



Published in final edited form as:

*J Chem Theory Comput.* 2016 September 13; 12(9): 4534–4548. doi:10.1021/acs.jctc.6b00300.

## Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies

Petra Kührová<sup>§</sup>, Robert B. Best<sup>‡</sup>, Sandro Bottaro<sup>†</sup>, Giovanni Bussi<sup>†</sup>, Jiří Šponer<sup>§,£,\*</sup>, Michal Otyepka<sup>§,£</sup>, and Pavel Banáš<sup>§,£,\*</sup>

<sup>§</sup>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, pavel.banas@upol.cz

<sup>‡</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

<sup>†</sup>Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, 34136 Trieste, Italy

<sup>£</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic

### Abstract

The computer aided folding of biomolecules, particularly RNAs, is one of the most difficult challenges in computational structural biology. RNA tetraloops are fundamental RNA motifs playing key roles in RNA folding and RNA-RNA and RNA-protein interactions. Although state-of-the-art Molecular Dynamics (MD) force fields correctly describe the native state of these tetraloops as a stable free-energy basin on the microsecond time scale, enhanced sampling techniques reveal that the native state is not the global free energy minimum, suggesting yet unidentified significant imbalances in the force fields. Here we tested our ability to fold the RNA tetraloops in various force fields and simulation settings. We employed three different enhanced sampling techniques, namely temperature replica exchange MD (T-REMD), replica exchange with solute tempering (REST2), and well-tempered metadynamics (WT-MetaD). We aimed to separate problems caused by limited sampling from those due to force-field inaccuracies. We found that none of the contemporary force fields is able to correctly describe folding of the 5'-GAGA-3' tetraloop over a range of simulation conditions. We thus aimed to identify which terms of the force field are responsible for this poor description of TL folding. We showed that at least two different imbalances contribute to this behavior, namely, overstabilization of base-phosphate and/or sugar-phosphate interactions and underestimated stability of the hydrogen bonding interaction in base pairing. The first artifact stabilizes the unfolded ensemble while the second one destabilizes the

\*corresponding authors.

#### SUPPORTING INFORMATION

Supporting Information includes detail description of used enhanced sampling techniques, setting of classical MD simulation used to seed the starting structures of REMD simulations, histograms of RMSD and eRMSD over all replicas, time evolution of the stem/loop populations in T-REMD and REST2 simulations, time evolution of structural dynamics in WT-MetaD simulations and detail description of the modifications applied to the clustering algorithm developed by A. Laio, detail description of the setting of the additional HBfix potential supporting hydrogen bonds, histograms of heavy atom distances of GC base pair hydrogen bonds in MD simulations compared to X-ray database, and AMBER library files including  $\chi_{OL3}CP$  force field.

folded state. The former problem might be partially alleviated by reparameterization of the Van der Waals parameters of the phosphate oxygens suggested by Case et al., while in order to overcome the latter effect we suggest local potentials to better capture hydrogen bonding interactions.

## Abstract

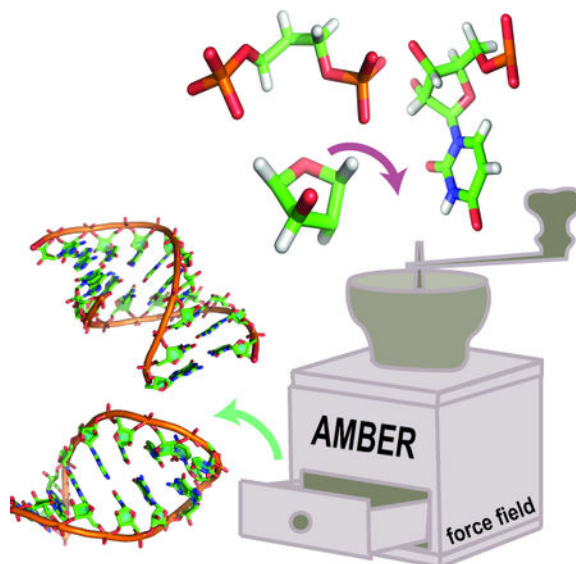


Table of Contents: The developments of enhanced sampling methods and force fields are highly mutually interrelated efforts. Here we used three different enhanced sampling techniques (temperature based replica exchange molecular dynamics, replica exchange solute tempering, and metadynamics) to fold 5'-GAGA-3' RNA tetraloop. We aimed to separate problems caused by limited sampling from those due to force-field inaccuracies, and identify which terms of the force field are responsible for poor description of tetraloop folding.

## Keywords

force field; RNA tetraloop; folding; enhanced sampling methods; REMD

## INTRODUCTION

Computer simulation of the folding of the biomolecules is one of the major challenges in biomolecular modelling. Simulations, together with experimental methods, help us to build a more complete picture of these nontrivial processes.<sup>1-7</sup> Computational studies of folding often use molecular dynamics (MD) simulations in conjunction with enhanced sampling methods. The accuracy, and in turn the predictive power of folding simulations critically depends on the quality of enhanced sampling techniques, accessible time scales, and most importantly on the force fields used. Although modern force fields often perform satisfactorily in unbiased simulations initiated from RNA experimental structures, this does not guarantee that the native state is the global free energy minimum within the force field

description, which would be required to achieve folding from the unfolded state. Therefore, contemporary force fields should be validated not only using unbiased MD simulations, which are typically trapped in a free energy basin corresponding to the starting structure, but also using enhanced sampling methods capable to describe e.g. folding of the RNA. The developments of enhanced sampling methods and force fields are thus highly mutually interrelated efforts. The enhanced sampling methods are designed to overcome energy barriers and provide a robust exploration of free energy landscape.<sup>8</sup> In the limit of converged sampling, the accuracy of enhanced sampling simulations is determined by the accuracy of the force field.<sup>9</sup>

There are many variants of enhanced sampling MD methods. The widely used temperature-based replica exchange MD<sup>10</sup> (T-REMD) benefits from multiple independent MD simulations running in parallel over a range of different temperatures. Exchanges of configurations between neighbouring temperatures are attempted at fixed time intervals and accepted according to a Metropolis-style algorithm to ensure canonical sampling at all temperatures. Another method based on multiple independent MD simulations is replica exchange with solute scaling (in a REST2 variant, see ref. <sup>11</sup>), which is a variant of Hamiltonian-based replica exchange MD (H-REMD).<sup>12</sup> This method is based on a modification of the potential energy, so that the interactions between solute atoms are scaled by a factor  $\lambda$ , solvent-solvent interactions remain unscaled, and solute-solvent interactions are scaled by an intermediate factor (in this case by  $\sqrt{\lambda}$ ). Scaling the energy by a factor  $\lambda$  is equivalent to a scaling of the temperature by  $1/\lambda$ . Thus, in case of REST2 only the solute atoms are effectively heated up. What is more important, the solvent-solvent interactions that typically contribute the most to the energy differences between replicas in T-REMD simulations do not contribute to exchanges. Another popular and conceptually very different enhanced sampling method is a well-tempered metadynamics (WT-MetaD).<sup>13-14</sup> WT-MetaD uses a history-dependent biasing potential on a few chosen coarse-grained degrees of freedom referred to collective variables (CVs). Suitable CVs help the systems to escape from the trap of the free energy minima and accelerate sampling of rare events. The well-tempered version of the metadynamics prevents overfilling of the free energy landscape and generally improves convergence of this technique.<sup>13</sup> Defining an appropriate set of CVs, which would allow the most relevant conformational changes to be described, is usually the most difficult challenge.<sup>15-16</sup>

RNA hairpins composed of a short loop at the tip of the Watson-Crick base-paired A-RNA stem are ideal testing systems for validation of force fields because of their small size and the significance of interactions other than Watson-Crick base pairs in stabilizing their structure. These abundant building blocks are indispensable for RNA folding<sup>17-18</sup> and play crucial roles in RNA-RNA and protein-RNA interactions. Hairpin loops containing four nucleotides in the loop region, denoted tetraloops (TLs), are notably stable and are often involved in tertiary contacts.<sup>17, 19-22</sup> Moreover, TLs play several biological roles in translation and transcription.<sup>17</sup> They have even been suggested to be involved in the emergence of spontaneous RNA catalysis in early stages of RNA prebiotic chemistry.<sup>23</sup> The most prominent families of TLs are 5'-GNRA-3' and 5'-UNCG-3' TLs (N stands for any nucleotide and R for purine). GNRA TLs can participate in protein-RNA complexes serving

as a recognition sites for RNA binding and are involved in many RNA-RNA and RNA-ligand interactions.<sup>24</sup>

Experimental results indicate that the RNA TL folding is more hierarchical compared to folding of small fast-folding proteins.<sup>25–26</sup> Two types of RNA TL folding mechanism have been suggested.<sup>27</sup> In the first model, the RNA chain collapses through non-specific base pairs to form a heterogeneous structural ensemble, followed by native-like or misfolded helix growth from the nucleation centre.<sup>27–28</sup> The second pathway is a zipping mechanism, in which the loop is closed by stable base pair, followed by subsequent zipping of the helical stem.<sup>27</sup> Despite their small size, a computational description of the folding of RNA TLs remains a challenge.<sup>29</sup> RNA TLs are precisely-shaped molecular building blocks with characteristic signature molecular interactions determining their consensus sequences.<sup>30</sup> The TL folding process results from a complex interplay of canonical base pairing, non-canonical interactions, stacking, solvation effects, backbone substates and electrostatic interactions between RNA and surrounding ions. The main obstacles to the accurate molecular description of the TLs folding are limited sampling in molecular simulations and limited accuracy of the force field.<sup>31</sup>

Current atomistic simulations of nucleic acids are still based on second-generation pair-additive force fields derived around twenty years ago.<sup>2–3</sup> There have been efforts to improve their performance by partial reparametrizations.<sup>32–38</sup> Most of these works attempted tuning of the uncoupled one-dimensional backbone dihedral potentials, which is the most straightforward refinement. These rather unphysical tweaks are used for final tuning of the force field once the other parameters are fixed, and we certainly cannot expect that they can provide a “perfect” force field. Rather, they are used to eliminate the most drastic force field failures, such as degradation of canonical helices. Regarding specifically RNA, one such refinement is the CHARMM36 tuning of the 2'-OH group dihedral parameters.<sup>39</sup> However, the CHARMM RNA force field still suffers from some under-stabilization of canonical A-RNA helices on the 10–100 ns time scale.<sup>40–41</sup> Thus, in contrast to good performance of CHARMM force field in the proteins and B-DNA simulations, variants of the AMBER Cornell et al. force field<sup>42</sup> are usually preferred for long RNA simulations. Due to the known principal limitations of pair-additive force fields, the efforts in CHARMM force field development has recently shifted to the promising design of a polarization force field, though its RNA version is not yet available.<sup>43</sup>

In the case of the AMBER nucleic acid force field, the 2007 bsc0 reparameterization<sup>35</sup> eliminated spurious  $\alpha/\gamma$  flips resulting in a progressive degradation of B-DNA structure during simulations. Although bsc0 is unnecessary for RNA simulations, longer simulations have proved that it also improves the RNA description.<sup>32, 40, 44</sup> In 2010 we identified a similar serious problem in RNA simulations, namely the formation of ladder-like structures in canonical A-form helices.<sup>45</sup> Subsequently we derived a reparameterization of the glycosidic torsion  $\chi_{OL3}$ <sup>32, 38</sup> which suppressed the ladder-like structures and became a part of the contemporary standard AMBER force field. Suppression of the RNA ladder-like structure required the DNA and RNA AMBER force fields to be separated, due to incompatible requirements on the  $\chi$  dihedral potentials in DNA and RNA. The bsc0 $\chi_{OL3}$  variant was tested on many RNA systems including RNA helices<sup>40, 44</sup> and RNA TLs.

