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

pubs.acs.org/JPCB

Article

A Computational Study of RNA Tetraloop Thermodynamics, Including Misfolded States

Published as part of The Journal of Physical Chemistry virtual special issue "Dor Ben-Amotz Festschrift".

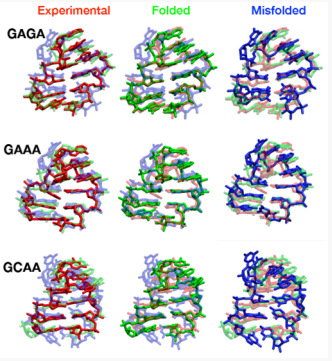
Gül H. Zerze, Pablo M. Piaggi, and Pablo G. Debenedetti*

Cite This: *J. Phys. Chem. B* 2021, 125, 13685–13695

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: An important characteristic of RNA folding is the adoption of alternative configurations of similar stability, often referred to as misfolded configurations. These configurations are considered to compete with correctly folded configurations, although their rigorous thermodynamic and structural characterization remains elusive. Tetraloop motifs found in large ribozymes are ideal systems for an atomistically detailed computational quantification of folding free energy landscapes and the structural characterization of their constituent free energy basins, including nonnative states. In this work, we studied a group of closely related 10-mer tetraloops using a combined parallel tempering and metadynamics technique that allows a reliable sampling of the free energy landscapes, requiring only knowledge that the stem folds into a canonical A-RNA configuration. We isolated and analyzed unfolded, folded, and misfolded populations that correspond to different free energy basins. We identified a distinct misfolded state that has a stability very close to that of the correctly folded state. This misfolded state contains a predominant population that shares the same structural features across all tetraloops studied here and lacks the noncanonical A-G base pair in its loop portion. Further analysis performed with biased trajectories showed that although this competitive misfolded state is not an essential intermediate, it is visited in most of the transitions from unfolded to correctly folded states. Moreover, the tetraloops can transition from this misfolded state to the correctly folded state without requiring extensive unfolding.



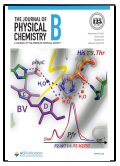
INTRODUCTION

A distinguishing feature of RNA is its dual ability to carry genetic information and to be catalytically active,[1−3] a characteristic often attributed to RNA's ability to fold into various structures.[4] A well-known challenge in *in vitro* RNA folding (or refolding) is the formation of nonnative low-free-energy conformers during the process.[5−7] These inactive or partially active conformers have been argued to have a free energy similar to that of the native configuration.[5,7] Moreover, they have been described as being separated from the native state by low free energy barriers, giving rise to rugged RNA folding landscapes.[5,7] The folding time scales of large ribozymes are often referred to as being glacial, compared to protein folding time scales, and this sluggishness has been argued to be the result of the presence of numerous kinetic traps in the folding landscape.[7] A range of terms are used interchangeably in the literature to describe these nonnative conformers in the context of the RNA folding problem,[6] such as alternative configurations, kinetic traps,[7] competing structures,[8] misfolded intermediates,[9] etc.

Folding free energy surfaces are informative representations to study these nonnative and native conformers. Computer simulations of RNA at atomistic resolution not only are an invaluable tool to obtain free energy surfaces (via advanced sampling)[10,11] but also provide detailed structures of the configurations that make up the free energy basins and transition states.[12,13] Much of the recent computational work on RNA folding at atomistic resolution has focused on identifying biases in RNA force fields and refining the models accordingly.[14−22] Some of this work has reported artificially overstabilized nonnative compact and/or partially folded structures.[15,18] The relative stabilities of native and nonnative configurations have then served as an implicit target and as a validation test for force field improvements.[20,23] While these efforts have led to significant success in improving the RNA force fields, reports of nonnative configurations have been mostly in the context of force field refinement; a thorough investigation of nonnative states is still needed.

Advanced sampling methodologies are a key ingredient in atomistic simulations of RNA folding, especially where the relative stabilities of native and nonnative configurations are the focus of study. Obtaining a reliable measure of relative

Received: September 11, 2021
Revised: November 28, 2021
Published: December 10, 2021



https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

stabilities is a formidable challenge in atomistically detailed calculations.[15,18,24] Collective variable-based advanced sampling, in particular metadynamics with strategically crafted order parameters, has made this goal more feasible.[22,25] However, recent metadynamics-based techniques for sampling RNA free energy surfaces have been dependent on collective variables that require *a priori* knowledge of the native structure.[25] In this work, we first describe a combined parallel tempering and metadynamics sampling technique that reliably samples the folding free energy surfaces of RNA tetraloops at the atomistic scale, requiring only knowledge, derived from experiments, that the stem folds into a canonical A-RNA configuration. We then applied our technique to study the folding free energy surface of a family of tetraloops, directing our focus to competitive nonnative states. We identified a competing misfolded state that is shared by all of the tetraloops that we studied. A structural clustering analysis of this misfolded state revealed that it heavily populates configurations whose number of stacked bases is the same as in the correctly folded state, which is consistent with the competitive nature of this misfolded state, while consistently lacking a noncanonical AG base pairing in the loop.

Kinetics is an integral component in the description of nonnative states.[26−29] One important question related to nonnative states in RNA folding is whether such states are on-pathway (folding intermediates) or off-pathway (misfolded end states) states.[30−33] DePaul et al. built a Markov model for a tetraloop RNA by generating several nanoseconds of molecular dynamics (MD) data for 10 000 initial conditions (obtained from a 1 ns annealing simulation) and found several different folding pathways, as well as on- and off-pathway intermediates.[28] Huang et al. constructed a multiresolution Markov model for the same tetraloop with a similar simulation technique and found no long-lived (metastable) on- or off-pathway intermediates.[34] They instead found that the folded state serves as a center that connects multiple unfolded/misfolded states separated by large barriers. Given the lack of a clear consensus on the kinetic behavior relating misfolded states to folded and/or misfolded states, we also performed a kinetic analysis by taking advantage of long albeit biased trajectories (recovered from combined parallel-tempering and metadynamics simulations) where a number of transitions between folded, unfolded, and misfolded states were observed. We found that our identified misfolded state is not an irreversible end state. Instead, it acts more as a reaction intermediate. Although the folding process can take place without visiting this misfolded state (i.e., it is not an essential intermediate), folding transitions in which this state is an intermediate outnumber those in which it is not.

We mention in closing that machine-learning-based approaches hold considerable promise in the elucidation of RNA structures. For a recent example, see Townshend et al.[35]

METHODS

RNA Tetraloops. Small RNA motifs show many characteristics of large ribozymes, including slow folding.[7,8] RNA folds contain several such motifs, among which hairpin loops (or stem-loops) are the most abundant.[36,37] Hairpin loops contribute to the structural stability of RNA folds by capping double-stranded helices, serving as nucleation sites for RNA secondary structures,[38] and also involving in tertiary structure formation.[39,40] RNA tetraloops are autonomously foldable hairpin loop units (i.e., they are able to fold independently of the context) that contain four-nucleotide loop structures. A typical RNA tetraloop unit has a stem-loop architecture where the stem forms a double-stranded helix with Watson−Crick base pairs in A-configuration (A-RNA) and a four-nucleotide loop that changes the backbone direction. Certain sequence patterns in the loop part are found to be conserved among tetraloops.[41,42] The GNRA pattern (where N is any of the four nucleotides, and R is either adenine or guanine) is one of the most frequently occurring sequence pattern in rRNA hairpins.[41] A degree of structural diversity (or conformational heterogeneity) has been experimentally found for GNRA hairpins[43−48] and also observed in computer simulations.[14−22,28,34] Being an autonomous folding unit, combined with their relatively small size, tetraloops are ideal candidates for computational studies at atomistic resolution that can shed light on conformational heterogeneity leading to misfolded states.

