refolding events and G_{L4} *syn*/*anti* reorientation would require significantly more robust sampling.

CONCLUSIONS

In this study, we investigated two biologically important RNA hairpin loops using a series of hundreds of nanoseconds long WT-MetaD simulations, which allowed enhanced sampling of the conformational space along the selected CVs using memory-dependent biased potentials. The biased simulations were complemented by 15-μs-long unbiased MD simulations,

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

representing the cutting edge of currently accessible time scales for explicit solvent simulations of RNA.

Despite the insightful and careful choice of CVs, our WT-MetaD simulations achieved only semiquantitative convergence. Specifically, we observed neither reversible folding/unfolding events with sufficient number of interconversions during the simulations nor full sampling of the unfolded state. Consequently, the constructed FESs were sampled with rather qualitative accuracy, which hampered thorough analysis of the folding energetics because it would inevitably require a convergent FES with thermochemical accuracy of ∼1 kcal/mol (or even better) and lack of trapping the systems in the minima of the hidden variables. Thus, it seems that some more sophisticated choice of CVs would be required to achieve qualitative convergence sufficient for thermodynamic estimates. These findings indicate that the theoretical description of RNA folding thermodynamics still poses a challenge for computational chemistry, even for rather small RNA structures.

Nevertheless, both the WT-MetaD simulations and extensive unbiased MD simulations provided new insights into the conformational behavior of the UUCG and GAGA TLs, as well as information about the performance of the used ff99bsc0χ_{OL3} force field. We found that the GAGA TL unfolded in our WT-MetaD simulations via the same mechanisms as observed in our recent REMD simulations.[25] It should be noted that the GAGA TL remained stable on a 10 μs time scale, and we observed spontaneous unfolding/refolding events within such unbiased MD simulations. This may indicate that both the simulation setup and force field used adequately represented the structural features of the GNRA TLs.

In the case of the UUCG TL, we confirmed a crucial role of the orientation of G_{L4} nucleobase around its glycosidic bond in both unbiased MD and WT-MetaD simulations. The simulation data strongly supported the hypothesis that the G_{L4}-*anti* conformation represents a kinetic trap along the UUCG TL folding pathway. The WT-MetaD simulations suggested that once the G_{L4} was in *syn* orientation, the force field described the native state as a significant free energy minimum that was comparable in depth to the unfolded state in the CV space. However, the simulations did not allow the free energies of states with different G_{L4} orientations to be merged.

At the end, we have to admit that the present results can also be affected by the force field limitations. We were unfortunately not able to fully distinguish whether the rare occurrence of refolding events was caused by not enough robust sampling or by the low stability of the folded state in the used force field. Recently, Cheatham and co-workers pointed out both issues, i.e., sampling as well as force field, as the most challenging points in the simulations of even small RNA systems.[63] Our simulations provide some hints to the force field performance. More specifically, the FES of GAGA TL clearly revealed the native state as at least the local free energy minimum and suggested that the single stranded unfolded state might be to some extent overstabilized by the force field. A similar indication emerged also from our recent REMD simulations.[25] As we pointed out elsewhere, although the recent refinements of the backbone dihedral potentials definitely improved the performance of the nucleic acids force fields by eliminating several substantial imbalances, the force fields necessarily remain far from being perfect.[33,64] It would be entirely naive to assume that modifying the backbone torsional potentials could be sufficient to obtain a universally satisfactory nucleic acids force field. In fact, comparisons with rigorous quantum chemical calculations even suggest that the pair-additive approximation with conformation-independent parameters is, in principle, not flexible enough to allow parametrization of fully balanced description of the potential energy surface of nucleic acids.[33] In other words, even upon a complete reparameterization of all energy terms within the framework of the current force field model, we are unlikely to be able to derive a truly accurate nucleic acids force field.

■ ASSOCIATED CONTENT

Supporting Information

Details of the WT-MetaD convergence, the population of structures sampled in WT-MetaD simulations, the second UUCG WT-MetaD simulation, and WT-metaD simulations with alternatively chosen CVs; further analyses of classical 15-μs-long MD simulations. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00010.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: pavel.hobza@uochb.cas.cz. Tel.: +420 220 410311.
*E-mail: michal.otyepka@upol.cz. Tel.: +420 585634756.

Author Contributions

#S.H. and P.K. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grant P208/12/1878 (J.S., M.O., P.K.) from the Czech Science Foundation. This work was further supported by Operational Program Research and Development for Innovations−European Regional Development Fund (project CZ.1.05/2.1.00/03.0058), the Operational Program Education for Competitiveness−European Social Fund (CZ.1.07/2.3.00/20.0058, CZ.1.07/2.3.00/20.0155), and project LO1305 of the Ministry of Education, Youth and Sports of the Czech Republic (M.O., P.B., P.K.), and by Student Project IGA_PrF_2015_027 of Palacký University (M.O., P.B., P.K.). This work was also part of the Research Project RVO:61388963 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and was supported by Czech Science Foundation (P208/12/G016). (S.H., P.H.).