<sup>29, 31–32</sup> No undesired side-effects for this specific modification have been found so far, though the force field as a whole of course remains far from “perfect”. Another reparameterization of the  $\chi$  torsion was independently suggested by Yildirim et al., though the authors were not aware of the ladder problem.<sup>36</sup> We have shown that this reparameterization also prevents the spurious ladder-like structures albeit with moderate side effects because the potential in the *anti* to high-*anti* region is too steep. This causes some flattening of the A-form helix by underestimation of inclination and roll parameters.<sup>32</sup> Subsequently, Yildirim et al. extended their version by including bsc0 and reparameterization of the  $\epsilon$ ,  $\zeta$ , and  $\beta$  torsions (known as AmberTOR<sup>46</sup>). At least in our tests, this complete dihedral reparameterization caused canonical A-form helices to deteriorate.<sup>47</sup> In 2013, Zgarbova et al. derived a DNA-specific reparameterization of  $\epsilon/\zeta$  torsions denoted  $\epsilon\zeta_{\text{OL1}}$  that significantly improved the description of B-DNA and DNA G-quadruplexes.<sup>37</sup> Subsequently we showed that  $\epsilon\zeta_{\text{OL1}}$  also significantly improved the description of non-canonical sugar-phosphate backbone conformation at the active site of hairpin ribozyme.<sup>34</sup> However, the general applicability of  $\epsilon\zeta_{\text{OL1}}$  to RNA has not yet been clarified. Additionally, our study also indicated that RNA simulations might be improved by implementation of modified Van der Waals parameters of the phosphate group, as suggested by Case et al.<sup>34, 48</sup> This was later supported by Cheatham et al. in REMD simulations of tetranucleotides.<sup>49</sup> An alternative reparameterization of ff99 AMBER force field has been suggested by Chen&Garcia (further referred as ff99Chen) during their efforts to achieve a reversible folding of RNA tetraloops.<sup>33</sup> It has included modification of van der Waals parameters of bases with additional atom-pair specific modification of the Van der Waals combination rules for water-base interactions using the so-called NBfix scheme. The main aim was to eliminate the known overestimation of the stacking interactions.<sup>50–51</sup> It has been accompanied by an adjustment of the  $\chi$  dihedral potential. The original T-REMD study reported multiple folding events to the folded state with correct signature interactions for two out of three studied RNA TLs, including a GNRA TL. A subsequent benchmark study by Cheatham et al. (multi-dimensional REMD with restrained A-RNA stem below the loop) confirmed that the ff99Chen force field leads to an improvement over the standard AMBER force field for the TLs. Specifically, the ff99Chen force field had clearly the highest population of the native UUCG loop conformation among all the tested force fields (~10% and ~1% at 277 K and 300 K, respectively), based on analysis of the signature interactions.<sup>52</sup> On the other hand, the ff99Chen force field may lead to excessive stabilization of some non-native base-pairs and flattening (i.e., loss of inclination) of A-RNA in some other systems, underlining the fact how difficult is to obtain a force field that would simultaneously reproduce all properties of RNA molecules.<sup>52–53</sup> Some studies also noted that nucleic acids simulations are somewhat sensitive to the choice of explicit water model,<sup>40, 44, 54</sup> though variations of water models are unlikely to resolve the limited accuracy of the primary solute force field, which originates from its inability to reliably describe inherent conformational preference of nucleic acids.<sup>55</sup> In any case, due to the above-noted limited capability to fine-tune the force fields by dihedral potential reparameterizations, future efforts to modify the AMBER non-bonded terms are expected.

In this work, we study the conformational dynamics and folding of a small RNA hairpin, namely 5'-GAGA-3' TL. Our study investigates the effect of the oligonucleotide length,

ionic conditions, choice of the sampling technique, force field, and the definition of the folded native state on the description of the folding free energy landscape. The main aim is to identify possible force field artefacts precluding correct theoretical description of the TL folding. In total, we report eight T-REMD simulations with a cumulative time of 896  $\mu\text{s}$ , one REST2 simulation with a cumulative time of 32  $\mu\text{s}$ , and two 3- $\mu\text{s}$ -long WT-MetaD simulations, i.e., almost 1 millisecond of simulations on aggregate.

## METHODS

### STARTING STRUCTURES

**The folded state.**—The folded state of GAGA TL was taken from the 1.04 Å resolution X-ray structure of the sarcin-ricin loop from *E. coli* 23S ribosomal RNA (PDB ID 1Q9A<sup>56</sup>, residues 2658–2663) and capped by an additional one or two CG base-pairs yielding the final sequences 5'-gcGAGAgc-3' or 5'-cgcGAGAgcg-3', respectively. The native structure of GAGA TL, shown in Fig. 1, is stabilized by three H-bonds:  $G_{L1}(N2)\cdots A_{L4}(pro-R_p)$  (3BPh interaction;<sup>57</sup> it can alternate with  $G_{L1}(N1/N2)\cdots A_{L4}(pro-R_p)$  bifurcated H-bonds 4BPh interaction during the simulation),  $G_{L1}(N2)\cdots A_{L4}(N7)$ , and  $G_{L1}(O2')\cdots G_{L3}(N7)$ . 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.

**The unfolded states.**—The initial structure of a single strand was built up using the Nucleic Acid Builder package of AmberTools14<sup>58</sup> as one strand of an A-form duplex. Subsequently, a starting ensemble of different replicas was prepared using classical 640-ns-long MD simulation of the single strand of the 8-mer and 10-mer at 300 K, initiated from the above-mentioned structure. The 64 unfolded structures were taken in 10 ns long intervals. The protocol of the classical MD simulation of the unfolded state is described in Supporting Information.

The starting topology and coordinates of folded and unfolded states of GAGA TL were prepared using the tLEaP module of AMBER 12 program package. The structures were solvated using a rectangular box of TIP3P<sup>59</sup> water molecules with a minimum distance between box walls and solute of 10 Å for the unfolded state and 15 Å for the folded state, respectively, yielding ~7000 water molecules added and  $\sim 60\times 60\times 60$  Å<sup>3</sup> box sizes. The simulations were performed in either ~150 mM KCl salt or in K<sup>+</sup> net-neutral conditions. We used the Joung-Cheatham ionic parameters<sup>60</sup> (K<sup>+</sup>:  $r = 1.705$  Å,  $\epsilon = 0.1937$  kcal/mol, Cl<sup>-</sup>:  $r = 2.513$  Å,  $\epsilon = 0.0356$  kcal/mol) in all simulations except for the T-REMD simulation with the ff99Chen force field,<sup>33</sup> where we used the suggested parameterization of ion-ion and ion-water interactions developed by Chen and Pappu.<sup>61</sup> In WT-MetaD simulations we used Na<sup>+</sup> net-neutral condition with Åqvist parameters<sup>62</sup> ( $r = 1.87$  Å and  $\epsilon = 0.0028$  kcal/mol). All structures were minimized and equilibrated using the standard equilibration protocol described in Supporting Information.

### SIMULATION SETTINGS

**T-REMD settings.**—The T-REMD simulations were carried out using the AMBER suite of programs with several combinations of nucleic acid force fields, namely using

ff99bsc0 $\chi_{OL3}$  force field<sup>32, 35, 38</sup> (standard AMBER ff10-ff14 force field for RNA, for simplicity henceforth denoted as  $\chi_{OL3}$ ), AMBER ff99 with the Chen&Garcia<sup>33</sup> modification (ff99Chen), and  $\chi_{OL3}$  with Van der Walls modification of phosphate oxygen developed by Case et al. for phosphorylated aminoacids (labelled as  $\chi_{OL3}CP$ ).<sup>48</sup> The Case phosphate parameters were implemented as introduced by Mlýnský et al.<sup>34</sup> including adjustment of torsion parameters affected by altered 1–4 Lennard-Jones interactions due to VdW modification, so that they resemble the original torsion profiles as closely as possible (AMBER library file of this force field version can be found in Supporting Information). We used 64 replicas in all T-REMD simulations. The temperatures of the 64 replicas spanned the range of 278–500 K and were chosen to maintain an exchange rate of ~25%. The T-REMD simulations were performed at constant volume (using the NVT ensemble in each replica), with long-range electrostatics calculated using PME with 1 Å grid spacing and 10 Å real-space cutoff. Langevin dynamics with a friction coefficient of 2 ps<sup>-1</sup> was used as a thermostat in all replicas, and the exchanges were attempted every 10 ps. Each replica was simulated at least for 1  $\mu$ s, while in cases, where we observed some folding events or in the simulation started from the folded state we extended the simulation time-scale to 2  $\mu$ s per replica (see Table 1). The total simulation time accumulated by all T-REMD simulations was 896  $\mu$ s.

**REST2 settings.**—The replica exchange solute tempering (REST2) simulation<sup>11</sup> was performed with 32 replicas at 300 K with the  $\chi_{OL3}$  force field and scaling factor ( $\lambda$ ) values ranging from 1 to 0.6057. Thus the effective solute temperatures ( $T/\lambda$ ) ranged from 300 to ~500 K, i.e., in the range corresponding to our T-REMD simulations. The starting structures were identical to the above-mentioned unfolded states for T-REMD simulations (replicas 33 to 64, see Figures S3, S4, and S5 in Supporting Information). The values of  $\lambda$  were chosen to maintain an exchange rate ~25%. Exchanges were attempted every 10 ps.

**Well-tempered metadynamics settings.**—The WT-MetaD simulations started from the native structure of 10-mer GAGA TL. At the beginning, the structure was equilibrated using a 5 ns NpT MD run with the Nosé-Hoover thermostat and isotropic Parrinello-Rahman (Andersen) barostat. After equilibration, two independent 3- $\mu$ s-long WT-MetaD simulations were performed with a 2 fs time step at a temperature of 300 K using two CVs. The first CV involved all native hydrogen bonds (called  $H_{core}$  CV) occurring in GAGA TL (both in stem and loop regions) and the second CV accounted for the all atom root-mean-square-deviation (RMSD) of the first ( $G_{L1}$ ), third ( $G_{L3}$ ), and fourth ( $G_{L4}$ ) nucleotides of the loop from the native state. The WT-MetaD simulation with the Gromacs code 4.5.5<sup>64</sup> was performed using the PLUMED 1.3 plugin.<sup>65</sup> This simulation setup corresponded to the settings published in ref.<sup>31</sup> and in fact one of the present simulations was a prolongation of the simulation published there.

**HBfix: local potential energy function selectively supporting native hydrogen bonding.**—To elucidate the influence of the strength of hydrogen bonds on folding of the GAGA TL, namely to test whether underestimated hydrogen bonding in the force field might affect propensity of the folding, we performed a set of T-REMD simulations with an additional locally-acting potential, increasing the stability of the base pairing and the

signature interactions. The additional potential, henceforth denoted HBfix, is applied to heavy-atom distances of the *native* hydrogen bonds. Thus, six hydrogen bonds are involved in base pairing in the stem of 8-mer TL together with further three signature interactions of the loop (see Supporting Information for more details of the settings of the restraints in AMBER simulations and list of used native hydrogen bonds).

This short-range potential is composed of two flat-bottom harmonic restraints with linear extensions having opposite signs of their curvatures (Figure 2). The slopes of the linear extensions of the restraints are equal except for the sign and thus cancel each other. As a consequence, these restraints sum into a constant everywhere except for a small region, where at least one of the flat-bottom restraints is quadratic. In our particular case, the additional potential is non-constant from 3 to 4 Å, where it climbs from 0 kcal/mol to  $\eta$  kcal/mol, where the  $\eta$  value is a controllable parameter stabilizing the given hydrogen bond. As noted above, the potential acts locally. Outside the interval 3 to 4 Å, it does not result in any forces and thus does not perturb the equilibrium distribution. It stabilizes the H-bonds once formed but does not promote the target interactions when the interacting atoms are not in a close contact. It results in a short-range native-interaction energy bias that aims to compensate under-stabilization of the H-bonding by the AMBER force field due to missing polarization effects.<sup>66</sup> We performed T-REMD simulations of 8-mer with  $\chi_{OL3}$  force field using two different values of the  $\eta$  parameter: a) 0.5 kcal/mol per hydrogen bond and b) 1.0 kcal/mol per hydrogen bond. Note that the simulations with a value of 0.5 kcal/mol did not lead to any visible increase of population of the folded state (the results of  $\chi_{OL3}HB_{0.5}$  simulation can be found in the Supporting Information). In addition, we used the same  $\eta = 1.0$  kcal/mol potential with  $\chi_{OL3}$ -CP force field (see Table 1). Apart from these additional HBfix potentials, the setup of these two T-REMD simulations was identical to the above-described protocol of other T-REMD simulations.