Modeling. The full sequences of single-stranded GNRA tetraloops studied here are GGCGAGAGCC, GGCGAAAGCC, and GGCGCAAGCC which will be referred to hereafter as GAGA, GAAA, and GCAA, respectively. Unfolded initial coordinates of the tetraloops were generated using the Nucleic Acid Builder tool of AmberTools.[49] Each sequence was modeled with the nucleic acid force field developed recently by Shaw and co-workers[17] combined with the TIP4PD water model[50] using the library files provided by Kuhrova et al.,[20] which will be referred to as DESRES ff hereafter. A single copy of each tetraloop was solvated in a truncated octahedron box volume of 135 nm^3. Na^+ ions were added to provide electroneutrality, and the salt concentration was adjusted to 1 M (NaCl). Ions were modeled by CHARMM22 parameters.[51] Simulation input files were generated using AmberTools[49] and then converted to a format compatible with GROMACS.

Refining atomistic force fields for more accurate modeling of RNA is an active area of research.[15−18,20,21,25,52] We chose the DESRES ff as it showed promising improvements in the representation of a variety of different RNA types (including tetraloops) validated against experimental data.[17] We also did our own force field testing where we also considered the most recent Amber RNA force field variant, the modified Amber ff99bsc0χ_{OL3}.[53−56] We observed a strong initial condition dependence and lack of convergence after simulating the systems for long periods of time with this force field (Figures S11 and S12). Similar to previous reports with this force field, we observed that this hysteretic behavior is the result of the RNA being stuck in arbitrary configurations for long periods of time (Figure S12A), likely due to unbalanced interactions in the force field.[15,18] We provide a further discussion of the results with the Amber ff99bsc0χ_{OL3} force field in the Supporting Information. We did not observe these problems with DESRES ff (Figures S1 and S2).

Sampling Methods. Parallel-tempering in the well-tempered ensemble combined with well-tempered metadynamics (PTWTE-WTM)[57−62] was used for sampling the tetraloop folding. In the framework of the well-tempered ensemble,[60] the potential energy was used as a collective variable in well-tempered metadynamics using Gaussian kernels with a 600 kJ/mol width and a 1.5 kJ/mol initial height. The deposition stride for the Gaussians was set to 2000 steps, and a bias factor of 25 was employed. The temperatures of the 14 replicas, ranging between 300 and 475 K, were

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

distributed geometrically. The average replica exchange acceptance ratio was 35%.

In the WTM part of PTWTE-WTM, we biased a similarity-based order parameter (Q) that only requires knowledge that the stem folds into a canonical A-RNA configuration whose crystallographic information is available in the literature.[63] We defined Q only for the stem part of tetraloops following a protocol analogous to the one that we defined for B-DNA recently.[64] The order parameter Q is defined as

$$Q = \frac{1}{N_{\mathrm{nb}_{\mathrm{stem}}}} \sum_{(i,j)} \frac{1}{1 + \exp(\gamma(r_{ij} - \lambda r_{ij}^{0}))} \quad (1)$$

by adapting the generalized definition of Q,[65,66] for the interstrand contacts. The sum runs over $N_{\mathrm{nb}_{\mathrm{stem}}}$, which is total number of (stem) atomic pairs (i, j) that are considered in contact where the atom i is from the strand ending at 3′ and atom j is from the strand ending at 5′. Any heavy (i.e., non-hydrogen) nucleobase (nb) atom of the 3′ strand was considered in contact with a heavy nb atom of the 5′ strand if the distance between them is less than 5 Å in the reference native structure. r_{ij}^{0} and r_{ij} are the distances between i and j in the reference native structure and in any given instantaneous configuration, respectively. γ in the smoothing function was taken as 50 nm^{-1}, and the adjustable parameter λ was taken as 1.5.[66] The reference native structures (double-stranded A-RNA configurations) for stem portions were created by the Nucleic Acid Builder tool of AmberTools[49] which uses X-ray fiber diffraction data.[63]

The initial Gaussian height for Q was set to 1.8 kJ/mol with a bias factor of 25 for the WTM sampling. The Gaussian width was set to 0.01. Since Q is defined strictly between 0 and 1, interval limits were applied together with restraining potentials to avoid accumulating systematic errors at the boundaries of Q (at Q = 0.01 and 0.99 for lower and upper boundaries, respectively) .[67] The force contribution from metadynamics bias acting on Q was set to zero beyond these limits, and harmonic restraining potentials with a spring constant of 75 000 kJ/mol were applied at the defined boundaries.[68]

After initial solvation, all tetraloop systems were equilibrated with 100 ps *NVT* simulations (T = 300 K) followed by 100 ps *NPT* simulations (T = 300 K, P = 1 bar). Prior to starting the PTWTE-WTM simulations, unbiased *NVT* simulations of each replica were performed for 200 ps in order to equilibrate the potential energy of the replicas. Production simulations were run at *NPT* conditions, where the temperature was maintained constant at any given replica temperature using a Nosé–Hoover thermostat[69,70] with a 1 ps time constant. Atmospheric pressure (1 bar) was maintained using an isotropic Parrinello–Rahman barostat[71,72] with a time constant of 2 ps. Electrostatic interactions were calculated using the particle-mesh Ewald method[73] with a real space cutoff distance of 1 nm. A cutoff distance of 1 nm was also used for the van der Waals interactions.

All simulations were performed using GROMACS (version 2016.3)[74,75] patched with the PLUMED (version 2.3.1) enhanced sampling plugin[76,77] for metadynamics sampling.

Convergence. As a stringent test of convergence, we performed two independent PTWTE-WTE simulations for one of the tetraloop hairpins (GAGA). In the first one, we initiated all replicas from fully unfolded configurations (initial condition 1, IC1) whereas in the second one, we initiated all the replicas from correctly folded configurations (IC2). The correctly folded configurations are the first of the 10 model structures deposited in 1ZIG PDB entry. For each IC, the sampling of Q as a function of time/replica at 300 K is presented in Figure S1. Initial 500 ns/replica simulation time is discarded as equilibration, and the remainder of trajectories is used for analysis. The results from IC1 are reported in the main text, and IC2 is used to report the deviation between two ICs. Free energy surfaces as a function of Q and RMSD are presented in Figure S2, together with the difference between them (as the error). As can be seen in Figure S2, the difference between two simulations with independent ICs (after 1 μs/replica sampling in both) is not greater than $\pm kT$ in any region of low free energy (FE < 20 kJ/mol), which lends confidence on the convergence of our simulations. All tetraloops in this work were simulated at least for 1 μs per replica.

An important point is that we did not observe the sampling of low-RMSD configurations using plain parallel-tempering (that is, without metadynamics bias on Q) (Figure S3) within a reasonable sampling time per replica for any tetraloop hairpins we studied. This highlights the requirement for an advanced sampling technique beyond parallel tempering.

