JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Computer Folding of RNA Tetraloops? Are We There Yet?

Petra Kührová,[§] Pavel Banáš,*[,§,†] Robert B. Best,[‡] Jiří Šponer,[†,∥] and Michal Otyepka*[,§,†]

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

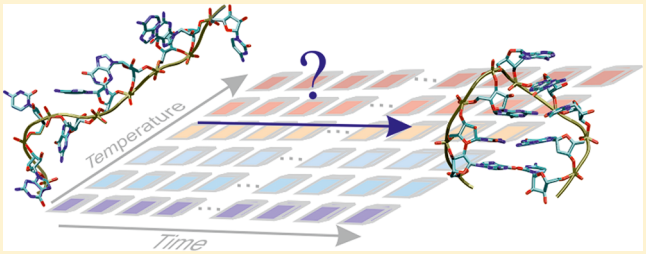
[‡]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, United States

[†]Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

[∥]CEITEC − Central European Institute of Technology, Masaryk University, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

Supporting Information

ABSTRACT: RNA hairpin loops represent important RNA motifs with indispensable biological functions in RNA folding and tertiary interactions, with the 5′-UNCG-3′ and 5′-GNRA-3′ families being the most abundant. Molecular dynamics simulations represent a powerful method to investigate the structure, folding, and function of these tetraloops (TLs), but previous AMBER force fields were unable to maintain even the native structure of small TL RNAs. Here, we have used Replica Exchange Molecular Dynamics (REMD) with our recent reparameterization of AMBER RNA force field to study the folding of RNA hairpins containing representatives UNCG and GNRA TLs. We find that in each case, we are able to reach conformations within 2 Å of the native structure, in contrast to results with earlier force fields. Although we find that the REMD simulation runs of a total of ∼19 μs (starting from both folded and unfolded state) in duration for each TL are still far from obtaining a representative equilibrium distribution at each temperature, we are nonetheless able to map the stable species on the folding energy landscape. The qualitative picture we obtain is consistent with experimental studies of RNA folding in that there are a number of stable on- and off-pathway intermediates en route to the native state. In particular, we have identified a misfolded-bulged state of GNRA TL, which shares many structural features with the X-ray structure of GNRA TL in the complex with restrictocin, namely the bulged out A_{L4} base. Since this is the same conformation observed in the complex of the TL with restrictocin, we argue that GNRA TL is able to bind restrictocin via a "conformational selection" mechanism, with the R_{L3} and A_{L4} bases being exposed to the solvent prior to binding. In addition we have identified a misfolded-anti state of UUCG TL, which is structurally close to the native state except that the G_{L4} nucleotide is in an *anti*-conformation instead of the native *syn*. Our data suggest that the UUCG misfolded-anti state may be a kinetic trap for the UUCG folding.



INTRODUCTION

Tetraloops (TLs), containing four bases (henceforth L_1-L_4) in the loop, belong to functionally important RNA hairpin loops, which usually cap canonical RNA helices. TLs enable the reversal of backbone path direction required for formation of secondary and tertiary structures of RNA.[1−5] Among all possible combinations of TLs,[6] the YNMG and GNRA (Y stands for pyrimidine, N for any nucleotide, M for adenine or cytosine, and R for purine) families are the most abundant.[4] According to experimental structures, these TL families have precisely defined folds with a set of signature backbone topological features and intraloop tertiary molecular interactions, whose conservation determines their consensus sequences. They are conserved autonomous RNA motifs (i.e., they fold independently of their context) and are exceptionally thermodynamically stable.[7] TLs are biologically significant as they initiate the folding of RNA structures[4,5,8] and present important interaction sites for tertiary contacts to RNA or to proteins.[9−11]

The GNRA and UNCG (subfamily of the YNMG family) TLs have been suggested to be at least in some structural context evolutionary interchangeable,[12] and their structures share many similarities. For instance, the first (L_1) and fourth (L_4) nucleotides form a noncanonical base-pair, whereas the two remaining bases (L_2 and L_3) are unpaired. The sharp backbone turn occurs between the first and second nucleotides. Both UNCG and GNRA TLs display a striking degree of structural conservation across a wide range of experimental structures, including their multiple occurrences in large RNAs and ribonucleoprotein particles.[6,13−25] However, there are also differences in biological function of UNCG and GNRA TLs. The GNRA TLs are often involved in tertiary RNA−RNA and

Received: December 11, 2012

   dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

RNA-protein interactions and thus participate in a final shaping of overall RNA folds. In contrast, the UNCG TLs do not bind natural ligands except of cations and are rarely involved in tertiary RNA/RNA interactions. The UNCG TLs are known to initiate RNA folding and are therefore considered to be RNA "foldons".[2]

The structural adaptations in complexes of GNRA TL with receptor typically include changes of the TL receptor, while the TL itself remains unchanged.[26] There is, however, at least one exception, the GNRA TL complex with a ribotoxin restrictocin, where the GNRA structure is altered.[27] This finding raises the question of whether the restrictocin induces the conformation change or binds to a transiently unstructured GNRA TL.

Because of their biological relevance, and also their small size, RNA hairpins (including TLs) have often been used as model systems in simulation and theory[28−37] as well as in experiments.[38−41] Based on the experimental data, both two-state and multistate folding mechanisms have been suggested. The two-state folding hypothesis emerges primarily based on thermodynamic measurements,[38,40,42] whereas the multistate hypothesis is based on kinetic measurements.[38,43] Finally, parallel folding pathways were also suggested by Ansari et al.[40,41] It is quite possible that the different suggested scenarios reflect the difference in the way how the different experimental techniques monitor the folding process.

In several simulation studies, it has proved possible to reach native-like hairpin structures.[28−31,44−50] However, in many simulations, the TLs were found to be only marginally stable and exhibited large fluctuations around the native structure. Not all signature features of the native TLs have been reproduced. This appears to be at odds with atomic resolution experimental structures, which generally show the RNA hairpins to be very well-defined. Both UNCG and GNRA TLs belong to so-called RNA motifs, i.e., autonomously folding RNA building blocks that form precise 3D structures with a set of strictly conserved noncanonical interactions, determining their sequence signatures.[51,52] This discrepancy suggests that simulation studies may reflect remaining deficiencies in the force field parameters. Historically, force field development has primarily focused on providing a correct representation of regular RNA (or DNA) helices, with less focus on "non-canonical" structures which may be more challenging to describe.[53−55] Indications of such deficiencies are the instability of native structures in long equilibrium simulations and the failure of enhanced sampling methods such as REMD to find the exact native structure. For example, we have recently demonstrated a progressive loss of native structure in long simulations of both RNA hairpins with essentially all earlier versions of the AMBER force field.[56] We have shown that, in general, RNA simulations are affected by the quality of force fields considerably more than previously assumed, and in some cases evidently non-native[56,57] or even entirely degraded (including A-RNA duplexes)[56,58,59] structures may appear in MD simulations as the global free-energy minima. When the simulations are short, the force field artifacts may remain hidden, but in the case of longer time scales or using enhanced sampling techniques such as Replica Exchange Molecular Dynamics (REMD), the force field inaccuracies become evident.

Thus, a large effort has been made to refine the force field parameters for RNA simulations.[56,58,60−62] Recently, the new Olomouc parmχ_{OL3} (χ_{OL3}) correction[56,60] to the χ glycosidic torsion of the Cornell et al. ff99 force field[63] has been shown to be a promising force field for RNA hairpin TLs simulations, especially when combined with the earlier Barcelona parmbsc0 (bsc0) correction[64] to α/γ torsions. This bsc0χ_{OL3} version is currently the default AMBER RNA force field providing rather stable simulations of TLs on microseconds time scale.[56] It should be noted, that the force field has been extensively tested on other RNA systems like canonical A-RNA duplexes,[59] reverse kink-turns,[65] or the small preQ riboswitch aptamer domain.[66] Although we obviously do not claim that the force field is perfect, it is a visible improvement compared to the earlier force field versions.