**Data analysis.**—The outputs of T-REMD and REST2 trajectories were sorted either to follow continuous trajectories or given temperatures, and Hamiltonians, respectively. All these trajectories were analysed with the Ptraj module of the AMBER package and the simulations were visualized using a molecular visualization program VMD.<sup>67</sup>

The estimations of error bars were calculated using bootstrapping.<sup>68</sup> The estimations of errors were obtained both by resampling time blocks in all 64 coordinate following replicas (identical time-blocks were resampled in all replicas simultaneously) as well as resampling this set of replicas used to finally obtain a resampled population at given temperature. In time domain we used a moving block bootstrap where optimal size of the blocks was selected to be close to  $n^{1/3}$ , where  $n$  is number of snapshots.<sup>69</sup> Subsequently, the replicas following coordinates were resampled (with replacement) and confidence interval with 5% level of significance was calculated from this resampled population estimated from 1000 resample trials. The resampling of time-blocks is typically used to estimate uncertainties originating from the limited sampling in time domain. However, this bootstrap implementation allows us in addition to distinguish between cases where the native state is formed only in one replica versus the native state being repeatedly formed in multiple replicas. Although in both cases one would observe both folded and unfolded structures in the reference replica (following the reference temperature), the latter scenario corresponds to



populations that can be expected to be significantly closer to convergence. The above-described resampling of replicas takes this effect into account. This analysis is partly inspired by the analysis of continuous replicas that has been used in other works.<sup>70</sup>

For identification of the native state, we used two metrics: (i) presence of all native hydrogen bonds, i.e., all hydrogen bonds participating on the base pairing in the stem as well as signature interactions of the loop (presence of a hydrogen bond was inferred from a heavy-atom donor-acceptor distance less than 3.5 Å) together with all atom root mean square deviation (RMSD) calculated over all nucleotides except A<sub>L2</sub> (A<sub>L2</sub> stacking/unstacking fluctuations were observed in native state ensemble both theoretically and experimentally, 29, 32, 71–72 so its dynamics should not be considered as disruption of the native folded state) with cutoff 3 Å, and (ii) eRMSD metric. The latter metric was also used for clustering, see below. eRMSD is a recently introduced metric used to measure deviation between RNA structures.<sup>73</sup> This metric is based on selected geometrical properties and provides reliable information about differences in the base interaction network. The eRMSD was calculated using the baRNAb package [<https://github.com/srnas/barnaba>].

In order to identify dominant conformations sampled in the course of T-REMD simulations at a reference temperature of 300 K, we used a cluster analysis based on a clustering algorithm introduced by Rodriguez and Laio<sup>74</sup> in combination with the eRMSD metric. The main idea of this algorithm is that cluster centres are characterized by a higher density than their neighbours, and by a relatively large distance from any other points with higher local density. In order to apply the algorithm using the eRMSD metric, we had to introduce a slight modification of cluster selection, namely definition of cluster hull representing noise spread around given cluster (see Supporting Information for more details).

## RESULTS AND DISCUSSION

Despite attempts to refine the RNA force fields (see the Introduction),<sup>2–3, 32, 37–38</sup> carefully performed MD studies continue to report persistent problems in description of many RNA molecules, including the smallest RNA systems such as tetranucleotides and tetraloops.<sup>29, 52</sup> Although recent reparameterizations are typically capable of correctly describing the native conformational basins as locally stable free energy minima (at least on the μs time-scale), robust sampling methods might indicate more complex imbalances in the force field if the native basin is not the global free energy minimum. The continuing problems in RNA simulations are not surprising considering the simplicity of the force field, large differences between force field and QM descriptions of nucleic acids,<sup>55</sup> and also the fact that essentially all refinements are still based on twenty-years-old second generation force fields.<sup>2–3</sup>

We aim to study the performance of selected current RNA force field versions for description of the conformational dynamics and folding of a small and structurally well-defined RNA motif, namely, the 5'-GAGA-3' TL. We assess the effects of the sequence length, ionic strength conditions, sampling technique, force field version, and definition of the native state. We want to separate problems caused by sampling from problems due to the force field, and identify how different force field terms affect the outcome of the simulations.

In total we analyzed eight T-REMD simulations with a cumulative time of 896  $\mu$ s, one REST2 simulation with cumulative time of 32  $\mu$ s, and two 3- $\mu$ s-long WT-MetaD simulations, i.e., 934  $\mu$ s of the simulation time in total (see Table 1).

### **Definition of the native state may have a dramatic impact on the reported population of the folded state.**

A key aspect in assessing the results is the definition of the native state. We strictly require that the TL must have not only the correctly paired WC stem, but also that all its signature interactions are properly structured. We classify the structures as misfolded TL when the stem is properly paired but the TL region is not properly structured. Note that some earlier studies considered the correct base pairing of the stem as sufficient to claim the TL folding while others counted even the spurious base zipper-like structures as folded.<sup>75</sup> When comparing with our stringent definition of native state, both later cases would be considered as “false-positives”. Thus, definition of the folded state has decisive impact on the reported populations of folded structures. Similarly, as shown in the literature, RMSD can often be an unreliable criterion to separate folded, unfolded and misfolded states.<sup>73</sup>

We are aware of the fact that our definition of the folded state might be considered as excessively conservative. However, the requirement for the presence of all signature interactions in the folded state is motivated by strict conservation of the GNRA TL conformation including preservation of all signature interactions in the structural database revealed by X-ray crystallography and NMR spectroscopy.<sup>76–78</sup> It is well established that, in general, biochemically relevant structures of recurrent RNA motifs such as the GNRA TLs are strictly defined.<sup>20, 30, 57, 79–81</sup> Indirect but essentially indisputable indication of that is the perfect correspondence between the signature interactions and the consensus sequences, known also as the isostericity principle of RNA evolution.<sup>57, 79</sup> It proves that the evolutionary pressure aims to achieve precisely structured UCG and GNRA TLs. Although the spectroscopic studies<sup>71–72, 82–83</sup> reported some inherent flexibility of the TL folded state, the majority of the alternative conformations suggested by these studies including, e.g., base-fraying of the terminal base-pair or unstacking of the L2 nucleotide from the tip of the TL, still maintain all the signature interactions. They thus fit to our definition of the folded state. The only conformation suggested by fluorescence studies that is not compatible with structural data is “5'-stacked” conformation having L2, L3, and L4 nucleotides stacked on the 5'-end of the stem.<sup>71–72</sup> However, such structure has never been reported by X-ray or NMR, and we also did not find any support for it in the present simulations. The absence of this conformation in the structural database cannot be explained by the limitations of X-ray crystallography and NMR spectroscopy due to high abundance of the GNRA TL in the database. In other words, the absence of 5'-stacked conformation cannot be explained by e.g. crystal packing, as the GNRA was resolved in numerous structural contexts and at least some of them are not affected by any crystal contacts. Therefore, we require in our definition of the folded state strict presence of the signature interactions suggested by available structural data. It is in line with current dominant view in RNA structural bioinformatics and structural biology.<sup>80</sup> As noted above, we admit that our definition of the folded state may be to certain extent considered too strict and specific for the MD technique. However, we can afford to use such definition due to the unlimited structural resolution of MD. Nevertheless,

we admit that our definition of folded state is not necessarily fully compatible with the criteria (order parameters) utilized by diverse experiments. Due to the indirect nature of the order parameters (measured signals) used in some experiments, it would not be trivial to find an exact fully quantitative correspondence between the ensembles sampled in the simulations and in the spectroscopic measurements.

### Sequence length affects convergence of T-REMD simulations.

The length (number of nucleotides) of the studied RNA hairpin might have significant effects on the propensity for folding observed in enhanced sampling simulations. Longer sequences (hairpins with more base pairs in the stems) are expected to be thermodynamically more stable. Therefore, assuming converged simulations, we should obtain a higher population of the folded state for longer sequences at the reference temperature. On the other hand, the unfolded state of the hairpin with longer sequence has a larger available conformational space. Thus, its folding events would face a larger entropic barrier compared to hairpins with shorter sequences. An increase of the entropic barrier can particularly complicate the T-REMD method, since the method does not improve sampling over entropic barriers and folding events are expected to be less frequent. This leads to slower convergence of simulations with longer stems and thus possible misinterpretation of the comparison between two simulations considering different sequence lengths.

In order to test the effects of the sequence length, we first performed 1- $\mu$ s-long T-REMD simulations using  $\chi_{OL3}$  force field and net-neutral conditions for the 8-mer (5'-gcGAGAgc-3') and 10-mer (5'-cgcGAGAgcg-3') TL hairpin sequences. Alternative folding simulations of these sequences were recently reported in literature.<sup>29, 33, 75</sup> It is worth to note that the  $G_{300K}$  of folding at 300 K estimated by Turner parameters<sup>84</sup> equals to -0.9 and -3.5 kcal/mol for the 8-mer and 10-mer, respectively. Such values would correspond to a rather minor difference in the expected population of the native state, namely 82.0% and 99.7% (see Table 2) in the case of the 8-mer and 10-mer, respectively. It is worth to note that the Turner parameters<sup>84</sup> were derived from the optical melting experiments relying on a folded state definition that is not straightforwardly related to the structural data. On the other hand, as discussed in the previous section, structural and bioinformatics data give a significant support for high stability of native TL conformation including all signature interactions, and thus it is likely that high populations estimated by Turner parameters correspond mostly to the native TL fold including signature interactions.

We have obtained two basic results. First, we did not observe a single folding event in the two T-REMD simulations of the 8-mer and 10-mer (we analyzed the folding events in terms of our definition of the folded state, see above; see Figures S1 and S2 in Supporting Information). This observation contrasts with our previously reported 500-ns-long simulation of the same 10-mer sequence, where we observed a single folding event (in one replica).<sup>29</sup> The difference might be explained by the stochastic nature of the simulations, and confirms our original conclusions that the previously observed folding event was fortuitous. In other words, the native state is rarely sampled on the (sub)microsecond time scale of T-REMD simulations with the given simulation settings, regardless of the sequence length.

The second observation was a major difference between the propensities of the two hairpins to reach the misfolded state with the properly paired stem and unstructured loop. While the 8-mer simulation revealed a clear minor population of the native stem base pairing (16 – 44 % at 300 K after bootstrapping, see Figure 3), the 10-mer simulation sampled mostly unstructured single stranded states with a complete absence of the folded stem with concurrent formation of all three base pairs. Thus, the 8-mer sequence apparently showed a faster convergence and all subsequent T-REMD runs were carried out using the 8-mer hairpin. Poor sampling of stem formation was also reported by Cheatham and co-workers in their recent REMD study on the UNCG TL. They then decided to restrain the stem as they were primarily interested in the structure of the apical TL part of the hairpin.<sup>52</sup>

### **The ionic conditions ( $K^+$ net-neutral vs. 1M-KCl) may have only a subtle effect on convergence of the T-REMD simulations.**

Sampling in folding simulations of RNA TLs may also be affected by ionic conditions. Due to finite size of the simulation box, the concentration of monovalent counter-ions is often above physiological concentration even in net-neutral conditions. Thus, net-neutralization should be sufficient to stabilize the negatively charged RNA backbone.<sup>40, 44</sup> On the other hand, the presence of divalent ions or very high concentration of monovalent ions might significantly affect both kinetics and thermodynamics of RNA folding.<sup>85</sup> Because of the known difficulties in modeling divalent ions with non-polarizable force fields,<sup>86</sup> we have limited our investigation to monovalent ions only.

In order to examine the effect of ionic conditions, we compared T-REMD simulations of 8-mer GAGA TL in  $K^+$  net-neutralizing condition (the simulation discussed in the previous paragraph) and 1M KCl excess salt. Similarly to the  $K^+$  net-neutral simulation, we did not observe any complete folding event even in the 1M KCl salt excess, thus the native state is not sampled (or might be sampled only rarely) in T-REMD simulations on the microsecond time-scale regardless of the ionic conditions. Nonetheless, the simulations revealed slightly different propensity of the stem population, however, bootstrapping analyzes showed that the statistical significance of this difference was disputable. In particular, the stem was populated to 16–44% in net-neutral condition, while a slightly diminished population of 4–21% was observed in 1M KCl salt excess at 300 K (see Figure 3).

Detailed analysis revealed that the population of paired stems originated dominantly from preservation of prefolded stems structures present in some replicas in our starting pool of conformations. In other words, we faced some starting structure bias and limited convergence of some replicas (see Figure S2 in Supporting Information). As explained in the Methods, the starting structures were taken in 10 ns intervals from a standard 640-ns-long MD simulation at 300 K. Although this simulation started from a straight single-stranded conformation, spontaneous temporary ~120-ns-long formation of the stem appeared along this trajectory and, consequently, 13 replicas (~20%) started from such misfolded conformations, which were annotated by clustering analyzes as four-purine stack (see the corresponding section below). Some of these replicas lose the paired stem rapidly, while some of them were trapped in it for the entire simulation likely because they occupied

mostly temperatures below 300 K, and their unfolding was thus slowed-down by low temperature.