Analysis Methods. We quantified the free energy surfaces (FES) of tetraloops using two order parameters, one of which is the parameter biased during the course of the simulations (Q), and the other one is the heavy atom root-mean-square distance (RMSD) from the reference native structure. The first of the 10 solution NMR structures deposited as the Protein Data Bank (PDB) entries 1ZIG, 1ZIF, and 1ZIH[45] were used as the reference folded structures for GAGA, GAAA, and GCAA tetraloops, respectively.

The unbiased two-dimensional probability densities of order parameters Q and RMSD were obtained by reweighting all the biases deposited on the Q and potential energy coordinates by using the technique described by Tiwary and Parrinello[78] as implemented in PLUMED (version 2.4).[76] The unbiased probability density, $P(Q, \mathrm{RMSD})$, was then converted to free energy via the equation

$$F(Q, \mathrm{RMSD}) = -kT \ln P(Q, \mathrm{RMSD}) + C \quad (2)$$

where kT is the product of the Boltzmann constant and the temperature, and C is an immaterial constant. RMSD has been particularly useful in determining the structural differences that arise from the differences in the loop (Q only includes stem atoms, and RMSD includes both stem and loop atoms).

The number of stacked bases was calculated based on a pairwise coordination of centers of masses of adjacent nucleobases, where a 0.7 nm distance was used as a cutoff for stacking. The maximum number of stacked bases is 9 for strands of 10 nucleotides.

Clustering was performed based on structural similarity of the backbone heavy, i.e., non-hydrogen atoms, following the GROMOS algorithm,[79] using a 0.20 nm root-mean-square deviation (RMSD) cutoff distance. We note that the calculated intracluster RMSDs showed that they were below around 0.3 nm[80] as shown in Figure S14 for the folded cluster and the most populated misfolded cluster of the GAGA tetraloop.

Reactive trajectory analysis was done by calculating Q and RMSD of continuous trajectories, i.e., trajectories that jump in temperature but are continuous in time. Continuous trajectories were obtained by demultiplexing parallel-tempered trajectories using the demux code of the GROMACS package.[74]

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

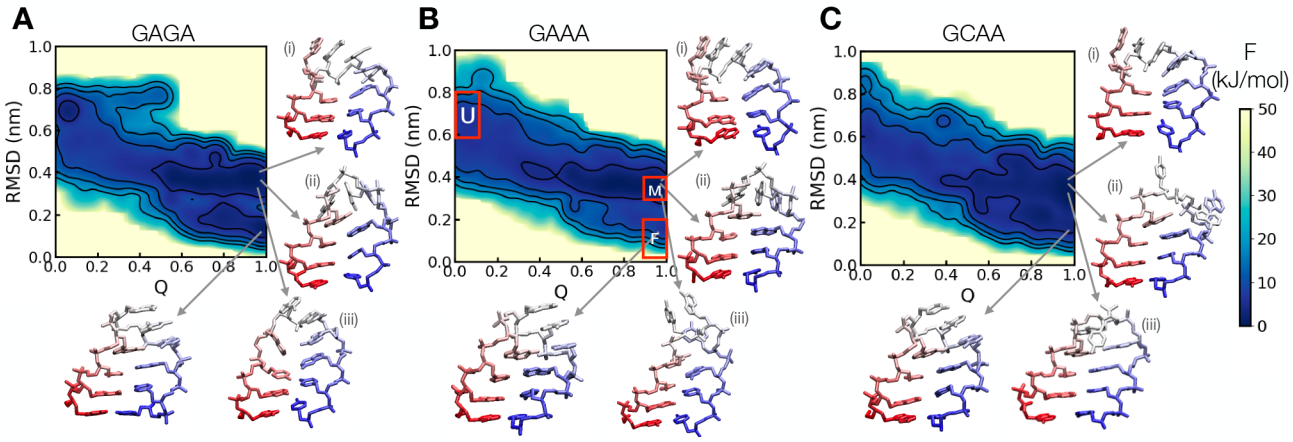


Figure 1. FES of tetraloops are calculated as a function of Q and RMSD at 300 K. Structural clustering was performed for each tetraloop at 300 K, specifically for subpopulations of M and F whose boundaries are shown in the middle panel. The basin F was represented by one predominant configuration in structural clustering for each tetraloop. A representative configuration from basin F is indicated by an arrow emanating from basin F for each tetraloop. We found some structural heterogeneity within the basin M for each tetraloop, where the most populated cluster i shares common structural features across all three tetraloops. Representative configurations of the three most populated structural clusters are shown in each panel. (A) FES of GAGA tetraloop. Percentages of clusters i, ii, and iii are 72, 10, and 4, respectively. (B) FES of GAAA tetraloop. Percentages of clusters i, ii, and iii are 74, 16, and 2, respectively. (C) FES of GCAA tetraloop. Percentages of clusters i, ii, and iii are 17, 13, and 12, respectively. The largest structural heterogeneity in state M is found for the GCAA tetraloop.

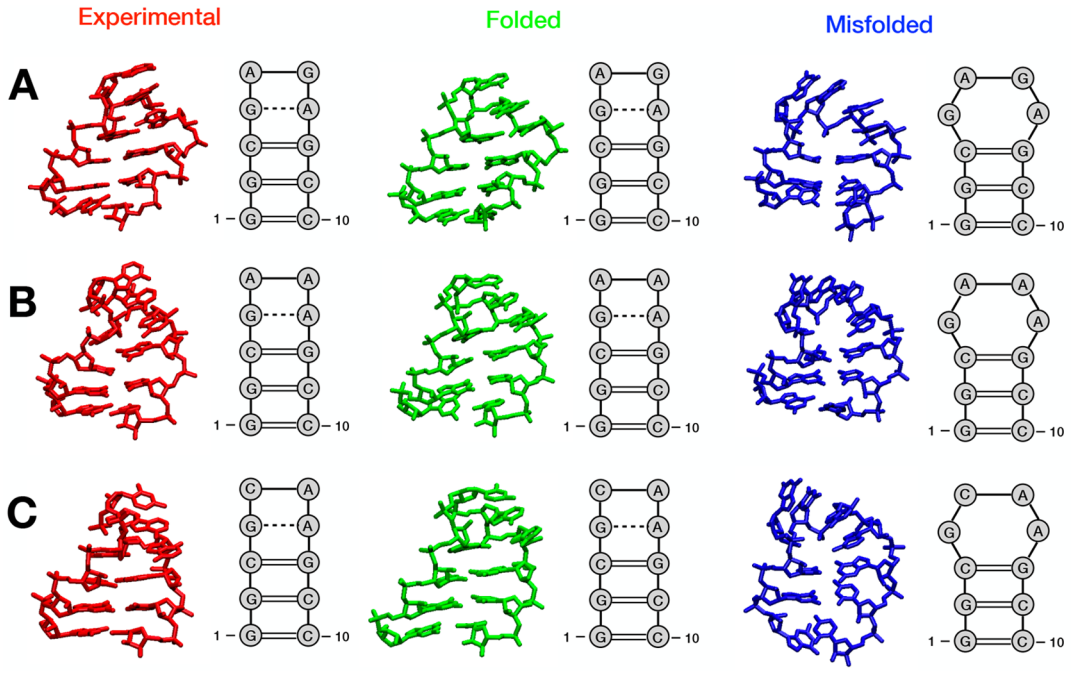


Figure 2. Experimentally found structures (red) and representative structures of the basin F (green) and the most populated basin M (blue) are compared side by side for GAGA (A), GAAA (B), and GCAA (C) tetraloops. Experimentally found (red) structure belongs to the first of the 10 model structures deposited as the PDB entries 1ZIG, 1ZIF, and 1ZIH[45] for GAGA, GAAA, and GCAA, respectively. For each configuration, a secondary structure annotation is provided on its right, where double lines denote canonical Watson−Crick base pairing and dashed lines denote noncomplementary (wobble) base pairs.