In the present study we investigate the folding and unfolding of two representative TLs (UUCG and GAGA) using REMD[67] in combination with the latest *ff*99bsc0χ_{OL3} force field. Although our REMD simulations are clearly not fully converged, despite a length of ∼0.5 μs per replica, we were able to obtain an outline of the important features of the folding free energy landscape. The *ff*99bsc0χ_{OL3} is able to correctly capture the native fold of both TLs as a global free energy minimum, with a marked reduction of flexibility. Furthermore, in REMD simulations starting from completely unfolded initial conditions, we reach structures within 2 Å of the native state, considerably closer than reported in earlier studies.[29,30] We also find a number of native-like intermediate structures on pathway to the native state, consistent with interpretations of kinetic experiments. Several of these structures are very close to experimental structures of the TLs in complex with other molecules, suggesting that these molecules may bind the RNA via a conformational selection mechanism.

METHODS

Starting Structures − The Folded State. The starting structure of UUCG TL was taken from the high-resolution NMR structure (PDB ID: 2KOC).[68] To form a stable 10 residue hairpin, we modified the AU base-pair present in the NMR structure to GC and removed the two terminal base-pairs to yield a final the cgcUUCGgcg sequence. The UUCG structure is tightly packed, forming a very stable core (U_{L1}, C_{L3}, G_{L4}) with U_{L2} base exposed to the solvent and the *trans* Watson−Crick/Sugar-Edge (*t*WS) G_{L4}/U_{L1} base-pair. The backbone turn occurs between the U_{L1} and U_{L2}, and it is stabilized by the base-phosphate interaction between the C_{L3} amino group and U_{L2} phosphate (C_{L3}(N4)···U_{L2}(*pro*-R_p), a type 7BPh interaction[69]). Moreover, there are three other UUCG signature H bonds: U_{L1}(O2′)···G_{L4}(O6), G_{L4}(N1)···U_{L1}(O2), and U_{L2}(O2′)···G_{L4}(N7). The C_{L3} base is stacked on the U_{L1} base. The U_{L2} and C_{L3} nucleotides adopt a C2′-endo sugar conformation, extending the backbone to help bridge the stem. Notably, the glycosidic torsion of the fourth nucleotide (G_{L4}) is in *syn* conformation

The GAGA TL was derived from the high resolution X-ray structure determined at 1.04 Å resolution of the sarcin-ricin loop from *Escherichia coli* 23S rRNA (PDB ID 1Q9A, residues 2658−2663)[19] and capped by two additional CG base-pairs, so that the sequence of the hairpin was cgcGAGAgcg. The structure of GAGA TL is stabilized by the presence of three H-bonds: G_{L1}(N2)···A_{L4}(*pro*-R_p) (3BPh interaction, which might be altered by 4BPh interaction during the simulation characterized by G_{L1}(N1/N2)···A_{L4}(*pro*-R_p) bifurcated H-bonds), G_{L1}(N2)···A_{L4}(N7), and G_{L1}(O2′)···G_{L3}(N7). The A_{L2}, G_{L3}, and A_{L4} bases form a purine triple base stack, and G_{L1} is base-paired with A_{L4} by the *trans* Hoogsten/Sugar-Edge

(*t*HS) A_{L4}/G_{L1} pattern. For a more detailed description of signature interactions of both TLs see our preceding paper Banas et al. and Figures 1 and 2 therein.[56]

Starting Structures – The Unfolded State. The unfolded conformations of both TLs were chosen from classical molecular dynamics simulations at constant volume at 800 K. RMSDs of initial structures of both TLs were at least 9 Å from the native structure.

Simulation Protocol. The REMD and Thermodynamic Integration (TI) simulations were carried out using the recent reparameterization of the AMBER force field for RNA - *ff*99bsc0χ_{OL3},[56,60] which has been thoroughly tested[56,60,65] and seems to result in the most stable folded structures among the available RNA force fields.[59,60,65,66] The system was solvated using a rectangular box of TIP3P waters with a minimum distance between the box walls and solute of 15 Å. The charge of the system was neutralized by adding Na^+ counterions (with radius 1.868 Å and well depth 0.0028 kcal/mol)[70] modeling minimal salt conditions (concentration of Na^+ ions ~0.2 M). The system was built in the tleap module of the AMBER simulation package.[71] In the case of REMD simulations, the system was converted using the Acpype tool[72] to Gromacs,[73,74] while the AMBER package was used for TI.[71] Prior to the production phase of REMD and TI, each TL was prepared using the following protocol: The TL was constrained, and the solvent molecules with counterions were allowed to move during a 1,000-step minimization followed by 10-ps-long MD run under $[NpT]$ conditions (p = 1 atm, T = 300 K). Then, the solute was relaxed through several minimization steps, with decreasing force constants applied to the backbone atoms. After the relaxation, each system was heated to 300 K within 100 ps. The PME method was used for treating electrostatic interactions, and all simulations except REMD production phases (see below) were performed under periodic boundary conditions in the $[NpT]$ ensemble at 300 K and 1 atm using a 2 fs integration step. The SHAKE algorithm with a tolerance of 10^{-5} Å was used to fix positions of all hydrogen atoms. A 10.0 Å cutoff was applied to nonbonded interactions, and coordinates were stored every 10 ps. The equilibrated TLs structures were chosen as starting structures for REMD and TI simulations.

REMD Setup. The temperatures of 48 replicas spanned a range of 270−546.5 K for GAGA and 271−486 K for UUCG, respectively. The simulation time for both TLs was 400 ns (for a total of over of ~19.2 μs of sampling per each TL). The temperatures of replicas, which slightly differ between both TLs, were chosen to maintain an exchange rate of ~20%. The actual exchange rate over the temperature range was between 21 and 30% during all REMD simulations (for more details see Supporting Information, Table S1−S2 and Figures S1−S6). All REMD simulations were performed at constant volume, i.e., using the $[NVT]$ ensemble in each replica, with long-range electrostatics calculated using PME with 1.2 Å grid spacing and 9 Å real-space cutoff. The system was propagated using Langevin dynamics with friction coefficient of 1 ps^{-1}, and replica exchange was attempted every 10 ps.

While the first simulation was run from the folded state, we ran also REMD simulations from unfolded UUCG and GAGA TLs with the same setup as in simulations started from the native state for at least 400 ns. In total we accumulated 38.4 μs of simulation time for each TL.

Thermodynamic Integration Setup. The Thermodynamic Integration (TI) simulations (details of methodology reviewed in ref 75) were used to estimate the free energy difference between native and misfolded-anti states of the UUCG TL having the glycosidic torsion of the G_{L4} in *syn*- and *anti*-conformation, respectively (see below). The NMR structure of the UUCG TL was used as a starting structure of the starting (disappearing) state, while a representative REMD snapshot of the misfolded-anti state was used as the initial structure of the ending (appearing) state. The system was split into two parts: (i) the appearing/disappearing part involving the U_{L1}, C_{L3}, and G_{L4} nucleotides (namely part of the strand between U_{L1}(C5′) and U_{L2}(C5′) and between C_{L3}(C5′) and G_{10}(C5′) atoms) that was described by unique coordinates in both starting and ending states and (ii) the rest of the system representing the common part, whose coordinates and velocities were shared by both states. In order to prepare the starting structures, the snapshot of the misfolded-anti state was superimposed with the native state structure and the disappearing/appearing atoms were pasted from misfolded-anti to the native structure, so that the rest of the systems share identical atomic coordinates. The disappearing/appearing atoms were described by soft-core van der Waals and soft-core electrostatics potentials. The α and β parameters corresponding to the van der Waals and electrostatics soft-core terms were set to 0.5 and 12 $Å^2$, respectively. The Hamiltonians of disappearing and appearing states were coupled by mixing parameter λ using the linear mixing rule. Nine-point Gaussian quadrature was used for the integration. Thus nine simulations with different λ values were performed, each involving 100 ps heating followed by a 5-ns-long production phase.

Data Analysis. The trajectories output from REMD simulations were sorted according to single temperatures. We also constructed continuous trajectories obtained by following each replica through the exchanges in temperature space.[73] The trajectories were analyzed with the Gromacs package[76,77] and the *Ptraj* module of the AMBER package.[71] The structural parameters of systems were monitored using the X3DNA program (http://rutchem.rutgers.edu/~xiangjun/3DNA/). The simulations were visualized using a molecular visualization program VMD.[78]

We used the number of native base-pairs and the RMSD distance of all atoms of the L_1-L_4 nucleotides as collective variables to characterize the overall structural states of TLs: (a) *the folded state* contains all native base-pairs (except the terminal GC base-pair of stem, for which we assume that it can be temporarily unpaired due to base fraying) and has the all-atom RMSD of TL < 2 Å, (b) *the misfolded state* contains all native GC base-pairs in stem (except the terminal base-pair, see above), while the L_1/L_4 pair is not in its native pattern, and (c) *the unfolded forms* do not have any native base-pairs. The cluster analysis was carried out using an *in-house* script. Clusters were obtained by grouping conformations for all replicas. We checked that the structures that meet the above-mentioned criteria for the folded state also involve all signature interactions listed above and thus can be considered as biologically relevant native states.