■ REFERENCES

(1) Woese, C. R.; Winker, S.; Gutell, R. R. Architecture of Ribosomal-RNA - Constraints on the Sequence of Tetra-Loops. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8467−8471.

(2) Zheng, M. X.; Wu, M.; Tinoco, I. Formation of a GNRA Tetraloop in P5abc Can Disrupt an Interdomain Interaction in the Tetrahymena group I Ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 3695−3700.

(3) Radhakrishnan, R. Coupling of Fast and Slow Modes in the Reaction Pathway of the Minimal Hammerhead Ribozyme Cleavage. *Biophys. J.* **2007**, *93*, 2391−2399.

(4) Wedekind, J. E.; McKay, D. B. Crystallographic Structures of the Hammerhead Ribozyme: Relationship to Ribozyme Folding and Catalysis. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 475−502.

(5) Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A. Structural Investigation of the GlmS Ribozyme Bound to its Catalytic Cofactor. *Chem. Biol.* **2007**, *14*, 97−105.

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

(6) Klein, D. J.; Ferre-D'Amare, A. R. Structural Basis of glmS Ribozyme Activation by Glucosamine-6-phosphate. *Science* **2006**, *313*, 1752−1756.
(7) Fujita, Y.; Tanaka, T.; Furuta, H.; Ikawa, Y. Functional Roles of a Tetraloop/Receptor Interacting Module in a Cyclic di-GMP Riboswitch. *J. Biosci. Bioeng.* **2012**, *113*, 141−145.
(8) Serganov, A.; Huang, L. L.; Patel, D. J. Structural Insights into Amino Acid Binding and Gene Control by a Lysine Riboswitch. *Nature* **2008**, *455*, 1263−U76.
(9) Smith, K. D.; Lipchock, S. V.; Livingston, A. L.; Shanahan, C. A.; Strobel, S. A. Structural and Biochemical Determinants of Ligand Binding by the c-di-GMP Riboswitch. *Biochemistry* **2010**, *49*, 7351−7359.
(10) Belanger, F.; Gagnon, M. G.; Steinberg, S. V.; Cunningham, P. R.; Brakier-Gingras, L. Study of the Functional Interaction of the 900 Tetraloop of 16 S Ribosomal RNA with Helix 24 within the Bacterial Ribosome. *J. Mol. Biol.* **2004**, *338*, 683−693.
(11) Correll, C. C.; Swinger, K. Common and Distinctive Features of GNRA Tetraloops Based on a GUAA Tetraloop Structure at 1.4 Angstrom Resolution. *RNA* **2003**, *9*, 355−363.
(12) Sahu, B.; Khade, P. K.; Joseph, S. Functional Replacement of Two Highly Conserved Tetraloops in the Bacterial Ribosome. *Biochemistry* **2012**, *51*, 7618−7626.
(13) Westhof, E.; Leontis, N. Atomic glimpses on a billion-year-old molecular machine. *Angew. Chem., Int. Ed.* **2000**, *39*, 1587−1591.
(14) Fiore, J. L.; Nesbitt, D. J. An RNA folding motif: GNRA tetraloop-receptor interactions. *Q. Rev. Biophys.* **2013**, *46*, 223−264.
(15) Varani, G.; Cheong, C.; Tinoco, I., Jr. Structure of an unusually stable RNA hairpin. *Biochemistry* **1991**, *30*, 3280−3289.
(16) Jaeger, L.; Michel, F.; Westhof, E. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **1994**, *236*, 1271−1276.
(17) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. Single nucleotide RNA choreography. *Nucleic Acids Res.* **2006**, *34*, 1481−1491.
(18) Allain, F. H.; Varani, G. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **1995**, *250*, 333−353.
(19) Correll, C. C.; Swinger, K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA* **2003**, *9*, 355−363.
(20) Heus, H. A.; Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **1991**, *253*, 191−194.
(21) 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.
(22) Leontis, N. B.; Westhof, E. Analysis of RNA motifs. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300−308.
(23) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(24) Chen, A. A.; Garcia, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820−16825.
(25) Kührová, P.; Banas, P.; Best, R. B.; Sponer, J.; Otyepka, M. Computer Folding of RNA Tetraloops? Are We There Yet? *J. Chem. Theory Comput.* **2013**, *9*, 2115−2125.
(26) Ma, H. R.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530.
(27) Mohan, S.; Hsiao, C.; Bowman, J. C.; Wartell, R.; Williams, L. D. RNA Tetraloop Folding Reveals Tension between Backbone Restraints and Molecular Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 12679−12689.
(28) Proctor, D. J.; Ma, H. R.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. Folding thermodynamics and kinetics of YNMG RNA hairpins: Specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry* **2004**, *43*, 14004−14014.
(29) Zhang, W. B.; Chen, S. J. Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states. *Biophys. J.* **2006**, *90*, 778−787.
(30) Chakraborty, D.; Collepardo-Guevara, R.; Wales, D. J. Energy landscapes, folding mechanisms and kinetics of RNA tetraloop hairpins. *J. Am. Chem. Soc.* **2014**, *136*, 18052−18061.
(31) Bottaro, S.; Di Palma, F.; Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res.* **2014**, *42*, 13306−13314.
(32) Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. Extensive Molecular Dynamics Simulations Showing That Canonical G8 and Protonated A38H(+) Forms Are Most Consistent with Crystal Structures of Hairpin Ribozyme. *J. Phys. Chem. B* **2010**, *114*, 6642−6652.
(33) Sponer, J.; Mladek, A.; Sponer, J. E.; Svozil, D.; Zgarbova, M.; Banas, P.; Jurecka, P.; Otyepka, M. The DNA and RNA sugar-phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15257−15277.
(34) 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 (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996**, *118*, 2309−2309.
(35) Zgarbová, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, 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.
(36) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(37) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wires Comput. Mol. Sci.* **2011**, *1*, 826−843.
(38) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, No. 020603, DOI: 10.1103/PhysRevLett.100.020603.
(39) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(40) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *J. Chem. Theory Comput.* **2012**, *8*, 2189−2192.
(41) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2013**, *16*, 163−199.
(42) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. Theoretical Perspectives on Protein Folding. *Annu. Rev. Biophys.* **2010**, *39*, 159−183.
(43) Ditzler, M. A.; Rueda, D.; Mo, J. J.; Hakansson, K.; Walter, N. G. A Rugged Free Energy Landscape Separates Multiple Functional RNA Folds Throughout Denaturation. *Nucleic Acids Res.* **2008**, *36*, 7088−7099.
(44) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(45) Correll, C. C.; Beneken, J.; Plantinga, M. J.; Lubbers, M.; Chan, Y. L. The common and the distinctive features of the bulged-G motif based on a 1.04 angstrom resolution RNA structure. *Nucleic Acids Res.* **2003**, *31*, 6806−6818.
(46) Nozinovic, S.; Furtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* **2010**, *38*, 683−694.
(47) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499−512.