RESULTS AND DISCUSSION

Free Energy Surfaces of GAGA, GAAA, and GCAA Tetraloops. The FES of GNRA tetraloops projected onto the order parameters Q and RMSD revealed major similarities between the three studied sequences (Figure 1A−C) at 300 K. We identified three distinct low-free energy regions, shared by all three tetraloops, one of which is an unfolded high-RMSD, low-Q state, whereas the other two are high-Q states. One of these high-Q states is a low-RMSD (≈0.2 nm) state, whereas the other has a comparatively higher RMSD (≈0.4 nm). For further analysis, we isolated the subpopulations belonging to these three free energy minima at 300 K. We first performed a structural clustering of these subpopulations, which showed that the high-Q, low-RMSD (≈0.2 nm) state represents the correctly folded state for all tetraloops, as it not only has the correctly folded stem but also has the noncanonical G-A pair formed between the first and last nucleotides in the GNRA loop, as in experimentally reported folded structures.[43,45] The

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

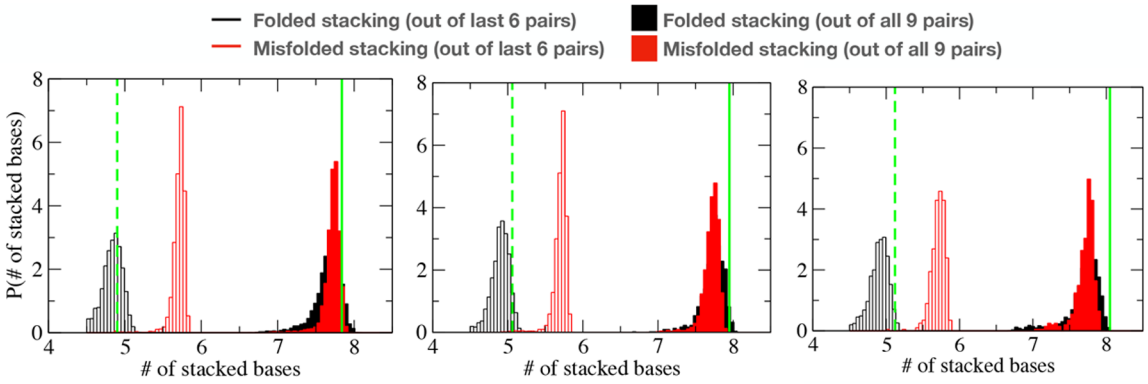


Figure 3. Normalized probability distributions of the number of stacked base pairs in correctly folded (black) and in misfolded (red) configurations for GAGA (left), GAAA (middle), and GCAA (right) tetraloops. Stacked pairs are separately counted for all pairs (filled bars) and for the last six pairs (unfilled bars). The misfolded population has the last six pairs stacked, whereas only five out of six pairs are stacked in the correctly folded population. The total number of stacked pairs is the same for both populations. Stacked pairs in their experimentally found native configuration are shown with green lines (solid line is for all pairs, broken line is for last six pairs). Calculations for experimentally found structure are performed for the first of the 10 model structures deposited as the PDB entries 1ZIG, 1ZIF, and 1ZIH[45] for GAGA, GAAA, and GCAA, respectively.

high-Q, high-RMSD (≈0.4 nm) state has its stem folded correctly (canonical A-RNA) as in the native state. However, there are major differences in the loop part, as can be seen in the structural clusters (Figure 1). A detailed structural analysis of folded and misfolded states is provided in the next subsection. We labeled the three states as U (unfolded, Q < 0.11, 0.6 < RMSD < 0.8), F (correctly folded, Q > 0.9, RMSD < 0.2), and M (misfolded, Q > 0.9, 0.3 < RMSD < 0.4), as marked on Figure 1B. Clustering of the subpopulation constituting the state F yielded a single cluster (see Methods for details of the clustering analysis), whereas we found multiple clusters for the state M (representative configurations from the three most populated clusters in the basin M are shown for each tetraloop in Figure 1). However, within this structurally heterogeneous state (M), we found one predominant configuration in all tetraloops that shares common structural features (cluster i in Figure 1 for each tetraloop). In order to aid the visualization of the structural differences between them, we present the experimentally found native structure, the folded structure obtained in our simulations (F), and the predominant misfolded structure found in the basin M in Figure 2, together with a secondary structure annotation. The annotation facilitates visualizing the lack of the wobble GA pair in misfolded configurations for all of the tetraloops (Figure 2). Further structural characterization of the predominant misfolded cluster in comparison with the correctly folded structure is presented in the next subsection.

We also calculated the stability of the F, M, and U states. Instead of calculating a melting curve based on a two-state folding assumption, we report free energy differences ΔF_{FU}, ΔF_{MU}, and ΔF_{MF} (where $\Delta F_{ij} = F_i - F_j$) since our transition states data suggest a more complicated mechanism than a two-state folding (see the Analysis of Reactive Trajectories subsection). We observe a monotonic nonlinear increase in both ΔF_{FU} and ΔF_{MU} as temperature increases (Figure S4A,B) indicating that both the M and F basins are destabilized with respect to the U state as temperature increases. Within the accuracy of the force field, we found that ΔF_{MF} is around 0 at room temperature. For GAGA, ΔF_{MF} is slightly positive, whereas for GAAA and GCAA it is slightly negative at 300 K, with the difference being within thermal fluctuations (i.e., $\pm kT$ ≈ 2.5 kJ/mol at 300 K) (Figure S4C). We note that the F state contains only one cluster composed of structures similar to the NMR structure.[45] However, a large degree of conformational diversity in the loop part of RNA hairpins has been detected in fluorescence experiments.[46,48] From these data, structural models have also been proposed where hairpins with alternative loop structures coexist with similar populations (same stem, different loop). Despite being in coexistence with the NMR structure, these alternative structures have been argued to be undetectable by conventional NMR since their high interconversion rates might be beyond the fast exchange limit.[46] Therefore, our findings with alternative loop structures with similar stability are in agreement with the fluorescence experiments.[46,48]

Structural Characterization of the High-Q States. One of the most significant differences between the correctly folded state (F) and the predominant cluster that is shared by all tetraloops in state M is the lack of noncomplementary G-A base pair (between the G and A of GNRA loop) in the misfolded (M) structure (Figure 2). Another notable difference is the arrangement of base stacking. There are a total of 10 nucleotides (n) in each tetraloop, i.e. the maximum number of stacked base pairs would be 9 ($=n-1$). In their correctly folded configurations, we found that the tetraloops have their first four nucleobases (including all three in 5′ strand and the first in the loop) stacked as a group and the remaining six are stacked as another group; i.e., the total number stacked base pairs is 8, which is consistent with the stacking observed in the experimentally found structures[45] (see also Figure 3, green data, for the number of stacked pairs in the experimentally found structures). The total number of stacked base pairs remains the same in the predominant M configuration; however, independently stacked groups are changed. In the misfolded configuration, the first three nucleobases (i.e., the three in the 5′ strand) are stacked as one group and the last seven nucleobases are stacked in the other group (all four in the loop and all three in the 3′ strand). This misfolded structure shares similar structural features with the "4-stack" configuration (all four loop bases are stacked on the 3′ side), as found in the work by Bottaro and Lindorff-Larsen by analyzing the tetraloop motifs found in the Protein Data Bank.[81] We quantify and present these stacking differences in Figure 3.