The time-scales required for relaxation of this starting structure bias differed between the studied ionic conditions (see Figure S2 in Supporting Information). While 1M KCl salt excess simulations escaped from the four-purine stack misfolded state on tens-of-ns time-scale in all replicas, most of the replicas of the  $K^+$  net-neutral simulation required hundreds of nanoseconds to relax this starting structure bias and some of them even remained trapped in this state. Therefore, the high population of the folded stem mostly originated from the starting structure bias. The observed difference between net-neutral and 1M KCl ionic conditions can be dominantly attributed to the different life-time of the four-purine stack misfolded state in different ionic conditions. It illustrates the complexity of sampling in T-REMD simulations which require in-depth monitoring of the individual replicas; for further details see Figure S2 in Supporting Information which gives time courses of the stem and stem+loop populations in the replicas. Nevertheless, it should be emphasized that some replicas formed the stem (and in particular the four-purine stack structure) spontaneously from true essentially straight single-strand configuration, so in the T-REMD 8-mer runs we observed also unbiased events of stem zipping (see Figure S2 in Supporting Information). Thus, we captured spontaneous stem zipping – unzipping processes and the four-purine stack definitely belongs to the folding landscape, as described by the force field. Taking all the data together, when we disregard the uncertainty in the four-purine stack population, the ionic conditions seem to have only a minor effect on the outcome of the GAGA T-REMD simulations. Nevertheless, the higher degree of trapping of the four-purine stack in net-neutral T-REMD suggests that under this ionic condition, RNA might be somewhat frustrated by limited amount of counter-ions. Thus, we hypothesize that it might be less flexible due to trapping in local free-energy minima, i.e., more prone to get trapped in some structures. It is worth noting that this suggestions, indirectly derived from the T-REMD simulations, could be consistent with lower flexibility of RNAs under low ionic strength reported by Woodson et al.<sup>87</sup> However, we re-emphasize that our simulations are still not quantitatively converged and further simulations on much longer time scales would be needed to clarify and quantify the effect.

### **The RNA force field inaccuracy is likely responsible for unsatisfactory description of the GAGA TL folding.**

In the above-described T-REMD runs, we did not observe any single folding event yielding into formation of a complete folded GAGA TL, though we would like to reiterate that one such event has been captured in our earlier study.<sup>29</sup> When considering all currently available data<sup>29, 52</sup> we suggest that computer simulations of the GAGA TL folding are still far from providing data consistent with the experimental observations. Even the latest  $\chi_{OL3}$  force field is not sufficient to describe the GAGA TL folding from the unfolded state with a satisfactory population of the fully folded state, i.e., both with completely formed stem and loop. As noted in the Introduction, Chen and Garcia recently proposed variant of the AMBER force field with reparameterized van der Waals terms.<sup>33</sup> Using T-REMD simulations of three 8-mer TLs in 1M KCl salt conditions, including a different variant of GNRA TL, they reported dozens of spontaneous folding events.<sup>33</sup> This observation sharply

contrasts with the behavior of the RNA GAGA TL in AMBER  $\chi_{OL3}$  force field observed here and earlier,<sup>29</sup> although it should be noted that we use slightly different sequence of GNRA TL and different metric for definition of the folded state. The data present in the first part of our paper indicate that the difference between our simulations and those by Chen&Garcia can be attributed neither to the length of the hairpin sequence nor the ionic conditions. Therefore, it seems that the reparameterization of the van der Waals parameters might be vital for improvement of the description of the RNA hairpin folding. However, the ff99Chen force field has not been tested on any other RNA systems and subsequently other groups have demonstrated that these parameters may cause also undesired side effects for some other RNA systems like tetranucleotides, kissing loop complex or A-RNA duplexes. 52–53

To obtain further insight, we decided to compare the  $\chi_{OL3}$  and ff99Chen force fields for the 8-mer GAGA TL in 1M-KCl salt excess, using an identical metric of folding. We did not see frequent re-folding events comparable to those reported by Chen and Garcia. We carefully checked the force field settings to rule out that this discrepancy might originate from an erroneous implementation of the force field parameters. It should be, however, noted that here we used slightly different sequence of the TL, namely 5'-GAGA-3' instead of 5'-GCAA-3' used in the original study. Thus, a direct comparison of the present simulations with the preceding work is still to some extent limited and further comparisons would be desirable. Nevertheless, as can be seen in Figure 3, the ff99Chen parameters significantly increased the stability of the stem, but not the stability of the complete stem-loop structure. We observed many events of RNA stem zipping, i.e., formation of GC base pairs, but typically without proper formation of the native conformation of the GAGA loop region (defined strictly by its signature interactions, see Methods). At 300 K, we observed 0.0–0.7% and 0% population of the fully folded native state with ff99Chen and  $\chi_{OL3}$ , respectively (see Figure 3B). The stem was populated to 55–76% and 4–21%, respectively (see Figure 3A), confirming that ff99Chen force field has much better propensity to fold the stem.

### **The reported differences in the literature may be affected by different definitions of the folded state.**

In order to explain the significant discrepancy between various T-REMD runs discussed above, we focused on the description of the folded state. In our study we defined native (folded) state by simultaneous fulfillment of several conditions: i) presence of all signature interactions in the loop region, ii) presence of all hydrogen bonds participating on base-pairing in the stem, and iii) all-atom RMSD calculated over all nucleotides except the  $A_{L2}$  below 3 Å. We excluded  $A_{L2}$  from the RMSD calculation as it does not participate in any signature interaction of the TL and its reversible (un)stacking is an inherent part of structural dynamics of the folded TL.<sup>29, 32, 71–72, 82, 88</sup> This allows us to use a rather strict cutoff of 3 Å for the rest of the structure. This cut-off is justified based on the observed fluctuations in T-REMD simulations initiated from the native state (in which TL remain folded), which fluctuate in the RMSD range 0.7 to 2.2 Å. In contrast, Chen and Garcia defined the native state by 4 Å cutoff of all heavy-atom RMSD from the native state. We thus reanalyzed our simulations using all-atom RMSD with cutoff 4 Å as the only criterion. We found that the

population of the folded state (within this simplified RMSD definition) significantly increased, namely to 12% and 38% in  $\chi_{OL3}$  and ff99Chen 1M KCl 8-mer simulations, respectively. We conclude that, although the ff99Chen reparametrization significantly improves the stacking interaction, and stabilizes the native form of the stem, the stability of the tetraloop might be spuriously sensitive to the metric used to define the native state. A similar observation has been made by Cheatham et al. for the UNCG TL.<sup>2</sup> In addition, the conclusions based on RMSD metric can be misleading when important parts of the structures are excluded from the overall RMSD calculation. For example, all our above-mentioned T-REMD simulations revealed significant population with RMSD around  $\sim 2$  Å when the RMSD was calculated only over heavy atoms of the stem and backbone atoms of the loop regions, respectively (see Figure S3 in the Supporting Information).

We would like to emphasize that all of the above analyses can be directly applied only to the simulations present in our work, since we could not analyze the trajectories obtained by the other groups.

The inadequacy of the simple RMSD metric is also discussed in the recent literature.<sup>2, 73</sup> Bottaro et al. suggested an alternative metric denoted  $\epsilon$ RMSD that is based on mutual position of coarse-grained nucleobases. This metric is significantly more robust compared to RMSD. Notably, in the case of 10-mer TL, even the spurious ladder-like structure<sup>32, 45</sup> is  $\sim 3.8$  Å in RMSD from the native conformation.<sup>29</sup> The  $\epsilon$ RMSD is able to clearly distinguish this spurious state as a non-native conformation. The ladder-like structure of the TL gives a large  $\epsilon$ RMSD of  $\sim 1.2$  while structurally similar RNA conformations are typically characterized by  $\epsilon$ RMSD below 0.5–0.7. We thus analyzed our T-REMD simulations using the  $\epsilon$ RMSD metric. The populations of the native state revealed by  $\epsilon$ RMSD metric is similar with populations obtained by our definition of the native state using full set of native hydrogen bonds in combination with the RMSD criterion. However, analysis of T-REMD simulations using  $\epsilon$ RMSD metric is significantly simpler compared to the cumbersome definition of the native state by signature interactions. Therefore, we suggest that  $\epsilon$ RMSD metric is a promising tool for analysis of nucleic acid simulations combining a simplicity comparable to the RMSD metric with significantly improved accuracy of identification of the native state from the non-native ones.

### **$\epsilon$ RMSD-based clustering.**

Motivated by the performance of the  $\epsilon$ RMSD metric in identification of the structural similarities, we performed a clustering analysis of all 300K replicas from the 8-mer T-REMD simulations using this metric.

We found that the hallmark of all T-REMD simulations starting from the unfolded state (i.e., all 8-mer T-REMD simulations, except the one starting from the folded state) was a significant population of the four-purine stack cluster. This was characterized by properly folded stem and continuous stacking of all four bases of the loop region as a 3'-overhang capping the stem (see Figure 4, which also includes simulations with force field modifications that are introduced later in the text). As detailed above, structures identified as four-purine stack cluster occurred already in starting structures of 13 replicas, i.e.,  $\sim 20\%$  of total pool of the starting structures (see replicas 31–44 in Figure S4 in Supporting

Information). This is because this state temporarily appeared during the MD simulation used for seeding the starting structures for the T-REMD simulations. Thus, high population of such structures during T-REMD runs could be affected by their presence in the initial ensemble (see Figure S4 in Supporting Information for time development of the clusters in coordinates following replicas). Nonetheless, although the populations might be affected by limited convergence and starting structure bias, we can still conclude that the four-purine stack is actually stabilized by the force field for two reasons. It was spontaneously formed in the classical simulation started from a straight single strand (used for seeding the starting conformations of the subsequent T-REMD simulations) and subsequently populated in T-REMD replicas. In addition, most importantly, we also observed several spontaneous folding events into such four-purine conformations during the T-REMD simulations from different conformations. In other words, we captured spontaneous unfolding and refolding of this particular conformation, i.e., there is an exchange with the unfolded ensemble in both directions. Thus, although the population of four-purine stack structure may be quantitatively overestimated due to their extensive presence in the set of starting replicas, the tendency of the force field to populate this conformation is indisputable.

Aside from the four-purine stack, almost all T-REMD simulations (except simulations stabilizing the native base pairs using the HBfix 1kcal/mol, introduced below) significantly populated the single stranded structures, i.e., structures lacking stem base pairing. Obviously, the ensemble of such structures is very diverse, so all clusters annotated as single stranded structures were sorted into two groups. In particular, single stranded structures stabilized by a wide variety of intramolecular interactions, mainly base-phosphate and sugar-phosphate (hydrogen bonding of 2'-OH with phosphate non-bridging oxygens) interactions were annotated as intermolecularly stabilized ssRNA (henceforth denoted as stabilized-ssRNA cluster, see Figure 4). The rest of single stranded structures, i.e., all structures having only stacking interactions between bases and lacking base pairing, base-phosphate and sugar-phosphate interactions, were further classified as the plain-ssRNA cluster (see Figure 4). The plain-ssRNA cluster includes not only A-form-like ssRNAs but also ssRNA structures with bulged bases. In addition, a prominent substate of the intramolecularly stabilized ssRNA state is 5'-intercalated cluster (see Figure 4), where the last two nucleotides at the 3'-terminus adopt a backbone conformation classified as 5d suite by Richardson nomenclature<sup>89</sup> while the 5'-terminal nucleobase is intercalated between them. It closely resembles a common simulation artifact observed for tetranucleotides.<sup>52, 90–91</sup> The last annotated misfolded cluster forms a stem, but in left-handed Z-form instead of A-form helix conformation and includes both stem guanines in *syn* orientation (Figure 4). Consistent with the other analyses, both  $\chi_{OL3}$  simulations (net-neutral and 1M KCl salt excess) as well as the ff99Chen simulation reveal negligible population of the native folded state (below 1%, with zero population within error bars calculated by bootstrapping).

Finally, we would like to note that we did not see any sign of ladder-like structure in any of our simulations, despite that all simulations were specifically monitored to detect such structures. This observation, together with several earlier-reported spontaneous reparations of ladders in  $\chi_{OL3}$  simulations of short duplexes,<sup>32</sup> suggests that the formation of the ladder-like structures in  $\chi_{OL3}$  force field documented by Cheatham et al. for UUCG TL at 277 K M-REMD reference replica might deserve a further analysis.<sup>52</sup>



### Several enhanced-sampling techniques (T-REMD, REST2 and WT-MetaD) point to the same results.

Due to the richness of the RNA conformational space, results of all simulation techniques, including enhanced sampling techniques, generally suffer from limited sampling. It seems to be a difficult task to achieve the ultimate convergence of the simulations even for small RNA systems.<sup>49, 52</sup> Thus it is worthwhile to carefully analyze the convergence of the simulations and preferably use several independent techniques to obtain data that are as robust as possible.