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877

(48) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.* **2009**, *37*, 4898−4918.
(49) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(50) Aaqvist, J. Ion Water Interaction Potentials Derived from Free-Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(51) Bussi, G.; Zykova-Timan, T.; Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* **2009**, *130*, No. 074101.
(52) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(53) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(54) Limongelli, V.; De Tito, S.; Cerofolini, L.; Fragai, M.; Pagano, B.; Trotta, R.; Cosconati, S.; Marinelli, L.; Novellino, E.; Bertini, I.; Randazzo, A.; Luchinat, C.; Parrinello, M. The G-Triplex DNA. *Angew. Chem., Int. Ed.* **2013**, *52*, 2269−2273.
(55) Piana, S.; Laio, A. A bias-exchange approach to protein folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(56) Bian, Y.; Tan, C.; Wang, J.; Sheng, Y.; Zhang, J.; Wang, W. Atomistic picture for the folding pathway of a hybrid-1 type human telomeric DNA G-quadruplex. *PLoS Comput. Biol.* **2014**, *10*, e1003562.
(57) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247−2254.
(58) Kuehrova, P.; Otyepka, M.; Sponer, J.; Banas, P. Are Waters around RNA More than Just a Solvent? - An Insight from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 401−411.
(59) Yang, X. J.; Gerczei, T.; Glover, L.; Correll, C. C. Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping. *Nat. Struct. Biol.* **2001**, *8*, 968−973.
(60) Zgarbova, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177−3189.
(61) Stadlbauer, P.; Trantirek, L.; Cheatham, T. E.; Koca, J.; Sponer, J. Triplex intermediates in folding of human telomeric quadruplexes probed by microsecond-scale molecular dynamics simulations. *Biochimie* **2014**, *105C*, 22−35.
(62) Colizzi, F.; Bussi, G. RNA Unwinding from Reweighted Pulling Simulations. *J. Am. Chem. Soc.* **2012**, *134*, 5173−5179.
(63) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492−499.
(64) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771−1782.

DOI: 10.1021/acs.jctc.5b00010
J. Chem. Theory Comput. 2015, 11, 3866−3877