RESULTS AND DISCUSSION

We carried out REMD simulations of two 10-nucleotide RNA hairpins containing UUCG and GAGA TLs in the TIP3P explicit water model using the recent *ff*99bsc0χ_{OL3} AMBER force field. The aim is to map conformational space and to analyze thermodynamics and mechanism of folding. The REMD algorithm runs multiple isothermal MD simulations in parallel at a sequence of increasing temperatures and attempts

dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

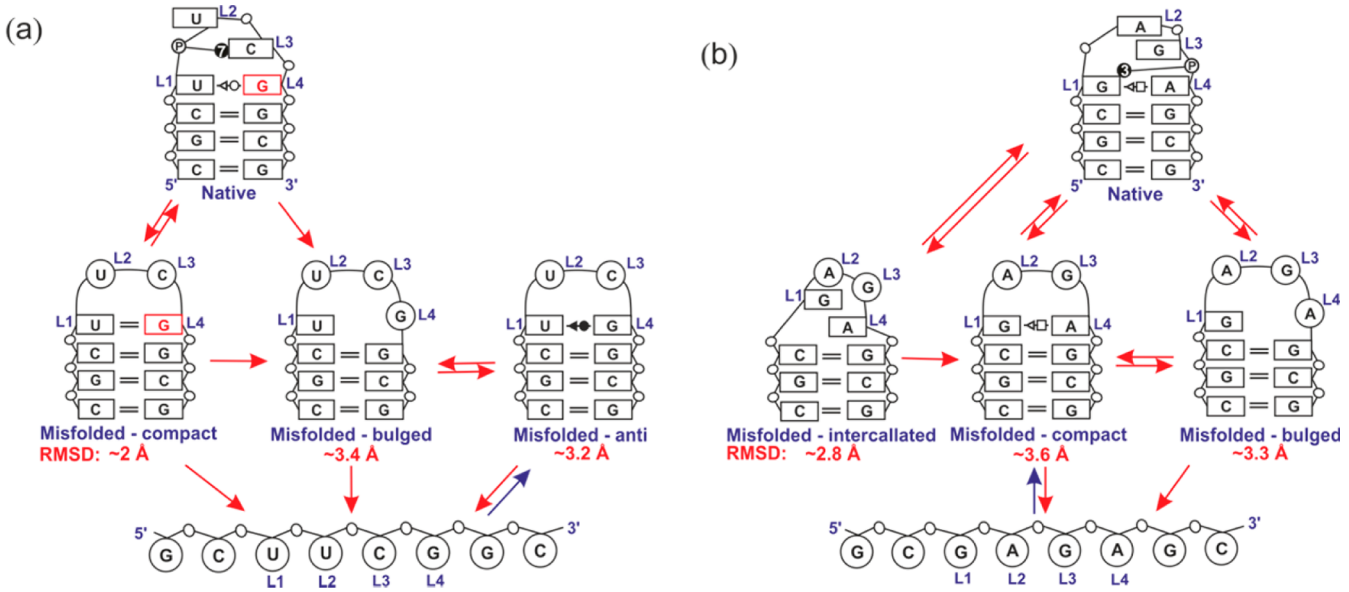


Figure 1. Conformational transition maps between folded, misfolded, and unfolded states of the investigated UUCG (a) and GAGA (b) tetraloops (TLs). The states are depicted as secondary structure schemes with base-pair and base−phosphate interactions annotated according to the standard classification.[69,83] Arrows between clusters indicate conformational transitions observed during the REMD simulations. The color of arrows indicate frequency of transitions: red corresponds to transitions frequently observed in our REMD simulations and blue stands for rare but still observed transitions (see Figure S7 for more information). Each misfolded state is accompanied by the average RMSD from the native state (shown in red); note that all misfolded states have RMSd below 4 Å. The loop residues are labeled as L_1−L_4. G_{L4} of the UUCG TL having *syn* orientation is shown in red. The nucleotides in circle are flexible, and their positions are not well-defined compared to the other nucleotides in the TL.

to exchange the structures between neighboring temperatures at regular time intervals. The probability of acceptance of these exchanges is determined by Metropolis probability criterion

$$P(i \leftrightarrow j) = \min(1, e^{-(\beta_j - \beta_i)(U_i - U_j)})$$

where P is the probability of an exchange between two neighboring replicas i and j, $\beta_i = 1/k_B T_i$ and $\beta_j = 1/k_B T_j$ where k_B is Boltzmann constant, T_i and T_j are the temperatures, and U_i and U_j are the potential energies of replicas i and j, respectively. REMD is a powerful computational technique offering more effective sampling than plain MD simulations,[79−81] by allowing the system to cross high energy barriers at higher temperatures. Although the method obviously brings some limitations compared to plain standard simulations such as a less accurate force field description in high temperatures or tendency of REMD to increasing crossing rate of enthalpic barriers but not in entropic barriers,[82] it facilitates sampling of unfolding and folding events. The enhanced sampling is also able to reveal force field inaccuracies and artifacts that remain hidden in plain MD simulations when they start from native structures. In contrast to the earlier folding studies of RNA TL, we use the latest *ff*99bsc0χ_{OL3} version of the AMBER force field,[60] which is the first force field shown to provide stable native structures in long simulations of RNA TLs on the microsecond time scale.[56] For each TL, we run two REMD simulations, one starting from the native structure and the other one from an unfolded single stranded structure with RMSD at least 9 Å from the native conformation. REMD simulations were run for 400 ns for each of 48 replicas (in total, an accumulated simulation time of ∼2 × 19 μs for each TL was obtained).

Identification of Folded, Misfolded, and Unfolded States in the Ensemble of Conformations. A large variety of TLs conformational states were sampled during the REMD simulations. We characterize the diversity of structures using a cluster analysis carried out for all REMD simulations starting from folded states.

To properly assess the quality of folding simulations, it is imperative to have the right criteria to define the folded state. Thus, the folded states of both TLs were identified strictly as only those structures satisfying all signature interactions (see Methods for their definition) of a given TL and sharing the same characteristic arrangement of stem and loop bases with the corresponding experimental structures. The other structures (not properly folded) generated within the REMD simulations were further divided into two classes, unfolded and misfolded structures, based on absence or presence of the canonical base-pairs in the stem. The structures with non-native conformation of TL but correctly formed stem duplex (with toleration of unpairing of the most terminal base-pair in the stem) were identified as misfolded. These structures adopt the native secondary structure, but the loop has locally a non-native tertiary structure arrangement. Finally, structures lacking pairing in the stem were identified as unfolded. The misfolded structures were further clustered according to the structure of the loop, mainly the conformation of the L_1/L_4 base-pair.

Three significantly populated misfolded states were observed in the simulations of the UUCG TL (Figure 1): (i) The misfolded-anti state, which is structurally relatively close to the experimental UUCG hairpin structure having RMSD ∼2 Å from the experimental structure but differs in glycosidic torsion angle (χ dihedral) of G_{L4} base, which is *anti* instead of the native *syn*, see Figure 2. (ii) The misfolded-compact state with the L_1/L_4 base-paired in tWW U_{L1}/G_{L4} manner instead of the tWS G_{L4}/U_{L1} of the native state (both having G_4 in *syn*). In this state, a higher conformational heterogeneity of U_{L2} and C_{L3} nucleotides with respect to the native state was observed and monitored by temperature B-factors (data not shown). (iii)

dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

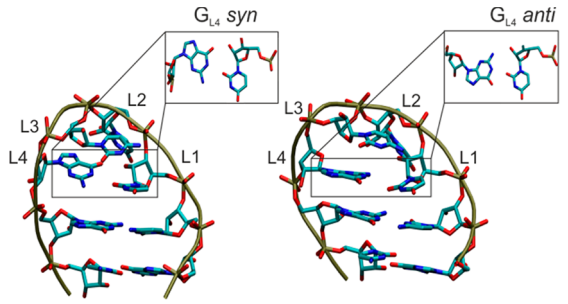


Figure 2. Comparison of structures of UNCG TL with G_{L4} in *syn* (native) and *anti* (misfolded-anti) conformations of UUCG TL. Details of the L_1/L_4 base-pairs are shown in the boxes.