The basic insight into estimation of accuracy achieved by the present level of sampling can be obtained by bootstrapping. Besides the standard block bootstrapping, where time-block data (at a given temperature) are resampled, we also resampled the set of coordinate-following replicas, from which the trajectory at given temperature is mixed (see Methods). Thus, the error estimations include not only inaccuracies originating from the limited time-scale, but also take into account inhomogeneities of the sampling among coordinate-following replicas. Note that this bootstrap analysis was able to reveal insignificance of the difference in folded stem population between 1M KCl salt-excess and  $K^+$  net-neutral simulations that originated from the different level of relaxation of the starting structure bias (see above).

An ultimate requirement of the converged populations of the native state would be having identical results obtained by T-REMD simulations starting from fully unfolded and fully folded states. Thus, we performed another T-REMD simulation of GAGA TL using the  $\chi_{OL3}$  force field and  $K^+$  net-neutral condition starting from the folded state. The population of the native state at 300 K progressively decreased during the simulation and was never recovered after unfolding (see Figure S2 in Supporting Information). Therefore, we extended this simulation up to 2  $\mu$ s per replica until the native structures were completely lost in all replicas and the population (at all temperatures/replicas) of the native state dropped to zero (note that Figure 4 reports total population calculated from the entire simulation time-scale). We cannot rule out that at longer time-scale we would be able to observe occasional short transient refolding events, as reported earlier.<sup>29</sup> However, the result is in clear qualitative agreement with the T-REMD simulations started from the unfolded state. All these T-REMD simulations together suggest that the canonical population of the native state under the employed force field description is negligible. In other words, the  $\chi_{OL3}$  force field can reach and temporarily stabilize the native state of the GAGA TL but does not capture it as the most favorable conformational basin. In fact, our data suggest that the same applies for the ff99Chen force field. Thus, all currently available force fields fail to describe the correct global free energy minimum of this TL.

In order to obtain an alternative insight into the folding propensity of the GAGA TL in  $\chi_{OL3}$  force field, we performed replica exchange solute tempering (REST2) simulation (see Methods). The REST2 run did not reveal any re-folding transitions (see Figure S5 in Supporting Information), supporting the conclusions based on the T-REMD simulations.

Finally, we also performed two independent well-tempered metadynamics simulations (WT-MetaD) of the 10-mer GAGA TL, where sampling of unfolding/re-folding dynamics was

enhanced using two carefully chosen collective variables (CVs). The same CVs were used in our recent study<sup>31</sup> where we showed that these CVs are able to guide the simulation through reversible unfolding/re-folding events. Both 3- $\mu$ s-long simulations resulted in qualitatively similar free energy landscape with clearly distinguishable free energy minima of single stranded, misfolded and native folded states (see Figure 5 and Figure S6 documenting structural dynamics in course of the WT-MetaD trajectories). Even our relatively long WT-metaD simulations are not able to provide rigorously converged estimates of the relative free energies of the different states. Nonetheless, they are in qualitative agreement with the T-REMD and REST2 simulations and suggest that the native folded state of the GAGA TL is not the global minimum under the force field description. The free energy of the native state was estimated by reweighting scheme to be  $\sim$ 3–4 kcal/mol above ssRNA and misfolded state (see Figure 5). It should be noted that while we speculated in our recent study<sup>31</sup> that the poor sampling of refolding events in WT-metaD simulations was caused either by still too simplified CVs or force field inaccuracies (or combination of both), the present data clearly suggest that the force field is not able to correctly describe GAGA TL folding.

Although all used enhanced sampling techniques, including simulations initiated from the folded state, point to the same conclusion, we would like to reiterate that none of the presented simulations is fully converged and our simulations could not reach steady state behavior.

### **Force fields may thermodynamically bias the simulations in favor of the ssRNA state by over-stabilization of BPh and sugar-phosphate interactions.**

As mentioned above, the hallmark of the T-REMD simulations of GAGA TL is a significant population of the intramolecularly stabilized ssRNA state involving structures lacking base pairing but stabilized by intramolecular base-phosphate and sugar-phosphate interactions. Therefore, we tested the effect of the van der Waals reparameterization of the phosphate's oxygen developed by Case et al.<sup>48</sup> This reparameterization, albeit originally developed for phosphorylated aminoacids, was shown in our recent study to improve description of the structural dynamics of the hairpin ribozyme active site.<sup>34</sup> Subsequently, Cheatham et al. showed that it also partially improved structural dynamics of tetranucleotides and UUCG TL.<sup>52</sup>

Here we used an implementation of these parameters described by Mlýnský et al.,<sup>34</sup> where VdW parameters of both bridging and non-bridging oxygens were modified, and all corresponding torsions were adjusted to take into account the effect of the modified 1–4 Lennard-Jones interactions (AMBER library file of this force field version can be found in Supporting Information). T-REMD 8-mer simulation using the combination of  $\chi_{OL3}$  force field with the phosphate reparameterization was able to sample three folding events in a few replicas (see Figure S2 in Supporting Information), yielding a native state population comparable to the ff99Chen simulation (Figures 3 and 4). The initial success motivated us to extend this simulation up to 2  $\mu$ s per replica. However, as in the ff99Chen simulation, bootstrap analysis showed that the population of the native state was still not significant (based on error bars representing the confidence interval calculated by bootstrapping at 5% level of significance, see Methods) at the present level of sampling; see error bars in the

Figure 3 that reach zero population at all temperatures in both ff99Chen and  $\chi_{OL3}CP$  simulations (note that zero population was observed in case of all other above-mentioned  $\chi_{OL3}$  simulations). This suggests that the phosphate reparameterization represents a step in the right direction but is not sufficient to radically change the force field performance. The results also suggest that excessive stabilization of stacking, which is corrected by the ff99Chen modifications, may only be part of the reason behind the poor sampling of the folded TL state with current force fields. Over-stabilization of the base-phosphate and sugar-phosphate interactions in the unfolded state may also be important.

### **Local structure-specific biasing potentials might compensate for the effect of underestimated stability of native hydrogen bonds.**

Another potential artifact of the empirical force fields that might preclude correct description of the RNA folding is underestimation of hydrogen bonding.<sup>50, 92</sup> In fact, part of the stem stabilization observed with the ff99Chen force field could be the increased stability of base pairing interactions, in addition to weakening of the stacking. Although the overstacking is a well-documented problem of contemporary force fields,<sup>50, 92–93</sup> a uniform rescaling of the VdW radii of the nucleobase atoms is inevitably accompanied by stabilization of base pairing, because the polar atoms involved in hydrogen bonding can come closer to each other, leading to a stronger electrostatic interaction. This is evident in the shorter heavy atom distances in hydrogen bonding interactions in the simulations with the ff99Chen force field (see Figure S7 in the Supporting Information), which might be an unintended consequence of the modification.

Considering the above results, we have attempted to stabilize the H-bonds with an additional term in the force field as a proof of principle. Instead of trying a global modification of the force field, which may be difficult to parameterize without incurring in undesirable side effects, we stabilized the native interactions locally and in a structure-specific manner. We aimed to strengthen hydrogen bonding interactions as gently as possible, to make minimal bias of the other parts of the folding landscape. We introduced a special additional HBfix potential affecting only a narrow range of heavy atom distances (between 3 and 4 Å, and supporting the hydrogen bonded state by simple potential energy bias as explained in Methods and Figure 2). Usage of 1 kcal/mol local bias per hydrogen bond applied to all native hydrogen bonds in the stem and loop region resulted in significantly increased population of the folded native state (Figures 3 and 4). Note, that in contrast to the other simulations, the error bars in this case did not reach zero and thus the observations are statistically significant over a wide range of temperatures (see Supporting Information for alternative T-REMD simulation using bias of 0.5 kcal/mol per hydrogen bond). Note that a value of 1 kcal/mol per hydrogen bond supporting the base pairing (3 kcal/mol for each GC base pair) approximately corresponds to the expected underestimation of the intrinsic hydrogen bonding interaction energy in the Cornell et al. based force fields, relative to quantum chemistry calculations.<sup>50</sup>

The slight improvement of the folding event sampling achieved with the Case et al.'s phosphate modification, and the significant effect of the HBfix potential motivated us to test these two modifications together. We hypothesize that while the phosphate's modification

destabilizes unfolded state of GAGA TL, the HBfix stabilizes the folded state, so a cooperative effect might be expected. Indeed, we observed that the population of the folded native state in 2- $\mu$ s-long  $\chi_{OL3}$ CP-HB simulation was significantly higher than the sum of the native state populations of  $\chi_{OL3}$ CP and  $\chi_{OL3}$ -HB simulations (see Figure 3). Thus we can conclude that these modifications act in a cooperative manner.

Structure-specific modifications targeting the native interactions are obviously not fully satisfactory. They require a priori knowledge of the native structure to be embedded into the definition of the potential energy function, i.e., they cannot be part of a transferable force field. Here we use them primarily to demonstrate the impact of an improved hydrogen bonding term on force field accuracy. On the other hand, when satisfactory reparametrization of the general force field is difficult, structure-specific modifications can be useful in studies of specific biomolecular systems.

We have also tried to estimate the effect of inclusion of non-native base-base hydrogen bonds via the same potential by reweighting. It does not alter the qualitative results, namely, after reweighting the native state would remain populated to 9.3% and 7.8% in  $\chi_{OL3}$ HB-1M and  $\chi_{OL3}$ CP-HB-1M simulations, respectively; cf. with the data in Figure 3. Our results thus show that introduction of a modest hydrogen bond correction significantly improves the population of the native state. This suggests that in future work it may be desirable to optimize a specific short-range hydrogen bond term, which would be mimicking the missing polarization effects accompanying hydrogen bonding. Such an optimization, which is beyond the scope of this work, could draw on existing quantum chemistry calculations, as well as experimental data for model systems in solution.

## CONCLUSIONS

We performed a series of enhanced sampling simulations of the GAGA RNA TL testing several force fields, simulation settings and different kinds of enhanced sampling techniques in their performance of the description of the folding process.

Despite the microsecond time-scale, we found that none of the utilized enhanced sampling techniques (T-REMD, REST2, or WT-MetaD) was able to converge the sampling of the GAGA TL conformational ensemble and thus the reported simulations could not reach a steady state behavior. In addition, it should be noted that simulations using different force fields, involving a different sequence length, or having different ionic conditions might converge on a different time-scale, for example, due to their different propensity to form the RNA stem. However, different techniques and simulations settings used in our study pointed to the same qualitative conclusions, which allowed us to qualitatively estimate the stability of GAGA native folded state in different conditions (force field, sequence length and ionic conditions).

Our simulations confirmed the observation suggested in recent literature<sup>29, 31, 33, 49, 52</sup> about the inability of the  $\chi_{OL3}$ <sup>35, 38, 42</sup> force field to correctly describe balance between folded and unfolded states of the GAGA TL. The  $\chi_{OL3}$  force field provides reasonable description of the folded state basin and consequently stable MD simulations of the TL on several- $\mu$ s time

scale as repeatedly shown in literature for many other RNA systems.<sup>31, 41, 53, 94–95</sup> However, the enhanced sampling techniques unambiguously reveal that the folded state of TL does not correspond to the global free energy minimum described by the force field. The simulations were able to form the correctly-paired A-RNA stem, but not the native TL structures. This means that we might face further problems in the force field, which do not appear to be correctable by refinements of the dihedral parameters. Thus, although the  $\chi_{OL3}$  force field appears to do a very good job in simulations that start from known atomistic experimental structures of folded RNAs, it is not sufficiently robust to study RNA folding landscapes.

We have also tested the recently introduced modification of the van der Waals parameters by Chen&Garcia.<sup>33</sup> However, for the presently studied system, these modifications did not improve the folding landscape when considering the structure of the entire stem-loop.

When comparing different simulations we found that the common RMSD criteria of the folded state can be misleading and we strongly suggest monitoring the TL folding either with a full set of native interactions, or using the recently introduced eRMSD metric.<sup>73</sup> By specific examples we demonstrate how not-sufficiently-stringent definitions of the folded state may bias the interpretation of the folding simulations of RNAs.