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

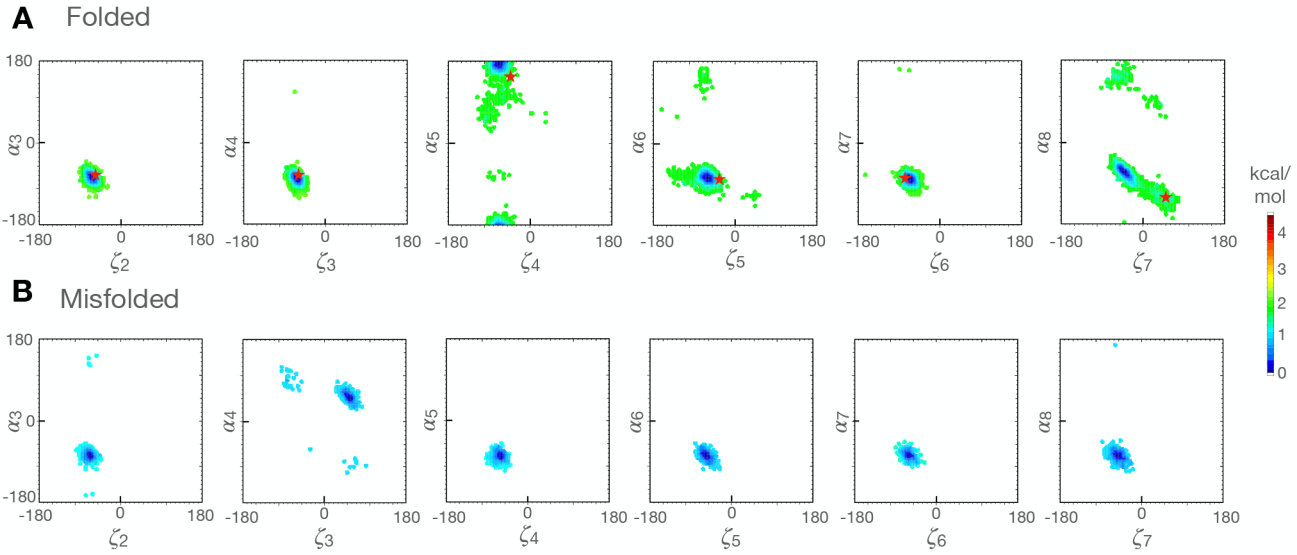


Figure 4. Free energies projected on α and ζ dihedrals of the GAGA tetraloop for the entire subpopulation in basin F (A) and for the most populated cluster in basin M (B). Dihedral angles in the native state, which are marked with a red star on each of the top panels, belong to the first of the 10 model structures deposited as the PDB entry 1ZIG.[45]

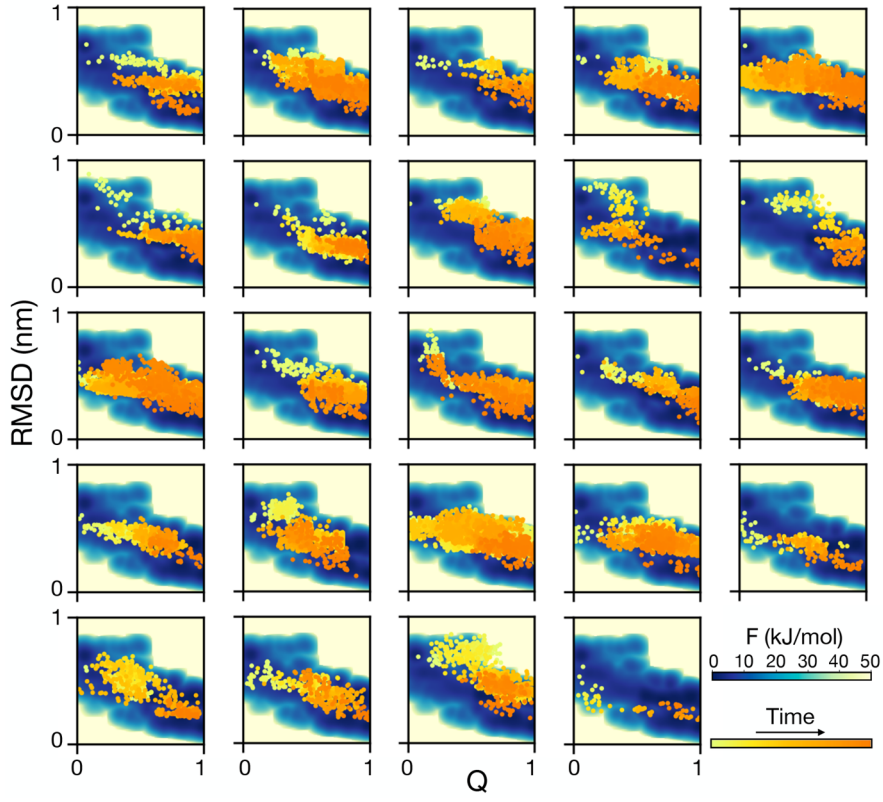


Figure 5. Paths of each reactive trajectory (starting from U and landing in F) found for GAGA tetraloop are illustrated on its two-dimensional (Q vs RMSD) FES (evaluated at 300 K). Only 4 trajectories achieved such a transition without visiting the state M (second row and fourth column, fourth row and second column, fourth row and fifth column, fifth row and fourth column) out of 24 total transitions.

To be able to accommodate the same number of stacked pairs in a different arrangement, this misfolded configuration adopts a twist in the backbone (between nucleotides 3 and 4, where the nucleotides are numbered starting from the 5′ end), different from the correctly folded configuration. We quantified this twist by measuring the backbone conformation around the phosphodiester bond, as this conformation is the most relevant to the backbone compaction and twists.[16,25] We measured backbone torsion angles α (for nucleotide i) and ζ (for nucleotide $i-1$) as shown in Figure 4 for the GAGA tetraloop (see Figures S5 and S6 for GAAA and GCAA tetraloops, respectively), both for the misfolded and folded populations. In the correctly folded state, the first four nucleobases are stacked in the canonical A-RNA form, where the $\alpha(i)/\zeta(i-1)$ angles