The misfolded-bulged state lacking the L_1/L_4 base-pairing, with G_{L4} bulged and U_{L1} retaining its native position.

Among the misfolded states, the most populated structures are the misfolded-bulged (~5% at 271 K) and misfolded-anti (~6% at 271 K) conformations (populations of native folded and unfolded state are ~73% and ~13%, respectively) albeit the observed estimates of the populations are compromised by limited convergence of REMD simulations (see the next section). The misfolded-anti state was also observed in an earlier study of the folding/unfolding of UUCG hairpin by Garcia and Paschek.[29] However, in their study this misfolded state was included in the native state, defined using a very loose 4 Å RMSD criterion for heavy atoms of the stem and backbone atoms of the loop. This would include both the native state and all the misfolded states (Supplemental Figure S8). It is worth noting, that in our REMD simulations starting from unfolded UUCG TL conformation, we obtained the misfolded-anti structure as a dominant misfolded state (see the next section "Convergence and Sampling Efficiency of the REMD Simulations").

The GAGA TL simulations also revealed three major misfolded states: (i) the misfolded-compact state with L_1/L_4 base-pair in the native *c*HS conformation but non-native conformation of L_2 and L_3 nucleotides, (ii) the misfolded-bulged state that similarly to the UUCG misfolded-bulged state lacked the L_1/L_4 base-pairing as L_4 nucleotide was bulged while L_1 nucleotide retained its native conformation, and (iii) the misfolded-intercalated state, with L_4 stacked between L_1 and adjacent base-pair of the stem (Figure 1). The populations of misfolded-bulged, misfolded-compact, and misfolded-intercalated structures are ~4%, ~6%, and ~2% at 270 K, respectively; however, it should be again noted that these estimates are not converged due to limited sampling as discussed in the next section (populations of native folded and unfolded states equal to ~78% and ~9%, respectively).

Convergence and Sampling Efficiency of the REMD Simulations. The replica exchange method is increasingly used to improve sampling in MD simulations of biological systems. However, despite REMD having been shown to be very effective in enhancing conformational sampling in small model systems like small peptides, the efficiency of REMD for larger structures, such as the TLs, in explicit water models has not been tested extensively.

In order to analyze the convergence of our REMD simulations, we performed two sets of REMD simulations starting from either folded or unfolded states (see Methods). The results show that, despite the extensive sampling, our simulations do not obtain a representative sample of the equilibrium distribution. Fully "converged" REMD simulations should reveal similar relative populations of folded, misfolded, and unfolded states with respect to the temperature regardless the starting structures. In the case of REMD simulations starting from the folded states, the relative population of the (initially) folded state rapidly decreased at the expense of a rising population of unfolded states at high temperatures. Within ~100 ns the temperature profiles of relative populations of folded, misfolded, and unfolded states were apparently stable and only slightly varied with increasing length of the simulations (see Supporting Information Figures S9 and S10). This could give a false impression that the REMD is converged. However, in the case of REMD simulations starting from the unfolded states, the populations of the unfolded state remained dominant at all temperatures, while only a small population of folded states appeared at the final part of the

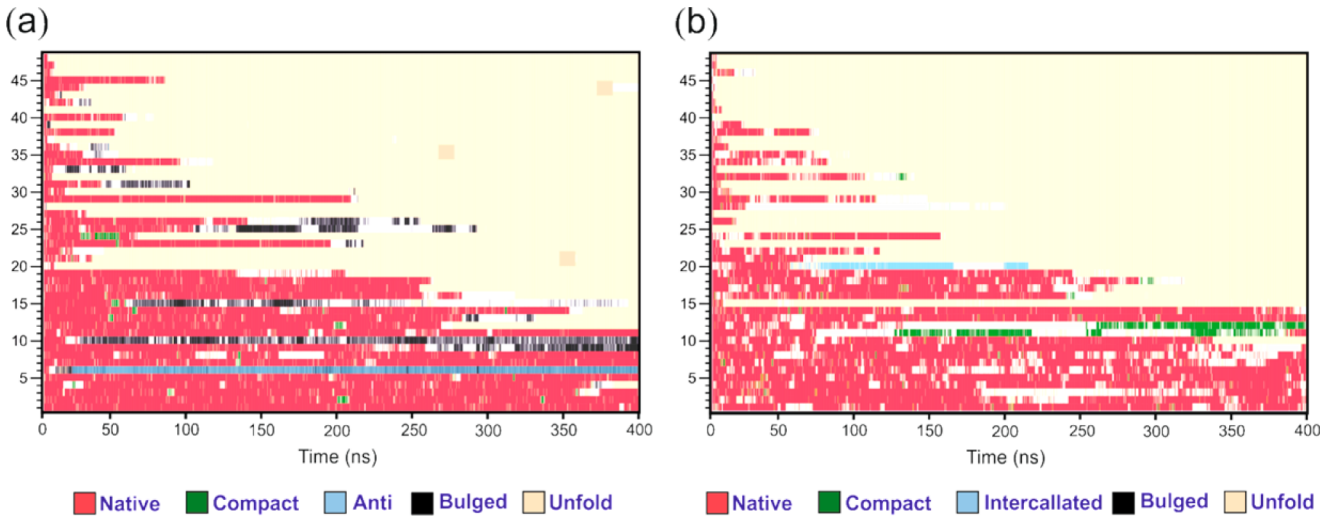


Figure 3. The time evolution of folded, misfolded, and unfolded states over REMD simulations started from the folded state for (a) the UUCG tetraloop and (b) the GAGA tetraloop. Each of the horizontal stripes corresponds to one of 48 unique replica simulations. The stripes are sorted by the mean temperature of unique replica simulation. The uppermost stripe has the highest mean temperature, while the lowermost stripe corresponds to the lowest mean temperature. The color of the stripes corresponds to the given state or is white when no cluster was identified.

dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

simulations. In addition, we monitored the number of transitions between the folded, misfolded, and unfolded states in the REMD starting from the folded structures. We found that at least some conformational transitions between folded, misfolded, and unfolded states were relatively rare, which caused imperfect convergence of the simulations despite the 400 ns time scale (Figure 3). We suggest that the simulation times necessary for proper convergence should be notably longer than those suggested by earlier studies (30 ns,[45,48] 36 ns,[84] 100 ns,[30] 120 ns,[85] and 226 ns[29] per replica) and evidently even much longer than the 400 ns per replica used in our simulations.

The longer time needed for convergence in this study, relative to earlier ones, may be attributed to a number of reasons. The first is that we apply quite strict criteria for definition of the native folded state, requiring the presence of all signature hydrogen bonds in the loop region (L_1-L_4), while other works used a relatively unrestrictive RMSD metric, masking considerable deviations in the conformation of the loop region. Thus, e.g., the RMSD of the misfolded-anti state is only ∼2 Å from the native folded state and even the ladder-like structure (eliminated with the presently used *ff*99bsc0χ_{OL3} force field), a major artifact of older force fields without any biological relevance,[56,57,60] is ∼3.8 Å in RMSD from the native structure. If we use a 4 Å RMSD from the native structure as the only parameter for defining the native state, both structures would be indeed considered as native-like despite the fact they significantly differ from the native structure; cf. Figure 5 in ref 56 to visually compare structures of native and ladder-like structure of GNRA TL. Our strict definition of the folded state is motivated by the experimental evidence that the native structures of RNA TLs are very precisely defined, i.e., these RNA motifs are not flexible molecules possessing multiple competing substates.[6] Another possible contribution to the difference from earlier studies may be that the dynamics in the new force field is slower due to having higher barriers for torsional transitions. However, we note that it may also be an intrinsic property of RNA folding, as experimental folding relaxation rates for RNA hairpins are typically about an order of magnitude smaller than for small autonomously folding peptides[36]

We therefore cannot estimate equilibrium properties such as the folding free energy and melting temperature (for the sake of completeness, see Supporting Information for folding free energy and melting temperature calculated from our REMD simulations started from the folded state). We suggest that such computations are at the limit of the capability of contemporary computational techniques and would require additional enhanced sampling techniques. Rather than attempting to compute stabilities, we instead use the REMD sampling to identify stable states on the energy landscape and their interconversions and to demonstrate *ab initio* folding starting from the unfolded state.