One of the main aims of our work was to identify the possible sources of the force field artifacts that preclude correct description of the folded-unfolded balance of the TL. We show that at least two different effects likely contribute to incorrect folded/unfolded free energy balance of the TL, namely, excessive stabilization of the unfolded ssRNA state by intramolecular base-phosphate and sugar-phosphate interactions, and destabilization of the native folded state by underestimation of the native hydrogen bond interactions, including the stem base pairing. The former problem can be partially alleviated by alternative phosphate's oxygen van der Waals parameters developed by Case et al..<sup>48</sup> Although this Van der Waals term modification cannot be considered as a breakthrough in the RNA simulations, its positive effect on RNA simulations was shown by us and others also for other RNA systems.<sup>34, 49</sup> Until now it does not seem to be associated with any side effects. On the other hand, a generalized correction for underestimation of base-base hydrogen bonding (not associated with some side effects for other interactions) remains a challenge. We have nevertheless obtained substantially improved folding simulations of the GAGA TL using weak local structure-specific biases supporting the native interactions. In addition, we found that modification of hydrogen bonding acts cooperatively with the refined phosphate's parameters. We suggest that due to the persistent and principal limitations of the RNA force fields, such system-specific corrections represent a viable option to obtain stable native structures, with minimal perturbation to the transferable force field. In the long run, it would be desirable to develop an additional force field term for hydrogen bonding which captures this effect in a transferable manner.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENT

This work was supported by grant P208/12/1878 (J.S., M.O., P.K.) from the Czech Science Foundation. Further institutional funding was provided by project LO1305 of the Ministry of Education, Youth and Sports of the Czech Republic (M.O., P.B., P.K.). JS acknowledges support by Praemium Academiae. R.B. was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health. G.B and S.B received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013) / ERC Grant Agreement n. 306662, S-RNA-S.

## REFERENCES

1. Ditzler MA; Otyepka M; Sponer J; Walter NG, Molecular Dynamics and Quantum Mechanics of Rna: Conformational and Chemical Change We Can Believe In. *Accounts Chem Res* 2010, 43, 40–47.
2. Sponer J; Banas P; Jurecka P; Zgarbova M; Kuhrova 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. [PubMed: 26270382]
3. Cheatham TE; Case DA, Twenty-Five Years of Nucleic Acid Simulations. *Biopolymers* 2013, 99, 969–977. [PubMed: 23784813]
4. Best RB, Atomistic Molecular Simulations of Protein Folding. *Curr Opin Struc Biol* 2012, 22, 52–61.
5. Freddolino PL; Harrison CB; Liu YX; Schulten K, Challenges in Protein-Folding Simulations. *Nat Phys* 2010, 6, 751–758. [PubMed: 21297873]
6. Zhang T; Nguyen PH; Nasica-Labouze J; Mu YG; Derreumaux P, Folding Atomistic Proteins in Explicit Solvent Using Simulated Tempering. *J Phys Chem B* 2015, 119, 6941–6951. [PubMed: 25985144]
7. Pan AC; Weinreich TM; Piana S; Shaw DE, Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J Chem Theory Comput* 2016, 12, 1360–1367. [PubMed: 26866996]
8. Abrams C; Bussi G, Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy-Switz* 2014, 16, 163–199.
9. Banas P; Jurecka P; Walter NG; Sponer J; Otyepka M, Theoretical Studies of Rna Catalysis: Hybrid Qm/Mm Methods and Their Comparison with Md and Qm. *Methods* 2009, 49, 202–216. [PubMed: 19398008]
10. Sugita Y; Okamoto Y, Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem Phys Lett* 1999, 314, 141–151.
11. Wang L; Friesner RA; Berne BJ, Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (Rest2). *J Phys Chem B* 2011, 115, 11305–11305.
12. Liu P; Kim B; Friesner RA; Berne BJ, Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *P Natl Acad Sci USA* 2005, 102, 13749–13754.
13. Barducci A; Bussi G; Parrinello M, Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* 2008, 100, 020603. [PubMed: 18232845]
14. Laio A; Parrinello M, Escaping Free-Energy Minima. *P Natl Acad Sci USA* 2002, 99, 12562–12566.
15. Barducci A; Bonomi M; Parrinello M, Metadynamics. *Wires Comput Mol Sci* 2011, 1, 826–843.
16. Bussi G; Branduardi D, Free-Energy Calculations with Metadynamics: Theory and Practice. John Wiley and Sons, Inc, Hoboken, NJ, USA: 2015; Vol. 28.
17. Varani G, Exceptionally Stable Nucleic-Acid Hairpins. *Annu Rev Bioph Biom* 1995, 24, 379–404.
18. Uhlenbeck OC, Nucleic-Acid Structure - Tetraloops and Rna Folding. *Nature* 1990, 346, 613–614. [PubMed: 1696683]
19. Pley HW; Flaherty KM; Mckay DB, Model for an Rna Tertiary Interaction Front the Structure of an Intermolecular Complex between a Gaaa Tetraloop and an Rna Helix. *Nature* 1994, 372, 111–113. [PubMed: 7526219]

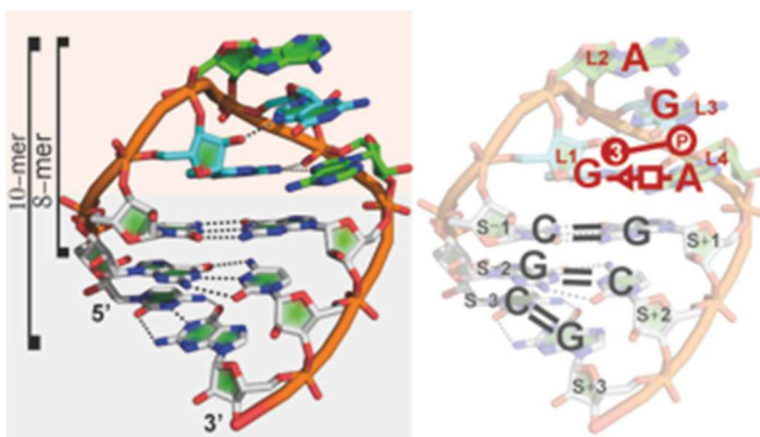
20. Xin YR; Laing C; Leontis NB; Schlick T, Annotation of Tertiary Interactions in Rna Structures Reveals Variations and Correlations. *Rna* 2008, 14, 2465–2477. [PubMed: 18957492]
21. Chauhan S; Woodson SA, Tertiary Interactions Determine the Accuracy of Rna Folding. *J Am Chem Soc* 2008, 130, 1296–1303. [PubMed: 18179212]
22. Marino JP; Gregorian RS; Csankovszki G; Crothers DM, Bent Helix Formation between Rna Hairpins with Complementary Loops. *Science* 1995, 268, 1448–1454. [PubMed: 7539549]
23. Stadlbauer P; Sponer J; Costanzo G; Di Mauro E; Pino S; Sponer JE, Tetraloop-Like Geometries Could Form the Basis of the Catalytic Activity of the Most Ancient Ribooligonucleotides. *Chem-Eur J* 2015, 21, 3596–3604. [PubMed: 25640446]
24. Jaeger L; Michel F; Westhof E, Involvement of a Gnra Tetraloop in Long-Range Tertiary Interactions. *J Mol Biol* 1994, 236, 1271–1276. [PubMed: 7510342]
25. Brion P; Westhof E, Hierarchy and Dynamics of Rna Folding. *Annu Rev Bioph Biom* 1997, 26, 113–137.
26. Antao VP; Tinoco I, Thermodynamic Parameters for Loop Formation in Rna and DNA Hairpin Tetraloops. *Nucleic Acids Res* 1992, 20, 819–824. [PubMed: 1371866]
27. Sorin EJ; Rhee YM; Nakatani BJ; Pande VS, Insights into Nucleic Acid Conformational Dynamics from Massively Parallel Stochastic Simulations. *Biophys J* 2003, 85, 790–803. [PubMed: 12885628]
28. Chakraborty D; Collepardo-Guevara R; Wales DJ, Energy Landscapes, Folding Mechanisms, and Kinetics of Rna Tetraloop Hairpins. *J Am Chem Soc* 2014, 136, 18052–18061. [PubMed: 25453221]
29. Kuhrova P; Banas P; Best RB; Sponer J; Otyepka M, Computer Folding of Rna Tetraloops? Are We There Yet? *J Chem Theory Comput* 2013, 9, 2115–2125.
30. Hsiao C; Mohan S; Hershkovitz E; Tannenbaum A; Williams LD, Single Nucleotide Rna Choreography. *Nucleic Acids Res* 2006, 34, 1481–1491. [PubMed: 16531589]
31. Haldar S; Kuhrova P; Banas P; Spiwok V; Sponer J; Hobza P; Otyepka M, Insights into Stability and Folding of Gnra and Uncg Tetra Loops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics. *J Chem Theory Comput* 2015, 11, 3866–3877. [PubMed: 26574468]
32. Banas P; Hollas D; Zgarbova M; Jurecka P; Orozco M; Cheatham TE; 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.
33. Chen AA; Garcia AE, High-Resolution Reversible Folding of Hyperstable Rna Tetraloops Using Molecular Dynamics Simulations. *P Natl Acad Sci USA* 2013, 110, 16820–16825.
34. Mlynsky V; Kuhrova P; Zgarbova M; Jurecka P; Walter NG; Otyepka M; Sponer J; Banas P, Reactive Conformation of the Active Site in the Hairpin Ribozyme Achieved by Molecular Dynamics Simulations with Epsilon/Zeta Force Field Reparametrizations. *J Phys Chem B* 2015, 119, 4220–4229. [PubMed: 25692537]
35. Perez A; Marchan I; Svozil D; Sponer J; Cheatham TE; Laughton CA; Orozco M, Refinement of the Amber Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys J* 2007, 92, 3817–3829. [PubMed: 17351000]
36. Yildirim I; Stern HA; Kennedy SD; Tubbs JD; Turner DH, Reparameterization of Rna Chi Torsion Parameters for the Amber Force Field and Comparison to Nmr Spectra for Cytidine and Uridine. *J Chem Theory Comput* 2010, 6, 1520–1531. [PubMed: 20463845]
37. Zgarbova M; Luque FJ; Sponer J; Cheatham TE, 3rd; Otyepka M; Jurecka P, Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput* 2013, 9, 2339–2354. [PubMed: 24058302]
38. Zgarbova M; Otyepka M; Sponer J; Mladek A; Banas P; Cheatham TE; 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. [PubMed: 21921995]
39. Best RB; Zhu X; Shim J; Lopes PEM; Mittal J; Feig M; MacKerell AD, Optimization of the Additive Charmm All-Atom Protein Force Field Targeting Improved Sampling of the Backbone Phi, Psi and Side-Chain Chi(1) and Chi(2) Dihedral Angles. *J Chem Theory Comput* 2012, 8, 3257–3273. [PubMed: 23341755]

40. Besseova I; Banas P; Kuhrova P; Kosinova P; Otyepka M; Sponer J, Simulations of a-Rna Duplexes. The Effect of Sequence, Solute Force Field, Water Model, and Salt Concentration. *J Phys Chem B* 2012, 116, 9899–9916. [PubMed: 22809319]
41. Banas P; Sklenovsky P; Wedekind JE; Sponer J; Otyepka M, Molecular Mechanism of Preq(1) Riboswitch Action: A Molecular Dynamics Study. *J Phys Chem B* 2012, 116, 12721–12734. [PubMed: 22998634]
42. Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 1996, 118, 2309–2309.
43. Savelyev A; MacKerell AD, All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J Comput Chem* 2014, 35, 1219–1239. [PubMed: 24752978]
44. 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. [PubMed: 26579919]
45. Mlynsky V; Banas P; Hollas D; Reblova K; Walter NG; 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. [PubMed: 20420375]
46. Yildirim I; Kennedy SD; Stern HA; Hart JM; Kierzek R; Turner DH, Revision of Amber Torsional Parameters for Rna Improves Free Energy Predictions for Tetramer Duplexes with Gc and Igic Base Pairs. *J Chem Theory Comput* 2012, 8, 172–181. [PubMed: 22249447]
47. Sponer J; Mladek A; Sponer JE; 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. [PubMed: 23072945]
48. Steinbrecher T; Latzer J; Case DA, Revised Amber Parameters for Bioorganic Phosphates. *J Chem Theory Comput* 2012, 8, 4405–4412. [PubMed: 23264757]
49. Bergonzo C; Cheatham TE, Improved Force Field Parameters Lead to a Better Description of Rna Structure. *J Chem Theory Comput* 2015, 11, 3969–3972. [PubMed: 26575892]
50. Banas P; Mladek A; Otyepka M; Zgarbova M; Jurecka P; Svozil D; Lankas F; Sponer J, Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J Chem Theory Comput* 2012, 8, 2448–2460. [PubMed: 26588974]
51. Schrodt MV; Andrews CT; Elcock AH, Large-Scale Analysis of 48 DNA and 48 Rna Tetranucleotides Studied by 1 Mus Explicit-Solvent Molecular Dynamics Simulations. *J Chem Theory Comput* 2015, 11, 5906–5917. [PubMed: 26580891]
52. Bergonzo C; Henriksen NM; Roe DR; Cheatham TE, 3rd, Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common Rna Force Fields. *Rna* 2015, 21, 1578–1590. [PubMed: 26124199]
53. Havrila M; Zgarbova M; Jurecka P; Banas P; Krepl M; Otyepka M; Sponer J, Microsecond-Scale Md Simulations of Hiv-1 Dis Kissing-Loop Complexes Predict Bulged-in Conformation of the Bulged Bases and Reveal Interesting Differences between Available Variants of the Amber Rna Force Fields. *J Phys Chem B* 2015, 119, 15176–15190. [PubMed: 26548477]
54. Besseova I; Otyepka M; Reblova K; Sponer J, Dependence of a-Rna Simulations on the Choice of the Force Field and Salt Strength. *Phys Chem Chem Phys* 2009, 11, 10701–10711. [PubMed: 20145814]
55. Kruse H; Mladek A; Gkionis K; Hansen A; Grimme S; Sponer J, Quantum Chemical Benchmark Study on 46 Rna Backbone Families Using a Dinucleotide Unit. *J Chem Theory Comput* 2015, 11, 4972–4991. [PubMed: 26574283]
56. Correll CC; Beneken J; Plantinga MJ; Lubbers M; Chan YL, 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. [PubMed: 14627814]