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

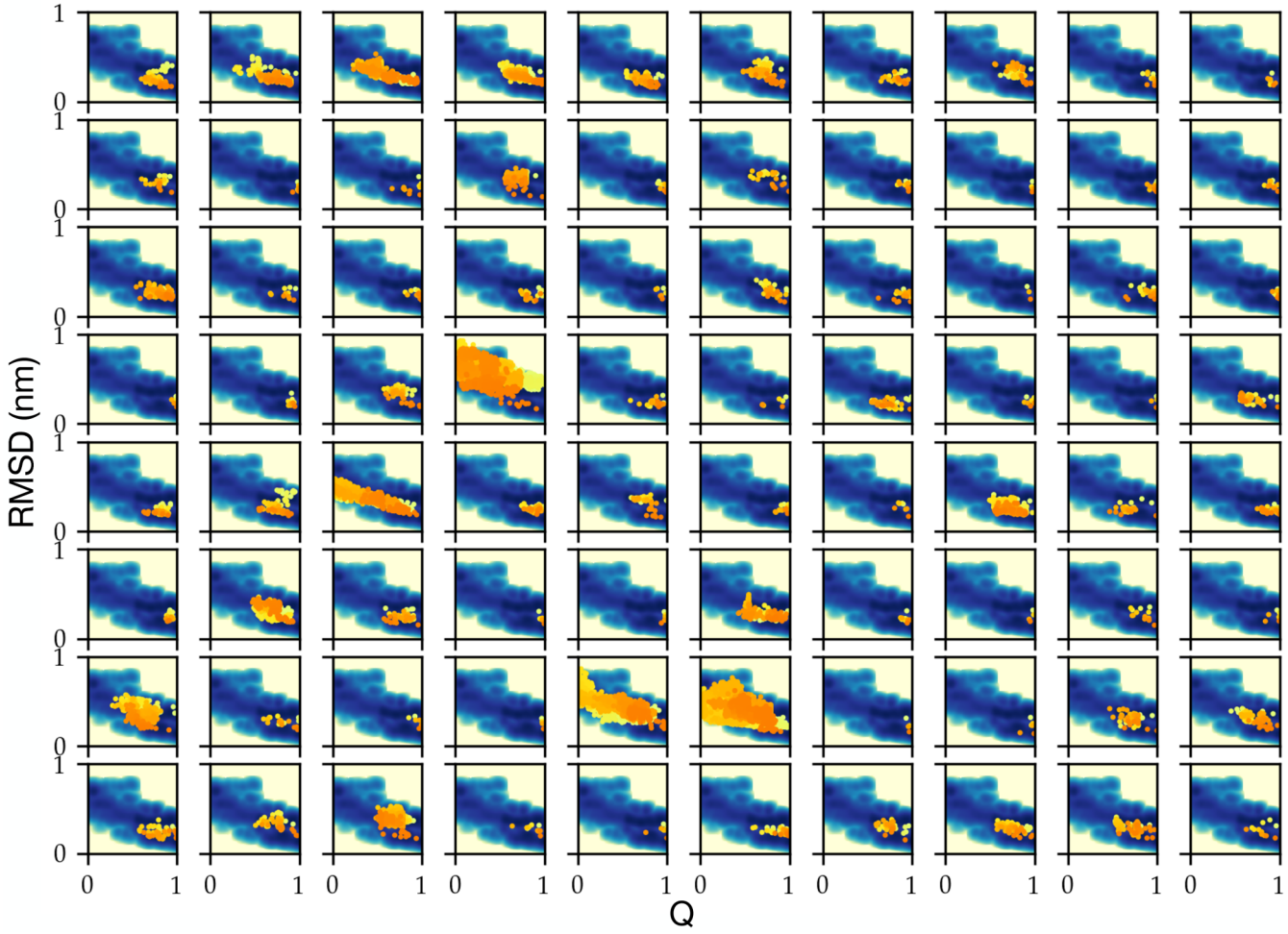


Figure 6. Paths of each reactive trajectory (starting from M and landing in F) found for the GAGA tetraloop are illustrated on its two-dimensional (Q vs RMSD) FES (evaluated at 300 K). Only 7 trajectories went through extensive unfolding (visiting $Q < 0.5$) in order to achieve the M to F transition, out of total of 80 M to F transitions.

are in gauche$^-$/gauche$^-$ conformation. The change in backbone direction (turn) is assisted with the switch in the α torsion of A_5 (from gauche$^-$ to trans conformation). In the misfolded state, on the other hand, the abrupt backbone direction change happens at the preceding nucleotide (G_4), involving both the ζ torsion of C_3 and the α torsion of G_4 (they become a gauche$^+$/gauche$^+$ rotamer). All nucleotides after G_4 (and including G_4) are stacked in this misfolded configuration, and their $\alpha(i)/\zeta(i-1)$ torsions stay as gauche$^-$/gauche$^-$.

Stacking has been reported as a major energetic contributor of nucleic acid folds' stability.[82] Our finding of extensive stacking in this misfolded configuration (the same number of base pairs as in the correctly folded structure) is therefore consistent with the free-energetically competitive nature of this misfolded state. However, one important remaining question is whether this misfolded configuration is a stable folding intermediate or a misfolded end state. To address this question, we analyzed the reactive trajectories.

Analysis of Reactive Trajectories. In an effort to place the misfolded state on the folding pathway, we analyzed each of our biased trajectories by identifying subtrajectories that start from the predefined unfolded state and land in the correctly folded state. Our aim here is to obtain insight into folding paths, and in particular to understand whether the misfolded state is located on this folding path. We note that while this reactive trajectory analysis provides information on the folding paths, the quantities derived from this analysis should not be overinterpreted. As a result of the PTWTE-WTM advanced sampling scheme, our time-continuous trajectories jump often in temperature space and also carry a metadynamics bias on Q and the potential energy. These elements can induce mechanistically artificial folding paths. But despite the biases, the reactive trajectory analysis is still useful for revealing the position of the misfolded state with respect to the folded state during our simulations. We also note that the two states are distinguishable on the RMSD dimension but not on the biased Q dimension.

We first obtained the continuous trajectories from the PTWTE-WTM simulations via so-called demultiplexing, that is, extracting continuous trajectories by tracing them in the temperature ladder.[83] Within the time-continuous trajectories, we identified every single subtrajectory that starts from the unfolded basin (definition in Figure 1B) and ends in the correctly folded state (definition in Figure 1B). The path between the start and end states is illustrated in Figure 5 in Q and RMSD dimensions as a function of time for each U to F reactive trajectory that we found in the GAGA tetraloop (see Figures S7 and S8 for the same analysis for GAAA and GCAA tetraloops, respectively.) We identified a total of 24 U to F

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

transitions for the GAGA tetraloop (Figure 5). Only 4 out of these 24 transitions landed in state F without visiting state M. All other transitions involved visiting state M first. This analysis suggests that the state M is not an end state, but it rather acts as a folding intermediate that might facilitate reaching the correctly folded state, as only 17% of the U to F transitions happened without visiting the state M. Although it facilitates reaching F, it is not a strictly necessary intermediate. Similar observations are also true for GAAA and GCAA tetraloops. We identified a total of 13 and 19 U to F transitions for the GAAA and GCAA tetraloops, respectively, where only 2 out of 13 (15%) and 1 out of 19 (5%) achieved this transition without visiting the M state (Figures S7 and S8).