UUCG Misfolded-Anti State. The misfolded-anti state is the most populated non-native structure at the lowest temperatures in REMD simulations of the UUCG TL starting from the folded structure. In addition, the misfolded-anti structure was spontaneously formed in one replica of the UUCG REMD simulations started from the unfolded structure, while the formation of the completely folded state was not observed in this REMD attempt at the present time scale. Note that the structure corresponding to the misfolded-anti state of UUCG TL only rarely occurs in the database of X-ray structures, and its rare occasions can be due to uncertainty in refinement or interactions with neighboring molecules. For instance, the UCCG TL in the L1 stalk of the *Thermus thermophilus* structure of 50S rRNA subunit (PDB ID 3V2F,[86] resolution 2.7 Å) contains G2147 in *anti*-conformation, while a more recent structure of the isolated L1 stalk with higher resolution (PDB ID 3U4M,[87] resolution 2.0 Å) revealed a *syn*-conformation of G2147.[68] Definitely, the native topology of UNCG TL requires *syn* G_4 geometry, and any real deviation would have to be imposed by some highly specific tertiary interactions, which obviously are absent in simulations. The available experimental data do not support the significant preference of UUCG TL for the misfolded-anti state seen in MD simulations for isolated TLs; however, they do suggest that it is accessible. There are several examples in which such a conformation is observed in complexes with other molecules (see Supporting Information and Supplemental Figures S11 and S13). There are two possible explanations of the observed simulation behavior: (i) Either the misfolded-anti state is artificially overstabilized by the force field or (ii) the misfolded-anti state is kinetically more accessible from the completely unfolded state than the native folded state, which in combination with a limited sampling might finally result in observed preference of the misfolded-anti state during the attempted UUCG TL folding process.

In order to test the former possibility, i.e. the force field bias, we performed the alchemical thermodynamic integration (TI). In this calculation, the G_{L4} guanine having *syn*-orientation of glycosidic torsion was disappearing with parameter λ, while the same guanine in *anti*-orientation was simultaneously appearing. We performed molecular dynamics of UUCG TL at nine different λ-values defining an alchemical transformation path from native G_{L4}-*syn* to the misfolded-anti (G_{L4}-*anti*) state. At each λ-value we monitored the mean force on the λ-coordinate, and subsequently we calculated the $\Delta G_{syn\rightarrow anti}$ free energy difference between the native and the misfolded-anti state by integrating the mean force over the λ coordinate (see Methods). We observed that (within the force field approximation) the native UUCG conformation with G_{L4} in *syn* is stabilized by ∼15 ± 8 kcal/mol over the misfolded-anti state having G_{L4} in *anti* (see the Supporting Information). It is evident that the calculated absolute ΔG value is substantially overestimated[88] (we note that the experimental free energy of folding of the whole TL is 3.8 kcal/mol[89]), but we may expect that the TI calculation should at least capture the correct trend. The TI calculation thus suggests that the native state is more stable than the misfolded-anti state taking into account all limitations of the used TI method, which are discussed in the Supporting Information. As both REMD and TI were performed in the same force field and the same MD setup, the obtained free energy difference between native and misfolded-anti states indicates that that the preferential formation of the misfolded-anti state is not due to a force field imbalance.

Taken together, we assume that the preferential formation of the misfolded-anti state in REMD starting from the unfolded state is caused by an insufficient sampling. As the *anti*-orientation of the glycosidic torsion corresponds to a significantly wider minimum on the potential energy surface than the native *syn*-orientation,[60] the *anti*-orientation is supposed to be entropically favored. Thus it is possible that the native folded state of UUCG TL having G_{L4} in *syn* is entropically disfavored over the misfolded-anti state; however,

it should be stabilized by the enthalpy, so that the native state is lower in total free energy than the misfolded-anti state. As most of the transitions between native folded, misfolded-bulged, and misfolded-anti states occurred at high temperatures, the effect of stabilizing enthalpy is diminished due to high temperatures, and thus the occurrence of native conformation with G_{L4} in *syn* is rare even at high temperatures. Note that any occurrence of the folded state would most likely lead to shifting of the given replica toward lower temperatures and thus in sufficiently long simulations the population of the folded state at lower temperature should be higher than the population of the misfolded-anti state. However, upon limited sampling and due to rareness of occurrence of G_{L4} *syn*-orientation at higher temperatures, the (unconverged) probability of the misfolded-anti state might be higher than the probability of the folded state. Similar trapped conformational states were also observed in REMD simulations of DNA hairpin loops.[33,84] This could explain the observed REMD simulations behavior and further underlines the sampling limitation of the method for even the smallest nucleic acids.

Misfold-Bulged States. The misfolded-bulged state was identified in the REMD simulations of both UUCG and GAGA TLs. In both systems, the L_4 base was bulged out, while the L_1 base retained its place. This is not surprising as the L_1 base has a bigger overlap with the adjacent terminal base-pair of the stem and is thus stabilized at its place by stacking. On the other hand, the overlap and corresponding stacking interaction of L_4 base with the adjacent base-pair of the stem is significantly weaker, and thus the L_4 is relatively free to unstack and bulge out. Note that the backbone between the stem and L_1 base has the same conformation as the backbone in an A-RNA duplex (1a family according to Richardson nomenclature[90]) and thus structurally matches the first nucleotide of the 3′-overhang that is known to stabilize the terminal base-pair of an A-RNA helix.[91]

Surprisingly, we observed almost the same propensity to form the misfolded-bulged state in both TLs. In contrast, experiments have only identified L_4-bulged structures in the GNRA TL, e.g. in complex with ribotoxin restrictocin.[27] No such experimental evidence is available in the case of the UNCG TL. This discrepancy might be explained as a consequence of potential force field artifact. Recently we showed that the stacking interactions are overestimated in empirical force fields with concomitant underestimation of hydrogen bonding.[92] We argued that this effect is most likely caused by anisotropy of the dispersion interaction which is not captured by the isotropic van der Waals term of the force fields.[92,93] Note that the L_1 base is fixed within the L_1/L_4 base-pair by three hydrogen bonds in UUCG TL ($U_{L1}(O2)\ldots G_{L4}(N1)$, $U_{L1}(O2)\ldots G_{L4}(N2)$, $U_{L1}(O2')\ldots G_{L4}(O6)$), while only by one in the case of GNRA TL ($G_{L1}(N2)\ldots A_{L4}(N7)$) (Figure 4). It is thus likely

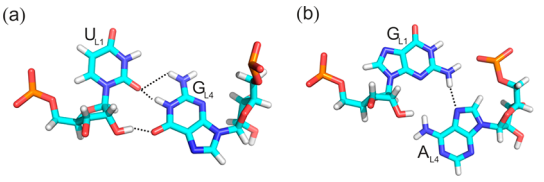


Figure 4. Signature H-bonds between L_1/L_4 base-pair (black dashed lines): (a) UUCG TL and (b) GAGA TL.

that the more realistic description of hydrogen bonding, i.e. the stronger interaction of each hydrogen bond, would only negligibly affect stability of the L_1/L_4 base-pair and consequently the population of the misfolded-bulged state in the GNRA TL. However, the L_1/L_4 base-pair of UNCG would be more affected by such an inaccuracy of the force field. We thus tentatively suggest that the population of the UUCG misfolded-bulged state might be artificially overestimated due to an underestimated hydrogen bonding interaction within the L_1/L_4 base-pair in empirical force fields. This could be consistent with excessive fraying of terminal base-pairs in the simulation of duplexes.[94]

Folding Pathway. From the REMD simulations initialized from the unfolded states, we can infer a putative folding pathway of UUCG and GAGA TLs. As explained above, although the simulations starting from the unfolded state were far from reaching full convergence of the folding process, we were able to observe formation of native or near-native structures (with RMSD <2 Å from the native structure) in a few replicas of both GAGA and UUCG TL REMD simulations. Although the transitions observed in REMD simulations usually span a range of temperatures, making it difficult to assign a specific temperature to a folding event, we can nonetheless sketch qualitative suggestions about the likely folding mechanism, based on the spontaneous formation of the native structure (or near-native in the case of UUCG, where the misfolded-anti state was formed instead of native as noted above) in these simulations.