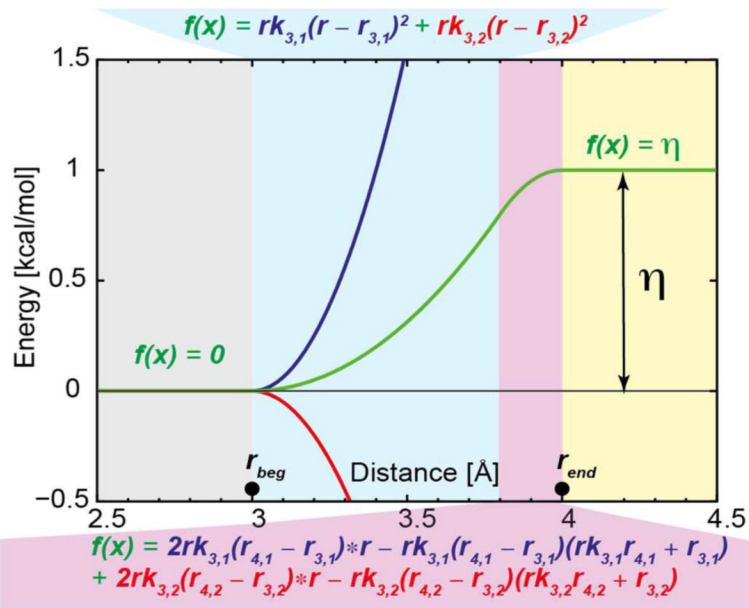


57. Zirbel CL; Sponer JE; Sponer J; Stombaugh J; Leontis NB, Classification and Energetics of the Base-Phosphate Interactions in Rna. *Nucleic Acids Res* 2009, 37, 4898–4918. [PubMed: 19528080]
58. Salomon-Ferrer R; Case DA; Walker RC, An Overview of the Amber Biomolecular Simulation Package. *Wires Comput Mol Sci* 2013, 3, 198–210.
59. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML, Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* 1983, 79, 926–935.
60. Joung IS; Cheatham TE, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J Phys Chem B* 2008, 112, 9020–9041. [PubMed: 18593145]
61. Chen AA; Pappu RV, Parameters of Monovalent Ions in the Amber-99 Forcefield: Assessment of Inaccuracies and Proposed Improvements. *J Phys Chem B* 2007, 111, 11884–11887. [PubMed: 17887792]
62. Aqvist J, Ion Water Interaction Potentials Derived from Free-Energy Perturbation Simulations. *J Phys Chem-US* 1990, 94, 8021–8024.
63. Leontis NB; Westhof E, Geometric Nomenclature and Classification of Rna Base Pairs. *Rna* 2001, 7, 499–512. [PubMed: 11345429]
64. 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. [PubMed: 26620784]
65. Bonomi M; Branduardi D; Bussi G; Camilloni C; Provasi D; Raiteri P; Donadio D; Marinelli F; Pietrucci F; Broglia RA; Parrinello M, Plumed: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput Phys Commun* 2009, 180, 1961–1972.
66. Sponer J; Jurecka P; Hobza P, Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. *J Am Chem Soc* 2004, 126, 10142–10151. [PubMed: 15303890]
67. Humphrey W; Dalke A; Schulten K, Vmd: Visual Molecular Dynamics. *J Mol Graph Model* 1996, 14, 33–38.
68. Hall P; Horowitz J, A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions. *Ann Stat* 2013, 41, 1892–1921.
69. Hall P; Horowitz JL; Jing BY, On Blocking Rules for the Bootstrap with Dependent Data. *Biometrika* 1995, 82, 561–574.
70. Henriksen NM; Roe DR; Cheatham TE, 3rd, Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J Phys Chem B* 2013, 117, 4014–4027. [PubMed: 23477537]
71. Menger M; Eckstein F; Porschke D, Dynamics of the Rna Hairpin Gnra Tetraloop. *Biochemistry-US* 2000, 39, 4500–4507.
72. Zhao L; Xia TB, Direct Revelation of Multiple Conformations in Rna by Femtosecond Dynamics. *J Am Chem Soc* 2007, 129, 4118–4119. [PubMed: 17373794]
73. Bottaro S; Di Palma F; Bussi G, The Role of Nucleobase Interactions in Rna Structure and Dynamics. *Nucleic Acids Res* 2014, 42, 13306–13314. [PubMed: 25355509]
74. Rodriguez A; Laio A, Clustering by Fast Search and Find of Density Peaks. *Science* 2014, 344, 1492–1496. [PubMed: 24970081]
75. Garcia AE; Paschek D, Simulation of the Pressure and Temperature Folding/Unfolding Equilibrium of a Small Rna Hairpin. *J Am Chem Soc* 2008, 130, 815–817. [PubMed: 18154332]
76. Heus HA; Pardi A, Structural Features That Give Rise to the Unusual Stability of Rna Hairpins Containing Gnra Loops. *Science* 1991, 253, 191–194. [PubMed: 1712983]
77. Orita M; Nishikawa F; Shimayama T; Taira K; Endo Y; Nishikawa S, High-Resolution Nmr-Study of a Synthetic Oligoribonucleotide with a Tetranucleotide Gaga Loop That Is a Substrate for the Cytotoxic Protein, Ricin. *Nucleic Acids Res* 1993, 21, 5670–5678. [PubMed: 8284214]
78. Olieric V; Rieder U; Lang K; Serganov A; Schulze-Briese C; Micura R; Dumas P; Ennifar E, A Fast Selenium Derivatization Strategy for Crystallization and Phasing of Rna Structures. *Rna* 2009, 15, 707–715. [PubMed: 19228585]

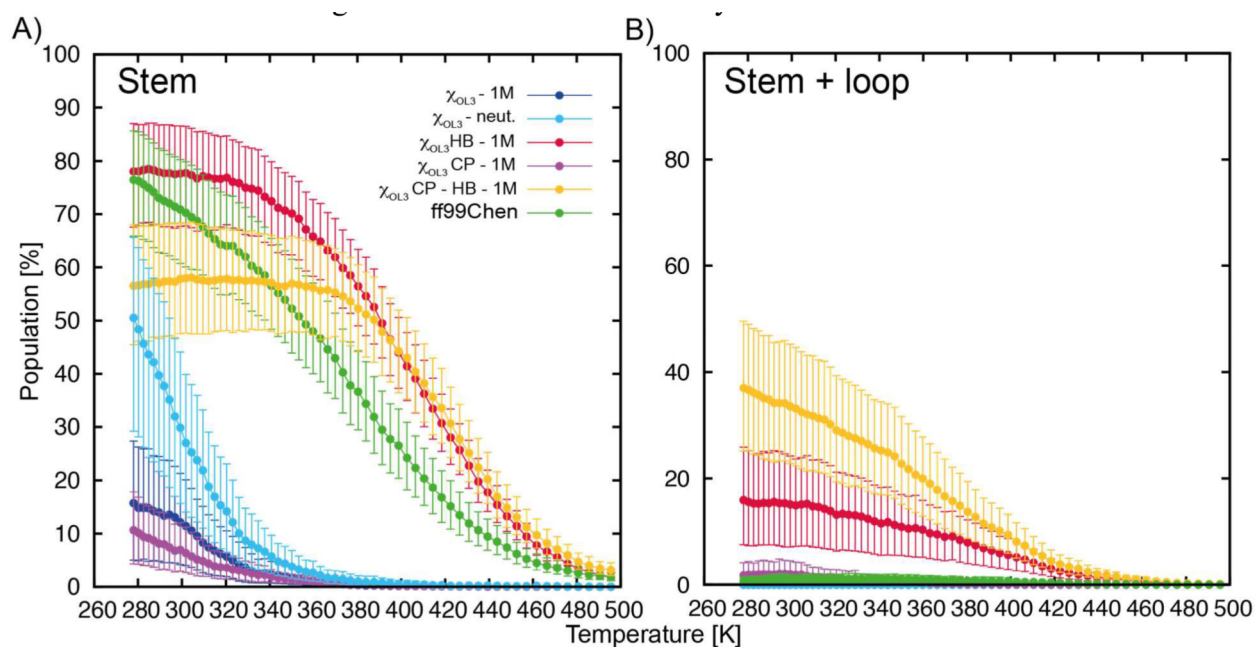
79. Sweeney BA; Roy P; Leontis NB, An Introduction to Recurrent Nucleotide Interactions in Rna. Wiley Interdiscip Rev: RNA 2015, 6, 17–45. [PubMed: 25664365]
80. Leontis NB; Lescoute A; Westhof E, The Building Blocks and Motifs of Rna Architecture. Curr Opin Struct Biol 2006, 16, 279–287. [PubMed: 16713707]
81. Sarver M; Zirbel CL; Stombaugh J; Mokdad A; Leontis NB, Fr3d: Finding Local and Composite Recurrent Structural Motifs in Rna 3d Structures. J Math Biol 2008, 56, 215–252. [PubMed: 17694311]
82. Sarkar K; Meister K; Sethi A; Gruebele M, Fast Folding of an Rna Tetraloop on a Rugged Energy Landscape Detected by a Stacking-Sensitive Probe. Biophys J 2009, 97, 1418–1427. [PubMed: 19720030]
83. Hairong M; Proctor DJ; Kierzek E; Kierzek R; Bevilacqua PC; Gruebele M, Exploring the Energy Landscape of a Small Rna Hairpin. J Am Chem Soc 2006, 128, 1523–1530. [PubMed: 16448122]
84. Mathews DH; Disney MD; Childs JL; Schroeder SJ; Zuker M; Turner DH, Incorporating Chemical Modification Constraints into a Dynamic Programming Algorithm for Prediction of Rna Secondary Structure. P Natl Acad Sci USA 2004, 101, 7287–7292.
85. Bizarro CV; Alemany A; Ritort F, Non-Specific Binding of Na<sup>+</sup> and Mg<sup>2+</sup> to Rna Determined by Force Spectroscopy Methods. Nucleic Acids Res 2012, 40, 6922–6935. [PubMed: 22492710]
86. Gresh N; Sponer JE; Spackova N; Leszczynski J; Sponer J, Theoretical Study of Binding of Hydrated Zn(II) and Mg(II) Cations to 5'-Guanosine Monophosphate. Toward Polarizable Molecular Mechanics for DNA and Rna. J Phys Chem B 2003, 107, 8669–8681.
87. Roh JH; Tyagi M; Aich P; Kim K; Briber RM; Woodson SA, Charge Screening in Rna: An Integral Route for Dynamical Enhancements. Soft Matter 2015, 11, 8741–8745. [PubMed: 26430908]
88. Ma HR; Proctor DJ; Kierzek E; Kierzek R; Bevilacqua PC; Gruebele M, Exploring the Energy Landscape of a Small Rna Hairpin. J Am Chem Soc 2006, 128, 1523–1530. [PubMed: 16448122]
89. Richardson JS; Schneider B; Murray LW; Kapral GJ; Immormino RM; Headd JJ; Richardson DC; Ham D; Hershkovits E; Williams LD; Keating KS; Pyle AM; Micallef D; Westbrook J; Berman HM, Rna Backbone: Consensus All-Angle Conformers and Modular String Nomenclature (an Rna Ontology Consortium Contribution). Rna 2008, 14, 465–481. [PubMed: 18192612]
90. Yildirim I; Stern HA; Tubbs JD; Kennedy SD; Turner DH, Benchmarking Amber Force Fields for Rna: Comparisons to Nmr Spectra for Single-Stranded R(GACC) Are Improved by Revised Chi Torsions. J Phys Chem B 2011, 115, 9261–9270. [PubMed: 21721539]
91. Condon DE; Kennedy SD; Mort BC; Kierzek R; Yildirim I; Turner DH, Stacking in Rna: Nmr of Four Tetramers Benchmark Molecular Dynamics. J Chem Theory Comput 2015, 11, 2729–2742. [PubMed: 26082675]
92. Sponer J; Zgarbova M; Jurecka P; Riley KE; Sponer JE; Hobza P, Reference Quantum Chemical Calculations on Rna Base Pairs Directly Involving the 2'-OH Group of Ribose. J Chem Theory Comput 2009, 5, 1166–1179. [PubMed: 26609626]
93. Brown RF; Andrews CT; Elcock AH, Stacking Free Energies of All DNA and Rna Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised Amber Parameters and Compared with Experiment. J Chem Theory Comput 2015, 11, 2315–2328. [PubMed: 26574427]
94. Aytenfisu AH; Liberman JA; Wedekind JE; Mathews DH, Molecular Mechanism for Preq1-Ii Riboswitch Function Revealed by Molecular Dynamics. Rna 2015, 21, 1898–907. [PubMed: 26370581]
95. Liberman JA; Suddala KC; Aytenfisu A; Chan D; Belashov IA; Salim M; Mathews DH; Spitale RC; Walter NG; Wedekind JE, Structural Analysis of a Class Iii Preq1 Riboswitch Reveals an Aptamer Distant from a Ribosome-Binding Site Regulated by Fast Dynamics. Proc Natl Acad Sci U S A 2015, 112, E3485–E3494. [PubMed: 26106162]