In a further effort to illuminate the role of this misfolded state in RNA folding, we analyzed the demultiplexed time-continuous trajectories to determine the subtrajectories that start from the state M and end in state F. This analysis revealed a significantly larger number of transition events (M to F) compared to U to F transitions. For GAGA, we found 80 M to F transitions (Figure 6), whereas the number of U to F transitions was 24 (Figure 5). Notably, most of the M to F transitions happened without extensive unfolding. We found that only 7 trajectories involve extensive unfolding (visiting $Q < 0.5$) in order to achieve the M to F transition, which supports the view that this M state may be an intermediate state. Findings were similar for the GAAA and GCAA tetraloops. The number of M to F transitions was significantly larger than the number of U to F transitions within the same set of demultiplexed trajectories, and most of the M to F transitions happened without significant unfolding. (See Figure S9 for GAAA which had 79 M to F transitions, only 9 of which had extensive unfolding, and Figure S10 for GCAA which had 149 M to F transitions, only 8 of which had extensive unfolding.) While our analysis is informative of the possible folding paths, a future direction of further study would be an unbiased analysis of kinetic quantities like rates and fluxes based on Markov state models[28,29,34] and/or diffusion maps[33] with data sets that allow direct quantification.

CONCLUSIONS

Tetraloops are computationally tractable, autonomously folding RNA units that have many characteristics of larger RNAs in their folding behavior, including an abundance of misfolded states whose stability is comparable to that of the native state. The strategy used here to find states that have free energies comparable to the folded state can be summarized as follows. First, appropriate order parameters are chosen to perform advanced sampling simulations and analysis. Then, the free energy surfaces are calculated and free energy basins are identified. Finally, the reactive trajectories are analyzed to find the location of the free energy basins on the folding pathway.

For the tetraloops that we studied here, we identified a misfolded configuration that has a correctly folded stem but a misfolded loop. This misfolded configuration (shared in all GNRA tetraloops simulated here) was found to be free-energetically competitive with the correctly folded state and had the same extensive stacking (but in a different arrangement) as in the correctly folded structure. This highly competitive nature of the stacking-stabilized misfolded state is in agreement with experimentally observed structural diversity in GNRA folds[46−48] and points to one possible structural origin of the competitiveness of the "misfolded" states. Our trajectory analysis suggested that this misfolded state might be an intermediate that facilitates folding rather than being an off-pathway intermediate. One important direction of future work is to explore the extent and/or importance of these nonnative (misfolded) configurations in *in vivo* and *in vitro* conditions where RNA binding assays are performed. The misfolded configurations that we described in this work only contained differences in the loop part of the structure. Similar nonnative configurations have been detected in experiments where a small molecule (malachite-green) binds to the GAGA tetraloop. The malachite-green-binding conformer of GAGA tetraloop has a loop structure different from the loop of the experimental native conformer found in aqueous solution.[47] The structural heterogeneity that may slow down finding the native state also provides RNA with structural versatility that can facilitate RNA to perform various functions,[84] as observed in the GAGA tetraloop and malachite green-binding example.

On the technical side, we point out that the recent improvements in RNA force fields made this detailed study of misfolded states possible. Our observations of the misfolded state are supported by fluorescence experiments[46,48] and are not force field dependent (see the Supporting Text and Figures S11−S13 for the Amber force field results, and also see refs 15 and 18 for similar misfolded configurations when using the Amber force field). We note that it is not straightforward to assess the influence of force field biases on the stability of these intermediates, emphasizing the need for direct experimental data for the comparison of alternative RNA folds. Combined experimental and simulation studies (where experiments can provide direct structural and thermodynamic characterization of the misfolded states) are therefore crucially needed contributions that can rigorously validate force fields against experimental data.[21,85,86] Another interesting technical point concerns the sampling efficiency, as plain PT fails to produce reliable results in a comparable simulation time. Obtaining further insight into the sampling efficiency of the current protocol compared to plain PT is an interesting subject for future work. The improvements in RNA models combined with sampling techniques such as those we used in this work are promising for future computational studies of more complicated RNA structures, like pseudoknots or riboswitches with crucial biological functions.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.1c08038.

Supporting text detailing the simulations with Amber ff99bsc0χ_{OL3} and 14 supporting figures showing the sampling of the order parameters as a function of time, the difference in the free energies, free energy differences as a function of temperature, free energies projected on dihedral angles, reactive trajectory paths, and RMSD fluctuations within the structure making up folded and misfolded clusters of GAGA tetraloop (PDF)

AUTHOR INFORMATION

Corresponding Author

Pablo G. Debenedetti − *Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0003-1881-1728; Email: pdebene@princeton.edu

Authors

Gül H. Zerze − *Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0002-3074-3521

Pablo M. Piaggi − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0003-0225-0081

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.1c08038

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

P.G.D. acknowledges the support of the National Science Foundation (Grant CHE-1856704). P.M.P. was supported by an Early Postdoc.Mobility fellowship from the Swiss National Science Foundation and by the center Chemistry in Solution and at Interfaces funded by the DOE under Award DE-SC0019394. The simulations presented in this work are performed on computational resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

■ REFERENCES

(1) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **1982**, *31*, 147−157.

(2) Bartel, D. P.; Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **1993**, *261*, 1411−1418.

(3) Cech, T. R. The ribosome is a ribozyme. *Science* **2000**, *289*, 878−879.

(4) Mortimer, S. A.; Kidwell, M. A.; Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **2014**, *15*, 469−479.

(5) Uhlenbeck, O. Keeping RNA happy. *RNA* **1995**, *1*, 4.

(6) Herschlag, D. RNA chaperones and the RNA folding problem. *J. Biol. Chem.* **1995**, *270*, 20871−20874.

(7) Treiber, D. K.; Williamson, J. R. Exposing the kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **1999**, *9*, 339−345.

(8) Bevilacqua, P. C.; Blose, J. M. Structures, kinetics, thermodynamics, and biological functions of RNA hairpins. *Annu. Rev. Phys. Chem.* **2008**, *59*, 79−103.

(9) Pan, J.; Woodson, S. A. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* **1998**, *280*, 597−609.

(10) Miner, J. C.; Chen, A. A.; García, A. E. Free-energy landscape of a hyperstable RNA tetraloop. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 6665−6670.

(11) Deng, N.-J.; Cieplak, P. Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophys. J.* **2010**, *98*, 627−636.

(12) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.* **2002**, *317*, 493−506.

(13) Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. Structural insight into RNA hairpin folding intermediates. *J. Am. Chem. Soc.* **2008**, *130*, 9676−9678.

(14) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578−1590.

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

(16) Gil-Ley, A.; Bottaro, S.; Bussi, G. Empirical corrections to the amber RNA force field with target metadynamics. *J. Chem. Theory Comput.* **2016**, *12*, 2790−2798.

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

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

(19) Banás, P.; Hollas, D.; Zgarbová, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E., III; 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.

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

(21) Bottaro, S.; Nichols, P. J.; Vögeli, B.; Parrinello, M.; Lindorff-Larsen, K. Integrating NMR and simulations reveals motions in the UUCG tetraloop. *Nucleic Acids Res.* **2020**, *48*, 5839−5848.

(22) Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banas, P.; Sponer, J.; Bussi, G. Fitting corrections to an RNA force field using experimental data. *J. Chem. Theory Comput.* **2019**, *15*, 3425−3431.

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

(24) Chen, A. A.; García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820−16825.

(25) Bottaro, S.; Banas, P.; Sponer, J.; Bussi, G. Free energy landscape of GAGA and UUCG RNA tetraloops. *J. Phys. Chem. Lett.* **2016**, *7*, 4032−4038.