The folding process of UUCG TL involves several steps (Figure 5a). At the beginning of the simulation, the fully unfolded state containing no base-pairs adopts a straight conformation with the initial RMSD from the experimental structure ∼9 Å. The first folding step involves a formation of a bend of a single strand, which is accompanied by a decrease of RMSD to ∼6 Å. Loop formation is initiated by making a contact between the L_2 and L_3 bases. The loop is then closed by a non-native canonical *c*WW base-pair between C_5 and G_{L4} (Figure 5a). This base-pair is subsequently broken, and the U_{L1}/G_{L4} base-pair is formed. However, G_{L4} binds U_{L1} in its *anti*-orientation, which results in the formation of the non-native *cis* Watson−Crick/Sugar-Edge (*c*WS) G_{L4}/U_{L1} base-pair. The native structure would require the *trans* Watson−Crick/Sugar-Edge (*t*WS) G_{L4}/U_{L1} base-pair with *syn* guanine orientation. The loop is stabilized by formation of native 7BPh interaction[69] between C_{L3}(N4) and U_{L2}(*pro*-R_p). Finally, the native canonical base-pairs of stem are formed by zipping from the loop outward to conformations having ∼2 Å RMSD (after omission of the flexible terminal base-pair) with respect to the native UUCG TL. These structures display the same characteristic arrangement of loop and stem bases except for the *anti*-conformation of the G_{L4} base and the corresponding *c*WS pattern of the G_{L4}/U_{L1} base-pair. Thus, this structure corresponds to UUCG misfolded-anti rather than to the native folded state, although the RMSD value is reasonably small and significantly below the loose RMSD criteria for the folded state used in the preceding studies.[29,30] A similar mechanism has been suggested in some recent studies.[28,38,40,41,43]

The misfolded-anti structure may correspond to a transiently trapped state (as discussed above) that has to undergo partial unfolding of the L_1/L_4 base-pair to form the native structure with the G_{L4} base in the *syn* conformation and *t*WS pattern of the G_{L4}/U_{L1} base-pair (Figure 5a). We observed this transition pathway in the reverse direction, i.e. from *syn*- (folded state) to

dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

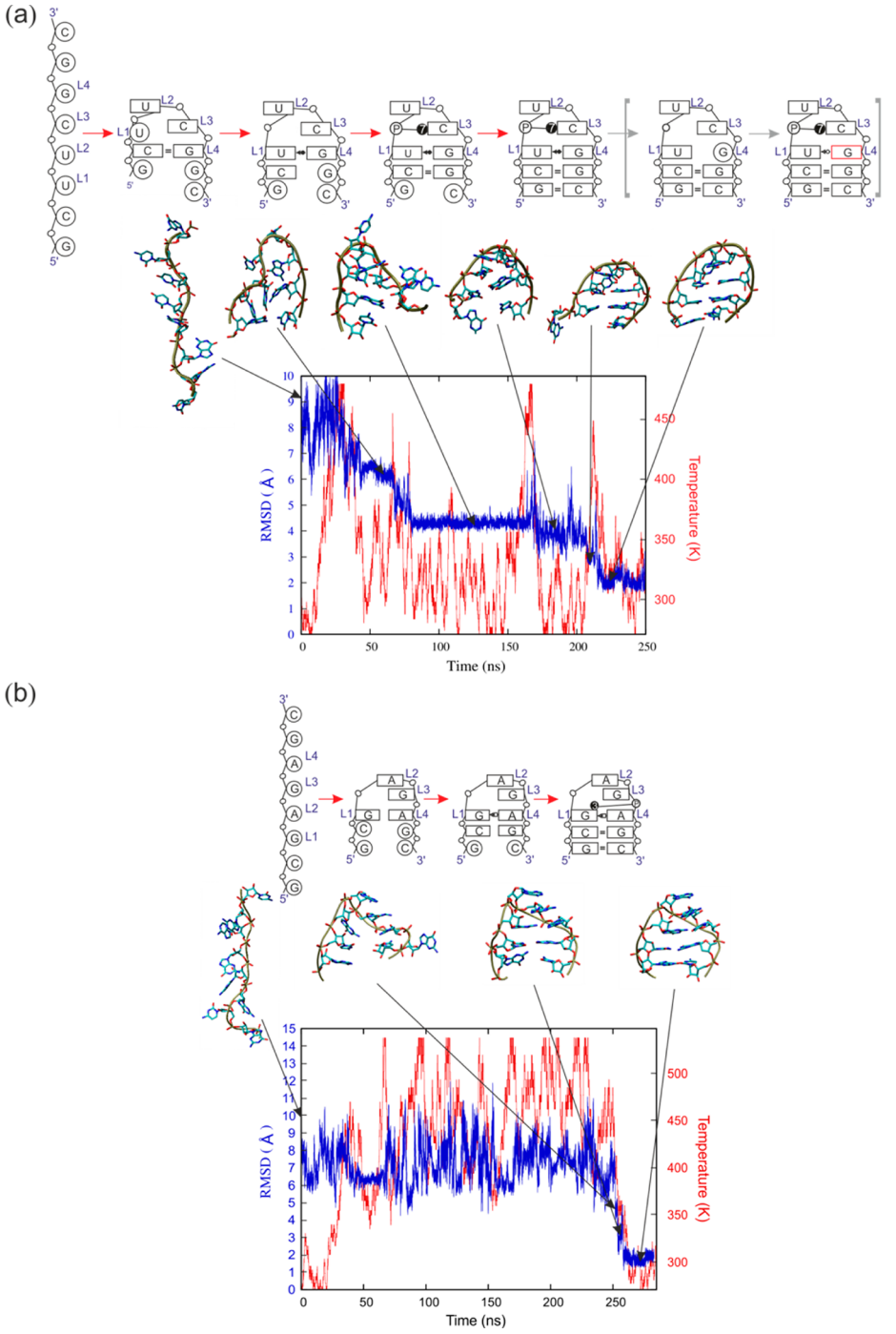


Figure 5. Schematic plots for the folding pathway of the studied tetraloops (TLs): (a) the UUCG TL and (b) the GAGA TL. The evolution of the all-atom RMSD distance of each TL relative to the native structure and the time series of simulation temperature are shown by blue and red lines, respectively. The nucleotides in circles are flexible, and their position is not well-defined compared to the other nucleotides (in boxes). The hypothetical step from the trapped misfolded-anti structure toward a native folded state not achieved (in the folding direction) in the simulations is shown in gray brackets (see also Figure 6); *syn*-guanosine is in red.

anti-conformation (misfolded-anti) of G_{L4}, in one replica of the REMD simulations initiated from the folded state, but we did not see it in the direction from misfolded-anti to native *syn*-conformation. The transition included disruption of the L_1/L_4 base-pair followed by the G_{L4} base bulging out, i.e. it proceeded via the misfolded-bulged state (Figure 6). After several nanoseconds, the G_{L4} base re-established the base-pair with U_{L1}, but in an *anti*-conformation. The L_4/L_1 base-pair then formed the *c*WS base-pair instead of the native *t*WS. Taking

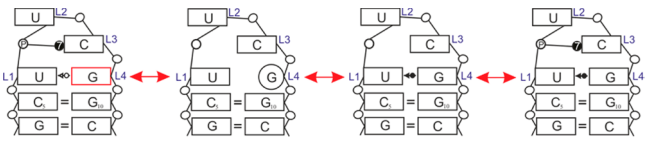


Figure 6. The transition pathway of changing *syn* conformation of L_4 to *anti* conformation in UUCG TL.

dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

together, the misfolded-anti state could represent the intermediate state of folding of both the native structure as well as the misfolded-anti state. As discussed above, the *anti*-conformation of G_{L4} in the misfolded-bulged state is likely entropically favored over the *syn*-conformation, which means that the folding pathway leading to the misfolded-anti structure proceeds through an entropically favored intermediate and thus is kinetically preferred over the path resulting in the native fold. Note that the native fold with G_{L4} in *syn* is thermodynamically favored over the misfolded-anti state as shown in our TI simulation. We can thus suggest that the misfolded-anti state most likely represents a kinetic trap for the folding of UUCG TL. This conclusion is in agreement with experimental observation of Bevilacqua and co-workers, who studied folding kinetics of UUCG and UUC8BrG TLs. The 8-bromoguanosine (8BrG) has limited rotation around glycosidic torsion with *syn*-conformation being favored over *anti*-conformation. The authors suggested that reduction of conformational variability of G_{L4} accompanied with stabilization of its *syn*-conformation increased stability of the TL.[38,95]

The folding pathway of the GAGA TL is very similar to the folding pathway of the UUCG TL. The starting structure is an unfolded straight single strand with RMSD ∼10 Å from the native conformation (Figure 5b). The first step of folding includes a formation of loop getting L_2-L_4 bases into a native position, which is associated with a decrease of RMSD to 5 Å. After this, the loop is closed by formation of the native L_1/L_4 base-pair and base-phosphate interaction between $G_{L1}(N2)\cdots A_{L4}(pro\text{-}R_p)$. Once the loop is stabilized by the base-pair and base-phosphate interaction (RMSD ∼3 Å), zipping of the remaining base-pairs of the stem occurs rapidly. Although a complete characterization of the pathways would require more events to be sampled, under equilibrium conditions, some features of the mechanism appear to be shared with experimentally inferred mechanisms for RNA: there are a number of long-lived, stable intermediates populated en route to the native state, and these are frequently stabilized by non-native interactions which need to be broken before the folding can proceed toward the correct structure.[96]

Unlike other force fields, which are prone to form artificial "ladder-like" structures as their most stable state in REMD,[56] the folded state formed in our folding simulations has the correct geometric characteristics of the folded RNA TL, except the *syn*/*anti* flip of G_{L4} in the UUCG TL as discussed above. This is in contrast to previous studies using the AMBER ff99 force field, for which even simulations initiated from the correctly folded structure rapidly formed ladder-like structures.[56]

CONCLUSIONS

Using the most recent reparameterization of the AMBER family of RNA force fields (ff99bsc0χ_{OL3}),[56,60,64] we were able to obtain the first all atomic REMD molecular dynamics simulation starting from fully unfolded conformations that resulted in the folded TL structures within 2 Å all-atom RMSD of the native structure. The folded structures have the TL signature interactions (except for the G_{L4} *anti*-orientation in UUCG TL). The previous pioneering studies applying the REMD technique to TL folding were only able to reach structures having ∼4 Å RMSD from the native state due to force field limitations. Such structures lack many key signature interactions in the loop region while often adopting deformed untwisted ladder-like stem shape. This suggests that the ff99bsc0χ_{OL3} force field[56,60] has visibly improved the potential for predictive simulations of RNA structural motifs, although the force field certainly remains far from perfect. Our work also demonstrates that simulation studies of the folding of nucleic acids should use considerably stricter criteria for the folded state than common until now.

Although we did not reach an equilibrated population of the folded state, when starting from unfolded structures, we still observed folding events in few replicas. The observed mechanism of folding shares some aspects for both studied UUCG and GAGA TLs. In both cases, the folding is initiated by formation of a bend in the unstructured RNA single strand that is followed by formation of transient interactions stabilizing the bend. Subsequently the L_1/L_4 base-pair replaces the transient contacts, after that the TL is "locked" by signature interactions, which in both cases comprise the base-phosphate interactions. The stem is then formed by a zipping mechanism. In addition, as we observed a rather limited convergence on the submicrosecond time scale, we suggest that although the REMD technique is a powerful tool for significantly increasing sampling, there still might be processes that are rare events even at high temperature. Consequently, the proper sampling of these processes cannot be overcome by the REMD scheme and either a longer simulation time or more powerful enhanced sampling method is required to achieve the full convergence for such processes. In other words, the presently affordable computer time scales do not allow fully converged folding of even the smallest nucleic acids systems.

Despite the limited convergence that precludes the quantitative estimation of melting temperature and free energy of folding, we were able to address some important questions related to the biological function of TLs. We have identified a GAGA misfolded-bulged state, which shares many structural features with the X-ray structure of GNRA TL in complex with restrictocin. Thus we suggest that the GNRA TL is able to transiently expose the A_{L4} base for possible base-pairing and subsequent restrictocin binding to the R_{L3} and A_{L4} bases that are both exposed to the solvent in this conformation. In other words, it appears that the GNRA TL is able to bind restrictocin via a "conformational selection" mechanism.

In addition we found that the folding of UUCG TL proceeds through a misfolded-bulged intermediate, from which either a native folded conformation or a misfolded-anti state having G_{L4}-*anti* (instead of native G_{L4}-*syn*) are accessible due to free *syn*/*anti* rotation of glycosidic torsion of bulged G_{L4}. On the basis of our observation of the preferential formation of the misfolded-anti state over the native state and the assumption of an entropic penalty of *syn*-orientation of bulged guanine, we argued that the misfolded-anti state is most likely a kinetic trap of the UUCG folding. These conclusions are in agreement with the enhanced folding rate upon 8-bromoguanosine mutation of G_{L4} in UUCG TL reported by Bevilacqua et al.[38,95] We can also relate our results to ultrafast spectroscopy experiments[97] that identified some features of misfolded states of GNRA TLs reminiscent of the misfolded-compact GAGA structure identified in this study.

ASSOCIATED CONTENT

Supporting Information

Details about convergence of our REMD simulations, non-native structures of studied TLs including their structural context obtained by bioinformatics search, details about our thermodynamics integration, and the calculation of folding/

dx.doi.org/10.1021/ct301086z | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

unfolding free energy and melting temperature of studied TLs obtained from present REMD. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: michal.otyepka@upol.cz (M.O.), pavel.banas@upol.cz (P.B.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by The Operational Program Research and Development for Innovations-European Regional Development Fund (project CZ.1.05/2.1/00/03.0058 of the Ministry of Education, Youth and Sports of the Czech Republic), the Operational Program Education for Competitiveness - European Social Fund (projects CZ.1.07/2.3.00/20.0017 and CZ.1.07/2.3.00/20.0058 of the Ministry of Education, Youth and Sports of the Czech Republic), project "CEITEC - Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068) from European Regional Development Fund, and the Grant Agency of the Czech Republic (grants no. P301/11/P558, P208/12/G016, P208/12/1878, and P305/12/G034). R.B. was supported by a Royal Society University Research Fellowship and the Intramural Research Program of the NIDDK, NIH.