**Figure 1.** 3D (left) and secondary structures (right) of the studied GAGA hairpin, consisting of an A-RNA stem (grey labels and background) and tetraloop (red labels and background). Black dashed lines show the native hydrogen bonds and the signature interactions. The base-pairs and the base-phosphate interaction are annotated according to the standard nomenclature.<sup>57, 63</sup> We used two sequence lengths of the RNA hairpin: 8-mer TL (5'-gcGAGAgc-3') and 10-mer TL (5'-cgcGAGAgcg-3').

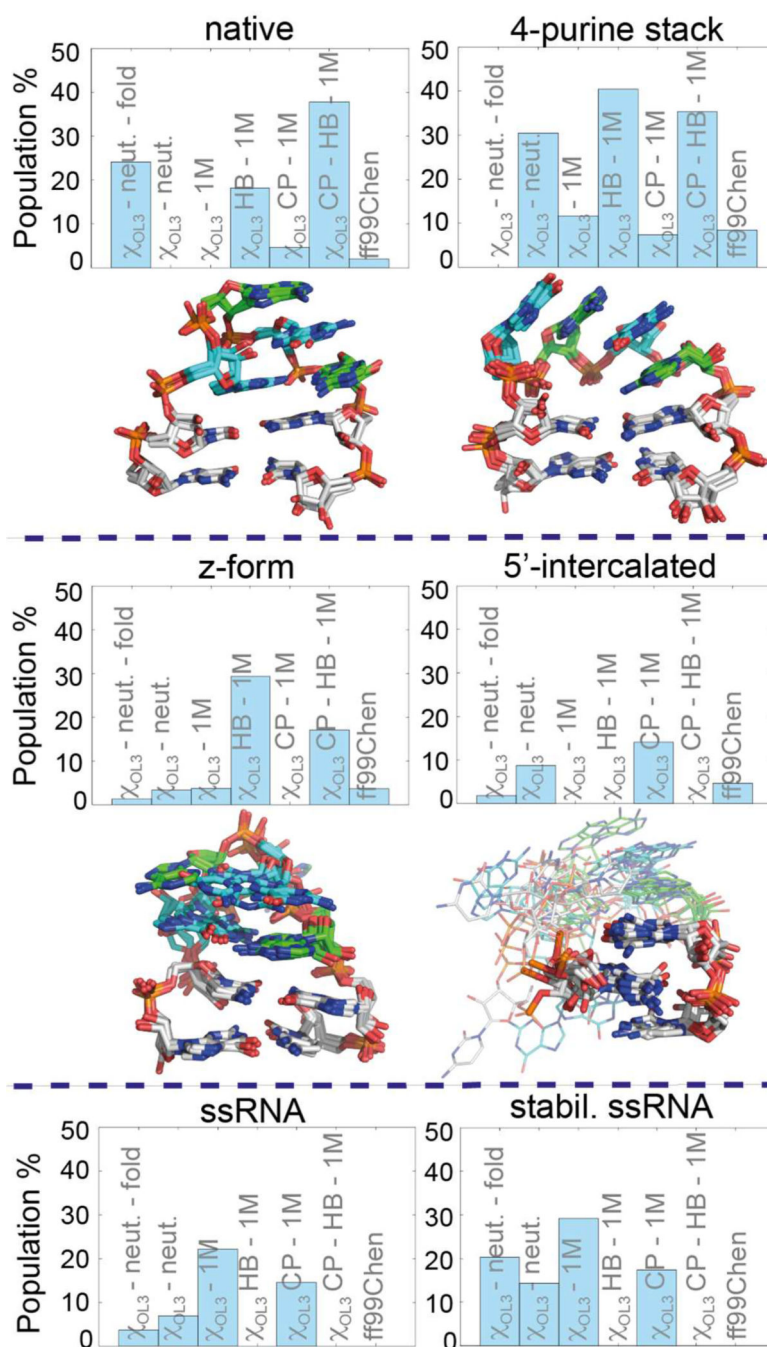
**Figure 2.**

Description of the HBfix potential (green curve) applied to heavy-atom distances in native hydrogen bonds to strengthen these interactions. The potential is constant (i.e., it provides zero forces) at all distances except the narrow region between 3 and 4 Å. It is composed of combination of two flat-well restraints with opposite sign of curvature and linear extensions that cancel each other at distances above 4 Å (red and blue curves). The parameters  $k_{3,1}$ ,  $k_{3,2}$  and  $r_{3,1}$ ,  $r_{4,1}$ ,  $r_{3,2}$ ,  $r_{4,2}$  values stand for spring constants of the quadratic parts of the flat-well restraints and distances enclosing the quadratic region, respectively (see more details in Supporting Information).



**Figure 3.**

The population of A) misfolded structures with only folded stem (including the terminal pair) and B) native structures as a function of temperature calculated over all T-REMD 8-mer simulations. The error bars were calculated by bootstrapping (using resampling of both time-blocks and coordinate following replicas, see Methods). Note that several simulations shown in the figure are discussed later in the text.



**Figure 4.**

The most populated GAGA 8-mer TL clusters observed in T-REMD simulations at 300 K based on clustering using the eRMSD metric. The bars present for each particular cluster its populations in the indicated 8-mer T-REMD simulation (note that several simulations are discussed later in the text, see Table 1 for their overview). All graphs, except the two bottom-most ones which correspond to ssRNA clusters, are accompanied by the overlapped structures representing cluster centers obtained in all T-REMD simulations. Solid bonds

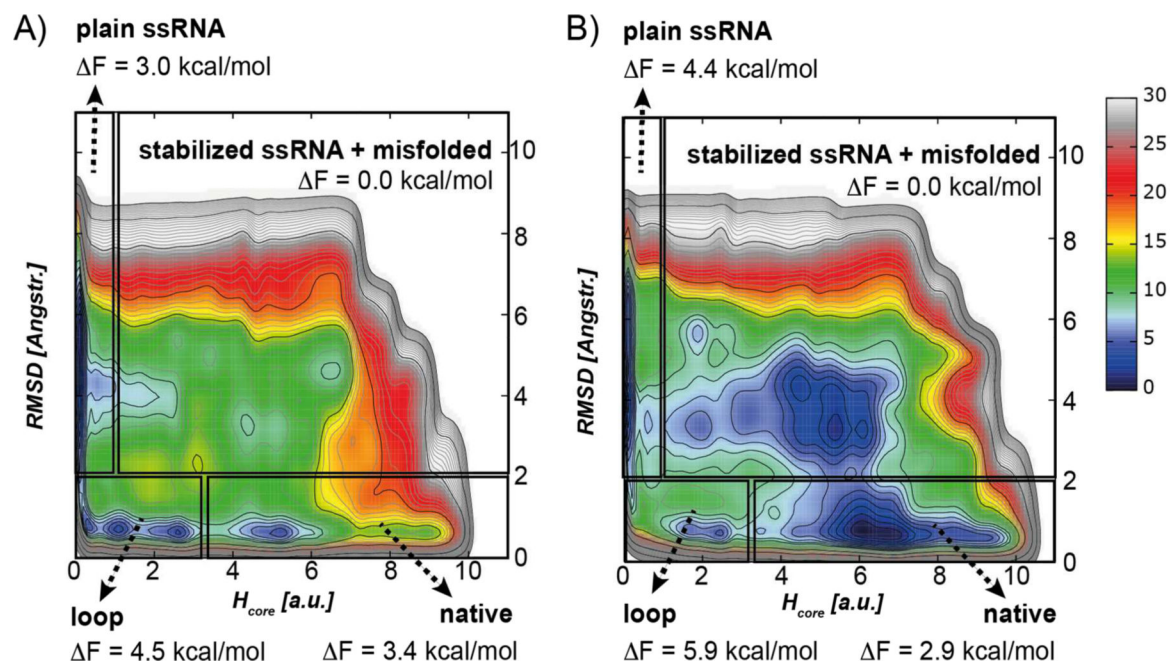
(stick) depict those parts of the structures that are shared in the given cluster while variable parts are shown faded (lines).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 5.**

Free energy surface plots of two independent 3- $\mu$ s-long WT-MetaD simulations of the GAGA TL using  $H_{core}$  and RMSD of the TL as CVs. The FES is contoured by 1 kcal/mol (thin lines) and 10 kcal/mol (thick lines). Energies lower than 20 kcal/mol relevant for structural dynamics are colored. The regions depicted on the FES by black boxes correspond to the conformational states populated during the simulations. The free energy values of conformational states were estimated by reweighting scheme.



**Table 1.**

List of all simulations of 8-mer GAGA TL with sequence 5'-gcGAGAgc-3' and 10-mer GAGA TL with sequence 5'-cgcGAGAgcg-3'.  $\chi_{OL3}$ -neut.-fold and  $\chi_{OL3}$ -neut.MetaD simulations started from the folded state, while all others were initiated from the unfolded state.

System	Method	Ionic conditions	Force field	Label	Length
8-mer	T-REMD	1M KCl	$\chi_{OL3}$	$\chi_{OL3}$ -1M	1 $\mu$ s x 64
			$\chi_{OL3}$ +HBfix 1 kcal	$\chi_{OL3}$ HB-1M	2 $\mu$ s x 64
			$\chi_{OL3}$ +HBfix 0.5 kcal	$\chi_{OL3}$ HB <sub>0.5</sub> -1M	1 $\mu$ s x 64
			ff99Chen	ff99Chen	2 $\mu$ s x 64
			$\chi_{OL3}$ -CP	$\chi_{OL3}$ CP-1M	2 $\mu$ s x 64
			$\chi_{OL3}$ -CP+HBfix 1 kcal	$\chi_{OL3}$ CP-HB-1M	2 $\mu$ s x 64
	K <sup>+</sup> net-neutral	$\chi_{OL3}$	$\chi_{OL3}$ -neut.	1 $\mu$ s x 64	
		$\chi_{OL3}$	$\chi_{OL3}$ -neut.-fold	2 $\mu$ s x 64	
	REST2	1M KCl	$\chi_{OL3}$	$\chi_{OL3}$ -1M-REST	1 $\mu$ s x 64
10-mer	T-REMD	K <sup>+</sup> net-neutral	$\chi_{OL3}$	$\chi_{OL3}$ -neut.-10mer	1 $\mu$ s x 64
	WT-MetaD	Na <sup>+</sup> net-neutral	$\chi_{OL3}$	$\chi_{OL3}$ -neut.MetaD	3 $\mu$ s x 2

**Table 2.**

Thermodynamic parameters of both studied TLs (8-mer and 10-mer) estimated by Turner et al. 2004 parameters<sup>84</sup> and the corresponding estimate of  $\Delta G$  of folding and expected population of the native state at 300 K.

	<b>8-mer</b>	<b>10-mer</b>
$G_{310K}$ [kcal/mol]	-0.1±0.2	-2.5±0.2
H [kcal/mol]	-24.1±2.4	-34.7±2.9
S [cal/mol/K]	-77.3±7.8	-104.0±9.5
$T_m$ [K]	311.8	334.1
$G_{300K}$ [kcal/mol]	-0.9±0.2	-3.5±0.2
$P_{\text{native},300K}$ [%]	82.0±4.4	99.7±0.1