(26) Brion, P.; Westhof, E. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113−137.

(27) Pan, J.; Thirumalai, D.; Woodson, S. A. Folding of RNA involves parallel pathways. *J. Mol. Biol.* **1997**, *273*, 7−13.

(28) DePaul, A. J.; Thompson, E. J.; Patel, S. S.; Haldeman, K.; Sorin, E. J. Equilibrium conformational dynamics in an RNA tetraloop from massively parallel molecular dynamics. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.

(29) Pinamonti, G.; Paul, F.; Noé, F.; Rodriguez, A.; Bussi, G. The mechanism of RNA base fraying: Molecular dynamics simulations analyzed with core-set Markov state models. *J. Chem. Phys.* **2019**, *150*, 154123.

(30) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. Exploring the energy landscape of a small RNA hairpin. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530.

(31) Sarkar, K.; Nguyen, D. A.; Gruebele, M. Loop and stem dynamics during RNA hairpin folding and unfolding. *RNA* **2010**, *16*, 2427−2434.

(32) Russell, R.; Das, R.; Suh, H.; Travers, K. J.; Laederach, A.; Engelhardt, M. A.; Herschlag, D. The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J. Mol. Biol.* **2006**, *363*, 531−544.

(33) Bottaro, S.; Gil-Ley, A.; Bussi, G. RNA folding pathways in stop motion. *Nucleic Acids Res.* **2016**, *44*, 5883−5891.

(34) Huang, X.; Yao, Y.; Bowman, G. R.; Sun, J.; Guibas, L. J.; Carlsson, G.; Pande, V. S. *Biocomputing 2010*; World Scientific, 2010; pp 228−239.

(35) Townshend, R. J.; Eismann, S.; Watkins, A. M.; Rangan, R.; Karelina, M.; Das, R.; Dror, R. O. Geometric deep learning of RNA structure. *Science* **2021**, *373*, 1047−1051.

(36) Chastain, M.; Tinoco, I., Jr. *Progress in Nucleic Acid Research and Molecular Biology*; Elsevier, 1991; Vol. *41*; pp 131−177.

(37) Moore, P. B. Structural motifs in RNA. *Annu. Rev. Biochem.* **1999**, *68*, 287−300.
(38) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; d'Aubenton Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1364−1368.
(39) Batey, R. T.; Rambo, R. P.; Doudna, J. A. Tertiary motifs in RNA structure and folding. *Angew. Chem., Int. Ed.* **1999**, *38*, 2326−2343.
(40) Jaeger, L.; Michel, F.; Westhof, E. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* **1994**, *236*, 1271−1276.
(41) Woese, C.; Winker, S.; Gutell, R. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8467−8471.
(42) Varani, G. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379−404.
(43) Heus, H. A.; Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **1991**, *253*, 191−194.
(44) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **1996**, *273*, 1678−1685.
(45) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H.; Pardi, A. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **1996**, *264*, 968−980.
(46) Menger, M.; Eckstein, F.; Porschke, D. Dynamics of the RNA hairpin GNRA tetraloop. *Biochemistry* **2000**, *39*, 4500−4507.
(47) Flinders, J.; DeFina, S. C.; Brackett, D. M.; Baugh, C.; Wilson, C.; Dieckmann, T. Recognition of planar and nonplanar ligands in the malachite green−RNA aptamer complex. *ChemBioChem* **2004**, *5*, 62−72.
(48) Zhao, L.; Xia, T. Direct revelation of multiple conformations in RNA by femtosecond dynamics. *J. Am. Chem. Soc.* **2007**, *129*, 4118−4119.
(49) Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T., III; Darden, T.; Duke, R.; Gohlke, H.; et al. *AMBER 14*; University of California, San Francisco, 2014.
(50) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(51) MacKerell, A. D., Jr; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(52) Haldar, S.; Kuhrova, P.; Banas, P.; Spiwok, V.; Sponer, J.; Hobza, P.; Otyepka, M. Insights into stability and folding of GNRA and UNCG tetraloops revealed by microsecond molecular dynamics and well-tempered metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 3866−3877.
(53) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(54) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(55) Zgarbová, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., III; 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.
(56) Steinbrecher, T.; Latzer, J.; Case, D. Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405−4412.
(57) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(58) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(59) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(60) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(61) Bonomi, M.; Parrinello, M. Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.* **2010**, *104*, 190601.
(62) Gil-Ley, A.; Bussi, G. Enhanced conformational sampling using replica exchange with collective-variable tempering. *J. Chem. Theory Comput.* **2015**, *11*, 1077−1085.
(63) Langridge, R.; Marvin, D.; Seeds, W.; Wilson, H.; Hooper, C.; Wilkins, M.; Hamilton, L. The molecular configuration of deoxyribonucleic acid: II. Molecular models and their fourier transforms. *J. Mol. Biol.* **1960**, *2*, 38−IN12.
(64) Zerze, G. H.; Stillinger, F. H.; Debenedetti, P. G. Thermodynamics of DNA Hybridization from Atomistic Simulations. *J. Phys. Chem. B* **2021**, *125*, 771.
(65) Best, R. B.; Mittal, J. Balance between α and β structures in ab initio protein folding. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.
(66) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.
(67) McGovern, M.; De Pablo, J. A boundary correction algorithm for metadynamics in multiple dimensions. *J. Chem. Phys.* **2013**, *139*, 084102.
(68) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. Protein folding and ligand-enzyme binding from bias-exchange metadynamics simulations. *Curr. Phys. Chem.* **2012**, *2*, 79−91.
(69) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255−268.
(70) Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A: At., Mol., Opt. Phys.* **1985**, *31*, 1695.
(71) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(72) Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(73) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(74) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(75) 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.
(76) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(77) Bonomi, M.; et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670−673.
(78) Tiwary, P.; Parrinello, M. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736−742.
(79) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide folding: when simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.
(80) Gonzalez-Aleman, R.; Hernandez-Castillo, D.; Caballero, J.; Montero-Cabrera, L. A. Quality threshold clustering of molecular dynamics: a word of caution. *J. Chem. Inf. Model.* **2020**, *60*, 467−472.

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695

(81) Bottaro, S.; Lindorff-Larsen, K. Mapping the universe of RNA tetraloop folds. *Biophys. J.* **2017**, *113*, 257−267.
(82) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564−574.
(83) Šponer, J.; Islam, B.; Stadlbauer, P.; Haider, S. In *Quadruplex Nucleic Acids As Targets For Medicinal Chemistry*; Neidle, S., Ed.; Annual Reports in Medicinal Chemistry, Vol. 54; Academic Press, 2020; pp 197−241.
(84) Dethoff, E. A.; Petzold, K.; Chugh, J.; Casiano-Negroni, A.; Al-Hashimi, H. M. Visualizing transient low-populated structures of RNA. *Nature* **2012**, *491*, 724−728.
(85) Bottaro, S.; Bussi, G.; Kennedy, S. D.; Turner, D. H.; Lindorff-Larsen, K. Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Sci. Adv.* **2018**, *4*, No. eaar8521.
(86) Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to understand atomistic molecular dynamics simulations of RNA and protein−RNA complexes? *Wiley Interdiscip. Rev.: RNA* **2017**, *8*, No. e1405.

https://doi.org/10.1021/acs.jpcb.1c08038
J. Phys. Chem. B 2021, 125, 13685−13695