REFERENCES

(1) Bevilacqua, P. C.; Blose, J. M. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79−103.
(2) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; Daubentoncarafa, Y.; Uhlenbeck, O. C.; Tinoco, I.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1364−1368.
(3) Uhlenbeck, O. C. *Nature* **1990**, *346*, 613−614.
(4) Woese, C. R.; Winker, S.; Gutell, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8467−8471.
(5) Wolters, J. *Nucleic Acids Res.* **1992**, *20*, 1843−1850.
(6) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481−1491.
(7) Sheehy, J. P.; Davis, A. R.; Znosko, B. M. *RNA* **2010**, *16*, 417−429.
(8) Varani, G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379−404.
(9) Marino, J. P.; Gregorian, R. S.; Csankovszki, G.; Crothers, D. M. *Science* **1995**, *268*, 1448−1454.
(10) Chauhan, S.; Woodson, S. A. *J. Am. Chem. Soc.* **2008**, *130*, 1296−1303.
(11) Jaeger, L.; Michel, F.; Westhof, E. *J. Mol. Biol.* **1994**, *236*, 1271−1276.
(12) Selinger, D.; Liao, X. B.; Wise, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5409−5413.
(13) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. *J. Mol. Biol.* **2000**, *304*, 35−42.
(14) Tishchenko, S.; Nikulin, A.; Fomenkova, N.; Nevskaya, N.; Nikonov, O.; Dumas, P.; Moine, H.; Ehresmann, B.; Ehresmann, C.; Piendl, W.; Lamzin, V.; Garber, M.; Nikonov, S. *J. Mol. Biol.* **2001**, *311*, 311−324.
(15) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *Nature* **2000**, *407*, 340−348.
(16) Scott, W. G.; Finch, J. T.; Klug, A. *Cell* **1995**, *81*, 991−1002.
(17) Correll, C. C.; Wool, I. G.; Munishkin, A. *J. Mol. Biol.* **1999**, *292*, 275−287.
(18) Correll, C. C.; Swinger, K. *RNA* **2003**, *9*, 355−363.
(19) Correll, C. C.; Beneken, J.; Plantinga, M. J.; Lubbers, M.; Chan, Y. L. *Nucleic Acids Res.* **2003**, *31*, 6806−6818.
(20) Correll, C. C.; Munishkin, A.; Chan, Y. L.; Ren, Z.; Wool, I. G.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13436−13441.
(21) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905−920.
(22) Szewczak, A. A.; Moore, P. B.; Chan, Y. L.; Wool, I. G. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9581−9585.
(23) Szewczak, A. A.; Moore, P. B. *J. Mol. Biol.* **1995**, *247*, 81−98.
(24) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191−194.
(25) Pley, H. W.; Flaherty, K. M.; Mckay, D. B. *Nature* **1994**, *372*, 68−74.
(26) Qin, P. Z.; Feigon, J.; Hubbell, W. L. *J. Mol. Biol.* **2005**, *351*, 1−8.
(27) Yang, X. J.; Gerczei, T.; Glover, L.; Correll, C. C. *Nat. Struct. Biol.* **2001**, *8*, 968−973.
(28) Bowman, G. R.; Huang, X. H.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2008**, *130*, 9676−9678.
(29) Garcia, A. E.; Paschek, D. *J. Am. Chem. Soc.* **2008**, *130*, 815−817.
(30) Zuo, G.; Li, W.; Zhang, J.; Wang, J.; Wang, W. *J. Phys. Chem. B* **2010**, *114*, 5835−9.
(31) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. *J. Mol. Biol.* **2002**, *317*, 493−506.
(32) Zhuang, Z. Y.; Jaeger, L.; Shea, J. E. *Nucleic Acids Res.* **2007**, *35*, 6995−7002.
(33) Portella, G.; Orozco, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 7673−7676.
(34) Hyeon, C.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6789−6794.
(35) Hyeon, C.; Thirumalai, D. *J. Am. Chem. Soc.* **2008**, *130*, 1538−1539.
(36) Zhang, W. B.; Chen, S. J. *Biophys. J.* **2006**, *90*, 765−777.
(37) Zhang, W. B.; Chen, S. J. *Biophys. J.* **2006**, *90*, 778−787.
(38) Ma, H. R.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530.
(39) Svoboda, P.; Di Cara, A. *Cell. Mol. Life Sci.* **2006**, *63*, 901−918.
(40) Ansari, A.; Kuznetsov, S. V.; Shen, Y. Q. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7771−7776.
(41) Ansari, A.; Kuznetsov, S. V. *J. Phys. Chem. B* **2005**, *109*, 12982−12989.
(42) Jung, J.; Van Orden, A. *J. Am. Chem. Soc.* **2006**, *128*, 1240−1249.
(43) Ma, H.; Wan, C.; Wu, A.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 712−716.
(44) Williams, D. J.; Hall, K. B. *Biophys. J.* **1999**, *76*, 3192−3205.
(45) Villa, A.; Widjajakusuma, E.; Stock, G. *J. Phys. Chem. B* **2008**, *112*, 134−142.
(46) Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wohnert, J.; Stock, G.; Schwalbe, H. *Nucleic Acids Res.* **2008**, *36*, 1928−1940.
(47) Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2006**, *34*, 697−708.
(48) DePaul, A. J.; Thompson, E. J.; Patel, S. S.; Haldeman, K.; Sorin, E. J. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.
(49) Riccardi, L.; Nguyen, P. H.; Stock, G. *J. Phys. Chem. B* **2009**, *113*, 16660−16668.
(50) Deng, N. J.; Cieplak, P. *Biophys. J.* **2010**, *98*, 627−636.
(51) Leontis, N. B.; Westhof, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300−308.
(52) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. *J. Math. Biol.* **2008**, *56*, 215−252.
(53) Fadrna, E.; Spackova, N.; Stefl, R.; Koca, J.; Cheatham, T. E.; Sponer, J. *Biophys. J.* **2004**, *87*, 227−242.
(54) Fadrna, E.; Spackova, N.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E.; Kulinski, T.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 2514−2530.
(55) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. *Acc. Chem. Res.* **2010**, *43*, 40−47.
(56) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.

(57) Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, 6642−6652.
(58) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(59) Besseova, I.; Banáš, P.; Kührová, P.; Košinová, P.; Otyepka, M.; Šponer, J. *J. Phys. Chem. B* **2012**, *116*, 9899−9916.
(60) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(61) Ode, H.; Matsuo, Y.; Neya, S.; Hoshino, T. *J. Comput. Chem.* **2008**, *29*, 2531−2542.
(62) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. *J. Chem. Theory Comput.* **2010**, *6*, 1520−1531.
(63) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(64) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.
(65) Sklenovsky, P.; Florova, P.; Banas, P.; Reblova, K.; Lankas, F.; Otyepka, M.; Sponer, J. *J. Chem. Theory Comput.* **2011**, *7*, 2963−2980.
(66) Banas, P.; Sklenovsky, P.; Wedekind, J. E.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2012**, *116*, 12721−12734.
(67) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(68) Nozinovic, S.; Furtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. *Nucleic Acids Res.* **2010**, *38*, 683−694.
(69) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 4898−4918.
(70) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(71) Case, A. D.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvai, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11*; University of California: San Francisco, 2010.
(72) Sousa da Silva, A. W.; Vranken, W. F. *BMC Res. Notes* **2012**, *5*, 367.
(73) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(74) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(75) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(76) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306−317.
(77) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(78) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(79) Klenin, K.; Strode, B.; Wales, D. J.; Wenzel, W. *BBA, Biochim. Biophys. Acta, Proteins Proteomics* **2011**, *1814*, 977−1000.
(80) Liwo, A.; Czaplewski, C.; Oldziej, S.; Scheraga, H. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134−139.
(81) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57−83.
(82) Beck, D. A. C.; White, G. W. N.; Daggett, V. *J. Struct. Biol.* **2007**, *157*, 514−523.
(83) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499−512.
(84) Kannan, S.; Zacharias, M. *Biophys. J.* **2007**, *93*, 3218−3228.
(85) Zhang, Y. F.; Zhao, X.; Mu, Y. G. *J. Chem. Theory Comput.* **2009**, *5*, 1146−1154.
(86) Polikanov, Y. S.; Blaha, G. M.; Steitz, T. A. *Science* **2012**, *336*, 915−918.
(87) Tishchenko, S.; Gabdulkhakov, A.; Nevskaya, N.; Sarskikh, A.; Kostareva, O.; Nikonova, E.; Sycheva, A.; Moshkovskii, S.; Garber, M.; Nikonov, S. *Acta Crystallogr., D* **2012**, *68*, 1051−1057.
(88) Spasic, A.; Serafini, J.; Mathews, D. H. *J. Chem. Theory Comput.* **2012**, *8*, 2497−2505.
(89) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 1045−1061.
(90) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA* **2008**, *14*, 465−481.
(91) Mohan, S.; Hsiao, C.; Bowman, J. C.; Wartell, R.; Williams, L. D. *J. Am. Chem. Soc.* **2010**, *132*, 12679−12689.
(92) Banas, P.; Mladek, A.; Otyepka, M.; Zgarbova, M.; Jurecka, P.; Svozil, D.; Lankas, F.; Sponer, J. *J. Chem. Theory Comput.* **2012**, *8*, 2448−2460.
(93) Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476−10493.
(94) Dršata, T.; Pérez, A.; Orozco, M.; Morozov, A. V.; Šponer, J.; Lankaš, F. *J. Chem. Theory Comput.* **2012**, *9*, 707−721.
(95) Proctor, D. J.; Ma, H. R.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. *Biochemistry* **2004**, *43*, 14004−14014.
(96) Chen, S. J. *Annu. Rev. Biophys.* **2008**, *37*, 197−214.
(97) Zhao, L.; Xia, T. B. *J. Am. Chem. Soc.* **2007**, *129*, 4118−4119.